Dynamic Chain Analysis by Bipartite Network for Medicine Selection

Xinming Yin\textsuperscript{1,2}, Yi Guo\textsuperscript{*}, Zhiwei Cao\textsuperscript{2} and Min Xiong\textsuperscript{2}

\textsuperscript{1} School of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, China
\textsuperscript{2} The Third Research Institute of Ministry of Public Security, Shanghai 200031, China
Email: guoyi@ecust.edu.cn

Abstract. The rapid development of society has brought about the uncertainty of social relations, and the structure of social networks is constantly changing. Chain forecast, as an effective method, plays an increasingly important role in understanding the dynamic nature of the network and determining future relationships, combining the structural characteristics of the present status of the network to foresee the possible existence of future network nodes, this paper proposes a Chain forecast method for Disease treatment and a Chain forecast method based on bipartite networks (such as treatment correspondent diagrams). In order to verify the forecast effect of the method, we selected several Chain forecast algorithms for check. The results prove that our proposed method is better than other methods based on Chain forecast.

Keywords. Bipartite network; chain analysis; medicine selection.

1. Introduction

Social network is a kind of topology, which is used to describe model relationships among objectives, individuals, enterprises, government organizations and communities in real life. A lot of complicated systems we encounter in various regions or in various ways may be visualized using social network tools\cite{1}. Social network also provides a way for us to understand the composition, evolution, and interrelations of these complicated systems. Network analysis of complex system can help to provide more valuable information, which can provide support for further research.

Chain forecast is mainly used to analyze the evolution status of future network, with the help of the features of nodes in the net and the observation results, it also predicts the probability of future connections between two nodes\cite{2}.

Among them, the past status of the social network is the basis for analyzing the future status of the network and determining what new changes will occur in the network.

The social relationships among individuals, enterprises, government organizations and communities will change over time. Therefore, the social networks that simulate the real-world relationships will change accordingly, and it will evolve to some extent as these relationships change. In the constant evolution of social networks, new Chains and nodes gradually appear on the networks where old Chains and nodes may also disappear from. Chain forecast is mainly used to determine the dynamic relationship of targets in social networks. Chain forecast can be applied in different fields \cite{3-6}. For example, in academic research networks, Chain forecast can help to find co-authors of a certain academic research; in online shopping networks, Chain forecast can help customers to recommend products; in protein interaction networks, Chain forecast can help to predict protein...
Interaction; In addition, Chain forecast can help to predict future citations, evaluates people’s activities and categorize them, and decide the relationship between diseases.

This paper aims to make a treatment method networking by available data collected from www.drugs.com, that mainly explains about which Medicines can cure which diseases. After completing this network, we can determine which diseases can be cured by a given Medicine through the Chain forecast method. The proposed internal Chain method can facilitate the forecast of Chains in Disease treatment. To prove the effectiveness of the method presented, four Chain forecast algorithms are chosen based on similarity to forecast the network. At the same time, the experimental results of the internal chain method are used for comparison with the results of the other four similarity-based Chain forecast methods. The presented method achieved better performances with respect to accuracy, recall rate and F-measure criterion shown in experiment results.

2. Chain Forecast
Chain forecast of social networks is mainly used to forecast the future structure of the network, which also can be defined as a topology diagram. Nodes represent data in the network, while Chains represent relationships. It is the main research content of Chain forecast that judges the future connection status of unconnected Chains, and the Chain forecast algorithm is responsible for assigning a certain forecast score to each node that may have connections.

The greater the predicted score value assigned between two nodes, the more likely that the two nodes will be connected in the future. The forecast edges are arranged according to the size of the forecast score. The front edges are the ones most likely to produce connection in the future. The accuracy of the forecast varies with the algorithm applied.

The dichotomous social network is a specific type of social network. In a dichotomous social network, nodes are distributed in two different cluster areas. The predicted Chain merely exists between nodes located in different cluster areas. There is no link relationship between nodes from the same cluster area. In practical applications, many social networks have a dichotomous network structure [7-9]. Since most Chain forecast methods are aimed at a single-mode network structure, it is difficult to apply traditional Chain forecast algorithms directly to bipartite networks. Therefore, based on previous research, it is suggested to convert the Chain forecast of the bipartite social network into the Chain forecast of the single-mode social network.

