Volume 6, Issue 6, June 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Link Prediction in Complex Network Using Fuzzy Logic

Heemakshi Malhi*, Mini Ahuja*
Computer Science & GNDU, Amritsar,
Punjab, India

Abstract— Complex networks has become significant part of the digital world. Many large scale problems can only be handled using complex networks. Evaluating the optimistic link in these networks is still a challenging issue. Link prediction in directed network is attracting growing interest among many network scientists. Compared with predicting the existence of a link, determine its direction is more complicated. It proposed efficient solution named Local Directed Path to predict link direction. By adding an extra ground node to the network, one can solve the information loss problem in sparse network, which makes the method effective and robust. As a quasi-local method, link prediction using fuzzy logic can deal with large –scale networks in a reasonable time. The overall objective is to evaluate various shortcomings in them.

Keywords— Complex Network, Link prediction, Link Prediction methods, Fuzzy logic, Fuzzy logic applications

I. INTRODUCTION

Complex network have grabbed a lot of recognition and like real world networks, they possess interesting and attractive features that help us to get a clear picture of them. Complex networks comprise of all systems where linkage, communication and interaction among various physical or logical inter-connected components merge to form complex and intricate network web. In view of their non-trivial and intricate network structure, such networks are usually referred to as complex and hence network topology Complex networks came into view in large range of aspects in social and natural sciences. Commonly spoken examples of the complex networks are internet, power grids. World Wide Web, transportation systems, ecosystems, food web, social networks, neural networks, grid and cloud computing, distribution networks etc. **[1]**.

II. LINK PREDICTION IN COMPLEX NETWORKS

In a given network what is major challenge encountered is inference of links whose occurrence is anticipated in future on basis of the edge, node and topological characteristics they possess thereby leading to a fundamental problem known as link prediction. In a network representation, at a particular instance of time, main objective is to predict what and all links in the network will occur in next instance of time. The nodes refer to people or entities in social terms and the edges represent linkage, interaction between the entities in social networks. The link prediction problem has been defined and illustrated in various ways if the link prediction problem is perceived on basis of data mining perspective, then it appears much similar to link mining procedure since various real world networks comprised of different diversity of entities and these entities are connected via multiple types of relationships. A major hurdle in process of link mining task is the problem related to mining of intensively linked datasets that are required to dig down to acquire knowledge about relationships and links**[12]**.

III. VARIOUS LINK PREDICTION METHODS

A. Traditional Link Prediction:

1) Similarity Measure Based Strategies

As connection expectation has its utilization in present world, because of which a lot of association anticipating strategies have been proposed. The comparability based procedures are more regularly utilized. There is a lot of procedures to find the similarity which focus on the information of hub qualities and framework structure. It's difficult to get the authentic hub qualities by virtue of security issues, so the procedures rely on upon equivalence of framework which is largely used. Liben-Nowell et al. [7] gave the significance of connection expectation. Join estimate was realized to portray the score (a, b). There are various procedures in light of likeness of hub like ordinary neighbor, particular extension. The quantity of neighbors that hubs ‘a’ and ‘b’ have in common in between them.

$$CN(a, b) = \Gamma(a) \cap \Gamma(b) \qquad \ldots (1)$$

Where Γ (a) and Γ (b) are the neighbor sets of node ‘a’ and ‘b’.

2) Classification Model Based Strategies

Join expectation issue can be considered as a gathering based issue. The arrangement model is adaptable than others displays when analysed. The segments which are removed from the equivalence based strategy can be utilized as a part of collection models for predicting links.

B. Heterogeneous Link Prediction

Research has been made on homogeneous systems, yet few of them demonstrates the work on heterogeneous systems for inducing the issue in connection expectation.

1) Normal Link Prediction

For taking care of the connection expectation in heterogeneous system there are two conceivable ways:(1)make beyond any doubt that all connections are dealt equally.(2)to study different sorts of connections and to discover relationship between them [4].

Davis et al. [6] demonstrated that triad statistics has higher computational many-sided quality and introduced the unsupervised strategy MRLP. It demonstrated that MRLP is superior to anything other old connection forecast techniques. As the unsupervised strategy was not adaptable so Davis presented the new multi-social managed technique. The effect score is as per the following:

$$\text{Flow}(a, b, c) = \text{score}(a).\ \alpha. \frac{\text{Weight}(a,b,c)}{\text{degree}(a,c)} +$$

$$\text{score}(a)\ \alpha \sum_{d \neq c}^{k} \left[\sigma(c,d).\frac{\text{weight}(a,b,c)}{\text{degree}(a,c)}\right] / (|E(a,b)| - 1) \qquad \text{.....}\ (2)$$

Where ‘a’ and ‘b’ are hubs, α is the Katz variable, score (a) is the likelihood of a connection between the source hub and hub a, and E (a, b) −1 is the quantity of connection structures between hub ‘a’ and ‘b’ with the exception of sort ‘d’.