As shown in figure 1(a), in the bipartite graph the single mode graph containing simply X nodes is represented as X-projection, and the single mode graph including simply Y nodes is represented as y-projection. In the bipartite graph, any two X nodes that have a neighbour relationship with the same Y node, then these two nodes establish a link in the X projection. For example, due to the X1 and X2 nodes have connections with the Y1 node in figure 1(a), the link between the X1 and X2 nodes shown in figure 1b is established in the X projection.

The easiest way to transform a bipartite network into a single-mode network is to build an unweighted network without considering the weight of the network, that is, without considering the cooperative frequency relationship between nodes. However, since the single-mode network has a larger amount of information than the bipartite social network, the bipartite network shown in figure 1 is converted into a weighted single mode network. This paper gives weight to a single-mode network based on the number of partnerships. For example, in figure 1, based on the situation that X1 and X2 have the same neighbour node, the weight calculation result between X1 and B is X2; Y1 is connected with X1 and X2 in the original graph. This paper refers to this weighting method as the weighting function [10], and is expressed as:

\[ w(u, v) = | N(u) \cap N(v) | \]

(1)

3. Related Work
So far, many Chain forecast methods have been proposed. Most of them are Chain forecast methods measured by similarity [11-12]. The principle point of the Chain forecast method based on similarity is
that the more amount of common neighbors of node X and node Y, the greater possibility of a connection between the two nodes in the future. In this paper, X and Y are the nodes in the network, and \( \Gamma(x) \) and \( \Gamma(y) \) are the set X nodes and neighbor nodes of the Y node in the network. The four most commonly used Chain forecast methods based on similarity are listed below:

1. Common Neighbors (CN): The premise of this similarity index is that there is a positive correlation between the probability that two nodes will produce connected edges in the coming days and the amount of collective neighbors of the two nodes [13]. Because of its simplicity, it is one of the most widely used similarity indicators in the Chain forecast. Its mathematical formula is as follows:

\[
CN(x, y) = |\Gamma(x) \cap \Gamma(y)|
\]

(2)

2. Jaccard Coefficient (JC): Jaccard Coefficient is a standardized form of public nearest neighbor similarity method [14]. It relates to the ratio of the number of common neighbors with the total number of all node neighbors. The larger the number of public neighbors, the more the score obtained by the similarity method. Its mathematical formula is as follows:

\[
JC(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|}
\]

(3)

3. Adamic/Adar (AA): This method was innovatively presented by Adamic and Adar, and it aims to quantify the degree of similarity between two web pages [15]. The basic idea of this method is that the contribution rate of a common neighbor node with a small degree is higher than that with a large degree. Therefore, the AA indicator sets a weight value to each node varying with the degree of the common neighbor node, and the weight is equal to 1 divide logarithm of the degree of the node. The AA indicator is defined as:

\[
AA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log|\Gamma(z)|}
\]

(4)

4. Preferential Attachment (PA): The premise of this similarity index is that the probability of two nodes in the network establishing a new connection is proportional to the number of neighbors of the node [16]. On the other hand, if the nodes have a larger number of neighbor nodes, the possibility of establishing a new connection is greater. Newman [17] believed that the possibility of cooperation

---

**Figure 1.** (a) bipartite graph case, (b) X Axis projection, (c) Y Axis projection.
between the X node and the Y node in the cooperative network is proportional to the number of collaborators. Its mathematical formula is as follows:

\[ PA(x, y) = |\Gamma(x)| \cdot |\Gamma(y)| \]  

(5)

4. Proposed Method

In the research of this paper, the initial data is collected from www.drugs.com. Data include medicines prescribed by doctors for diseases, percentages of users with various medicines, Medicine classification, pregnancy category, and so on. The nodes in the treatment network refer to Medicines and diseases. Links are established among the three most common drugs based on treatment. Assume that the percentage of other Medicines used is low, and the corresponding Chains are not included. Table 1 shows a specific part of the data collected from the website.

According to the disease Medicine relationship in Table 1, we establish a bipartite network with Medicines and diseases as node clusters. Figure 2 is a network example of establishing a Chain structure between selected diseases and the top three Medicines with the highest utilization rate of each disease. This paper attempts to explore a method to predict which Medicine can be prescribed as the Medicine corresponding to a disease through the related Chain forecast method based on the existing Medicine disease relationship.

| Disease   | Medicine                | Percentage |
|-----------|-------------------------|------------|
| asthma    | Flunisolide             | 10         |
|           | Levosalbutamol          | 10         |
|           | L-salbutamol            | 9          |
|           | Doxycycline             | 9          |
| Bronchiectasis | hydrochloride         |            |
|           | Doxycycline             | 7          |
|           | Salbutamol Sulfate     | 8          |
| Duodenitis | Omeprazole              | 6          |
|           | Pantoprazole            | 5          |
|           | Sterilization XR        | 10         |
| Sinusitis | Decylamine              | 8          |
|           | Xianliteng              | 9          |
|           | Sumatriptan succinate   | 8          |
| Migraine  | Sumatriptan             | 8          |
|           | Naratriptan             | 9          |

Table 1. Use part of data.

Figure 2. Medicine-disease network example.
According to the Disease-Medicine relationship in table 1, we have established a Bipartite Network with Medicines and diseases as node clusters. Figure 2 is an example of a network where we select a disease, establish the Chain structure between the disease and top three Medicines with the highest usage rate for this disease. The purpose of this study is to predict which Medicines can be written as the treatment corresponding to a certain disease based on the existing Medicine-disease relationship and related Chain forecast methods.

Unlike traditional methods, researchers have proposed many Chain forecast methods based on Bipartite Networks [18-20]. An internal Chain method for Chain forecast in the Disease-Medicine binary network is proposed. In this method, a private Chain called internal Chain is presented, and Chain forecast is made through these Chains. After the bipartite network is transformed into a weighted single-mode network, it indicates that the internal Chain connection in the original network has been determined. In a binary network, a node pair that does not yet have a connection is selected. If the Chain established between the pair of nodes does not cause a change in the topology of the projected graph, the Chain between the selected nodes is called an internal Chain.

The internal Chains proposed mainly refer to those Chains that do not yet have a connection in the bipartite network, and satisfy that when a connection is added between them, the projected image of the bipartite network will not change. In our research, all known Chains cannot be called internal Chains. At the same time, in order to ensure that necessary Chains are not missing and unnecessary Chains are not considered, the threshold is set. Among them, those Chains whose weight value is higher than the threshold configured are regarded as internal Chains.

5. Experimental Results

According to the disease Medicine relationship in table 1, we establish a binary network with Medicines and diseases as node clusters. Figure 2 is a network example of establishing a Chain structure between selected diseases and the top three Medicines with the highest utilization rate of each disease. The purpose of this study is to predict which Medicine can be written as the Medicine corresponding to a disease through the related Chain forecast method based on the existing Medicine disease relationship. This paper mainly analyzes the performance of the algorithms mentioned in the Medicine disease network. In addition, such network data can be obtained from www.Medicines.com Obtained on the official website. The network selected in this paper consists of 500 edges. In order to test the accuracy of the related algorithm in predicting, we randomly select 100 edges as the test group and another 400 Chains as the training group. Then, based on the disease Medicine network, the algorithm mentioned in this paper is used for Chain forecast. In addition, internal Chains are found among nodes that have not yet established the connection edge, and those Chains whose weight value is higher than the preset threshold value are defined as internal Chains. Next, the discovered internal Chains are sorted according to the weight value, and the Chains with the weight ranking in the top 100 are selected. Aiming to compare and analyze the prediction performance of the methods mentioned above, we choose four similarity based methods to predict the Chain of the disease Medicine binary network. The score (XJ) of each pair of unconnected nodes in the training set is calculated. Similar to previous studies, this fraction represents the similarity between nodes X and y. The higher the predictive score, the more likely the missing Chain will show up in the future. In addition, aiming to compare the prediction results, the common neighborhood (CN), JC, PA and Adamic/ADAR (AA) methods are used as the calculation functions of the fractional values. All forecast Chains are arranged in descending order according to the forecast score, and the first 100 Chains with higher score are selected as forecast results. Finally, we compare the forecast Chains of the internal Chain algorithm (IL), the common neighbor algorithm (CN), the JC algorithm, the PA algorithm and the Adam / ADAR algorithm (AA) with the previously extracted test set data. At the same time, the forecast accuracy, recall rate and F-measure are selected as performance evaluation criteria. Figure 3 describes the effect comparison results of five methods in forecast accuracy, recall rate and F-measure. As shown in figure 3, the presented method reaches the highest forecast effect among the three indicators. Compared with the other four similarity based Chain forecast algorithms, the algorithm presented in this paper reaches
a higher effect of Chain deletion recovery. To sum up, according to the forecast results, the algorithm presented in this paper reaches a high forecast success rate, which can more easily predict which Medicine can be used to cure a certain disease.