2) Moderator Link Prediction

Informal communities have become famous in present world. Numerous sites have been produced and individuals are having different number of records in those sites. A single individual can have various records in the meantime. Kong et al. [20] made his work on stay joins on different systems. At the point when the zone of the anchor connection is known then we extracted the heterogeneous components from numerous systems and gave an answer called MNA (multi-anchoring system).

C. Temporal Link Prediction

In this present system, most of interpersonal associations are dynamic and changeable, which bring more challenges for the prediction of connection.

1) Matrix and Tensor Decomposition Based Strategies

Daniel et al. [18] displayed the association between two link finding methods, framework based and tensor-based, for bipartite diagrams. The co-creation framework is genuinely a normal bipartite framework, where a general rule the type of nodes are experts and gatherings. They additionally exhibited how Katz procedure is expansive to bipartite diagrams, and utilized the truncated SVD to devise a versatile system which was then used to register a "truncated" Katz score. The disadvantage of the tensor-based system is that the computational cost is high. They likewise demonstrated the convenience of regular three-dimensional structure of temporal connection.

2) Time Series Model Based Strategies

Paulo et al. [20] as opposed to utilizing the time component utilized topological networks in order to expel the confinement of old systems. They cleared up headway of framework's using time arrangement issues. In any case, the time arranged model has a limited application, in light of the fact that the system must be associated with the condition where there are bottomless time course of action of event of connection. Firstly, frameworks which are comparative with time of the framework are figured using key likeness measure computation, for instance, PA, CN, etc. At that point, to anticipate the future values an arrangement of surely understood factual determining models, for example, moving normal irregular walk, and so on were utilized. Thus figured qualities will be utilized as a part of directed and unsupervised learning [15].

IV. FUZZY LOGIC IN LINK PREDICTION

Fuzzy logic involves method of reasoning much alike human reasoning. It produces acceptable yet define the result as a response to ambiguous, contradictory, out of bounds, incomplete, distorted fuzzy input. The concept of fuzzy logic resembles the task of decision making in human brain that comprises of all in between possibilities between digital terms Yes or No. which range in between yes or No. This was observed by fuzzy logic inventor Loffi Zodeh. Fuzzy logic is dependent on input possibilities levels to determine definite output [5]. Fuzzy logic may have two values and infers two possible solutions. Fuzzy logic is a flexible machine learning approach accompanied with multi –valued logic allowing intermediate values. Interpretation and execution of commands is taken care by inference mechanism seen in fuzzy logic approach. Usage of imperfect data in a reasonably sensible manner enables fuzzy logic to lessen complexity. Fuzzy logic also allows relative values such as not, yes, a little bit, no so much tec. Fuzzy logic systems suits the scenes where approximate or uncertain reasoning is involved. Dual-valued logic have only two possible values 0/1, right/wrong, yes/no, true/false etc. The way fuzzy logic handles the problems is much similar to how human brain make fast decisions. Fuzzy logic can be used and implemented in software as well as hardware or the combination of both [9]. Each part along with basic fuzzy logic operations will be described in more detail below. The formula for fuzzy logic is given as:

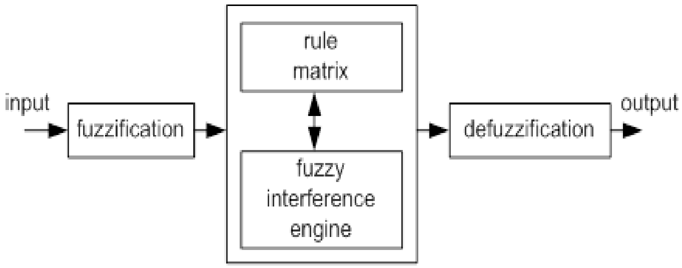

Figure 1: Fuzzy logic controller [12]

Algorithm of Fuzzy logic

a) Linguistic variables and terms are defined.
b) Membership functions are constructed for them.
c) Knowledge base of rules are constructed.
d) Crisp data is converted into fuzzy data sets using membership functions (fuzzification).
e) Evaluation of rules in the rule base is done (interface engine).
f) Results are combined from each rule (interface engine).
g) Output data is converted into non-fuzzy values (defuzzification) [17].