![Figure 3. Comparison of the effects of five different methods.](image)

6. Conclusion
So far, Chain forecast in binary networks has become the focus of research in many academic fields. Researchers from various fields continue to deepen research on this theme. In recent years, the Chain forecast of the binary network is mostly concentrated in the Disease-Medicine network. You can usually obtain a network dataset of Medicine recommendation relationships from the website (www.Medicines.com). The network dataset is a binary network. Compared with Chain forecast in a single mode network, predicting Chains in a bipartite network is a more complicated and difficult. In addition, Chain forecast algorithms based on single-mode networks cannot be applied to predict bipartite networks without conversion. Therefore, in order to improve the effect of the Chain prediction of the bipartite network, it is first necessary to convert the existing bipartite network into a single mode network. Compared with other classical algorithms, the internal Chain method with higher forecast accuracy can be better applied to Disease treatment. Similarly, similarity-based Chain forecast methods such as public neighbors, Jaccard coefficients (JC), priority connections (PA), and Adamic/Adar (AA) have also been applied to dichotomous networks. When considering the predictive effects of five different methods, a certain breakthrough has been made in which Medicines can be used to treat which diseases.

Acknowledgments
The projects funded by this paper are: National Key Research and Development Program (2018YFC0807105, 2018YFC0807106); Shanghai Science and Technology Talent Program (17XD1420400).

References
[1] Li F, He J, Huang G, et al. 2015 Node-coupling clustering approaches for link prediction Knowledge-Based Systems 89 (2) 669-680.
[2] Tan S, Wu J, Lü L, et al. 2016 Efficient network disintegration under incomplete information: The comic effect of link prediction Scientific reports 6 (1) 22916.
[3] Liu J 2004 An Introduction to Social Network Analysis (Beijing: Social Sciences Academic Press) pp 15-20.
[4] Zhou T, Wang B, Han X, et al. 2010 Social network analysis and its application in the prevention and control of propagation for public opinion and the epidemic Journal of Systems Engineering 25 (6) 742-754.
[5] Zhang Q, Shang M and Lü L 2010 Similarity-based classification in partially labeled networks International Journal of Modern Physics C 21 (6) 813-824.
[6] Vinayagam, A, Stelzl, U, Foulle R, et al. 2011 A directed protein interaction network for investigating intracellular signal transduction Science Signaling 4 (189) rs8.

[7] Lü L 2010 Link prediction on complex networks Journal of University of Electronic Science and Technology of China 39 (5) 651-660.

[8] Liu H, Lü L and Zhou T 2011 Uncovering the network evolution mechanism by link prediction SCIENTIA SINICA Phys., Mech. & Astron. 41 (7) 816-823.

[9] Cannistraci C V, Alanis-Lobato G and Ravasi T 2013 From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks Sci. Rep. 3 (2) 1613.

[10] Zhang Y 2011 Link Prediction in Directed and Weighted Networks (Xiangtan: Xiangtan University).

[11] Liu S, Ji X, Liu C, et al. 2014 A complex network evolution model for network growth promoted by information transmission Acta Phys. Sin. 63 (15) 158902.

[12] Tan F, Xia Y and Zhu B 2014 Link prediction in complex networks: A mutual information perspective PLoS ONE 9 (9) e107056.

[13] Lü L and Zhou T 2013 Link Prediction (Beijing: Higher Education Press) pp70-75.

[14] Zhou T, Lü L and Zhang Y 2009 Predicting missing links via local information The European Physical Journal B 71 (4) 623-630.

[15] Liu W and Lü L 2010 Link prediction based on local random walk Europhysics Letters 89 (5) 58007.

[16] Liu S, Liu H, Chen Q, et al. 2017 Link prediction algorithm based on network representation learning and random walk Journal of Computer Applications 37 (8) 2234-2239.

[17] Newman M E J 2001 Clustering and preferential attachment in growing networks Phys. Rev. E 64 (2) 025102.

[18] Bai M, Hu K and Tang Y 2011 Link prediction based on a semi-local similarity index Chinese Physics B 20 (12) 498-504.

[19] Cao Z, Zhang Y, Guan J, et al. 2018 Link prediction based on quantum-inspired ant colony optimization Scientific Reports 8 (1) 13389.

[20] Daminelli S, Thomas J M, Duran C, et al. 2015 Common neighbors and the local-community-paradigm for topological link prediction in bipartite networks New Journal of Physics 17 (11) 113037.