V. FUZZY LOGIC APPLICATIONS

The application areas of fuzzy logic are as given [11]**:**

A) It can be used in Automotive Systems

1) Automatic Gearboxes
2) Four-Wheel Steering

B) It can be used in Electronic Goods for consumers

1) Hi-Fi Systems
2) Photocopiers

C) It can also be used in Domestic Goods

1) Microwave Ovens
2) Refrigerators

VI. LITERATURE SURVEY

YuXiao Zhua et al. [8] propose a parameter subordinate index, which impressively enhances the forecast precision. At last, we demonstrate the pertinence of the proposed list to three genuine testing techniques: colleague testing, discretionary go testing and way based inspecting. In this we inspect exactly how to learn lacking hyperlinks with insignificant sum hubs, particularly hyperlinks in the test accumulation are of lower sum things than the typical discretionary examining.

S. Bondorf et al. [3] This paper describes the way to detect and prevent the intruder within the network. Intruder can jam the traffic. If this happens then the entire network will be rendered unusable. This is known as DDOS attack. With the advent of the technology network is exposed to more and more users. Because of which network is exposed to wide variety of users. The intension of the users is unpredictable. Hence network is exposed to many different types of challenges. This paper describes the problems and solutions associated with the wireless sensor network**.**

A.B Khan et al. [4] this paper suggests how to form a dataset for the given problem. Today database are huge so extracting data from the database using genetic algorithm is complex task. Learning techniques are employed in order to make algorithm collect the data which is required. Collecting data and training process is not a simple task. In this paper it is suggested that how we can collect and train the algorithm to achieve desired objective. Dataset formulation is easy using this algorithm. Supervised learning mechanisms are followed in this case.

J.J Whang et al. [10] the proposed paper describes community detection as an important task in network analysis. A community (also referred to as a cluster) is usually unified vertices which have extra associations intimate the set than exterior. In many social and information networks, such kind of communities perceptibly overlaps. In this paper, Preliminary outcomes display the seed development procedure outperforms other state-of-the-art overlapping community detection procedures in relations of creating unified groups and identifying ground-truth communities. We also show that different seeding approaches then existing approaches, and are thus effective in finding good overlapping communities in real-world networks.

Wang P.et al. [20] In this paper, endeavours to efficiently outline all normal work away at the connection expectation in interpersonal organizations. A sort of connection expectation strategies and connection forecast issues is proposed. Join forecast techniques are discussed, the topology-based measurements and learning-based strategies. Join expectation issues and applications will likewise be displayed. Potential headings and issues are also presented. The objective of the paper is to completely audit, investigate and talk about the best in class of the connection expectation in interpersonal organizations. A systematically class for link prediction techniques and issues is introduced. Regular utilization of connection expectation have also been addressed. Accomplishments and guides of some research groups are presented.

VII. PROPOSED METHODOLOGY

The detailed insertion process for the proposed approach is given below:

Let us assume a network G= (V, E) having 'n' nodes. The mathematical demonstration can be given by adjacency matrix A having elements $A_{ij}=1$ if there exists an edge from v_i to v_j or $A_{ij}=0$. Detection of various communities means separating various nodes into different groups in a network. By exploring various genuine systems and when numerous calculations were made it was detected that few nodes depends on the same community regardless of what algorithm has been applied. In a community A is a subset of M if we have:

$$\sum_{Wj\in Ar} Bij > \sum_{Wj\in A-Ar} Bij \qquad (3)$$

The definition discussed above is similar to the work presented by Radicchi et al. if A=M, else it demonstrates a strong structure in community. At some point if A is a solid community then it can also be said as its core. It is simple to figure out which node makes the core for any given community in a system. It is very simple to find which nodes make the core for a considered community in a network. As per Hu's definition a core in a network is important to find its community. The main motive is to find the community core in a network. On establishing relationship among links, nodes and communities in a network, we concluded that the higher degree nodes have strong agglomeration as compared the one with lower degree. Earlier Hu presented his work on setting every node and its irregular half neighbor to be a community. A node having maximal degree for building the community core in a given network is always important.

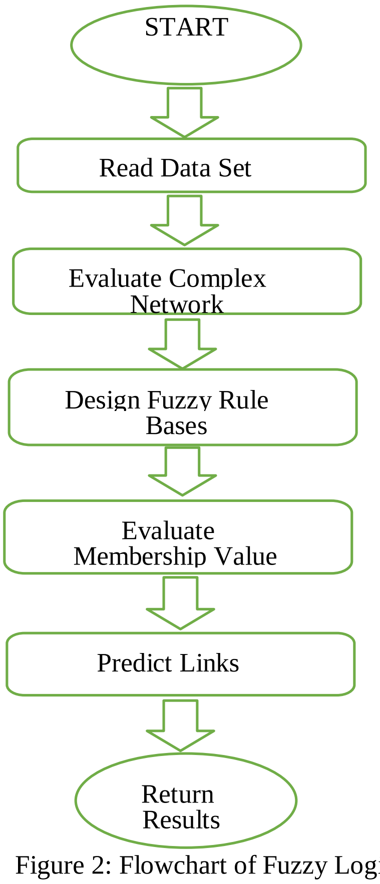

Figure 2: Flowchart of Fuzzy Logic

STEP I: To detect the community core.

The node w_i with maximal degree and its unnamed neighbors in W are detected at priority. M_i represent the set of unnamed neighbors of w_j. For every node w_j in M_i, the value α_j=| $M_i \cap M_j$|+1/d_j is calculated in which d_j denotes the degree of w_j. The node w_j with α_j> 0.5 present the core of community C_{ri}. If $|C_{ri}|/d_i$>0.5 and $|C_{ri}|$>=3 we can link community label to nodes in a C_{ri}. Else the node w_i is labelled symbol 'W'. This step need to be repeated: M=M – {w_i}

STEP II: Addition of nodes

Suppose M as the set of neighbors of C_{ri}. Detect the unnamed node w_k in M. M_k' represents set of neighbors of w_k moves into C_{ri} and is named if necessary conditions are meet.

(1) $|C_{ri} \cap M_k'|$ is equal or greater than $d_k/2$.

(2) $|C_{ri} \cap M_k'|$ is equal or greater than $|A_g \cap M_k'|$ and $|P_u \cap M_k'|$, where A_g is a detected community with label g, P_u is a set of unnamed nodes in M.

Let $C_{ri}=C_{ri}\cup\{w_k\}$ and update M by moving v_k's neighbors which are not in C_{ri} into M. Repeat this process M until there is node satisfying conditions (1) and (2). Let $M=M-C_{ri}$ and visit step I while M is not null.

III: Remove the nodes named 'X'.

If all the nodes have been marked or named, and there is no node named 'X', then all communities in a network have been found. Or for putting nodes into proper communities following steps are needed.

(1) Compute $\beta_s=|M_s'\cap P_y|/d_s$ for every node w_s in P_y, where P_y is a set of nodes named 'X' in network. Search for node w_s with minimal β_s in P_y.

(2) Move w_s into community g if $|A_g\cap M_s'|>|A_i\cap M_s'|$, for all detected communities i=g. If there exist various such communities g, compute $C_g=\sum d_t$ where w_t run over $A_g\cap M_s'$. Then node w_i is moved into community g whose C_g is maximal or node w_s is randomly paced over there.

(3) Let $P_y=P_y-\{w_s\}$ and repeat this process until set P_y is null.

The proposed algorithm generally extracts the best neighborhood from a found node and its neighbors' information. The task of detecting a residential area is comparable to coming a snowball. Which means this algorithm is known as the snowball algorithm. The running time of the proposed algorithm largely is dependent upon the computational need of the measures I and II. It's unimportant to locate a node with a maximal degree in O (n). The functioning time of constructing a residential area primary Cri is O (d_i+2m_i), wherever di is the degree of w_i and n_i is the amount of edges in Cri. The full time usage of this step is approximately O $(K_i^2 m_i)$, wherever m_i is the sum total number of neighborhood i and k_i may be the mean vertex degree in the extracted neighborhood [18]. The running time of the proposal algorithm is approximately O (m+n), wherever 'm' is the amount of vertices and 'n' the amount of edges in the network. The expression can be computed as:

$$triangle(a; x, y, z) = \begin{cases} 0\,, & a \le x \\ \dfrac{a-x}{y-x}\,, & x \le a \le y \\ \dfrac{z-a}{z-y}\,, & y \le a \le z \\ 0, & z \le a \end{cases} \quad \text{.... (4)}$$

VIII. RESULTS AND DISCUSSION

The proposed algorithm is tested on different images. The algorithm is applied using different performance indices Accuracy, Bit Error Rate (BER). To be able to implement the proposed algorithm, design and implementation has been performed in MATLAB using image processing toolbox with some performance metrics. Result shows our proposed approach gives better results than the existing techniques.

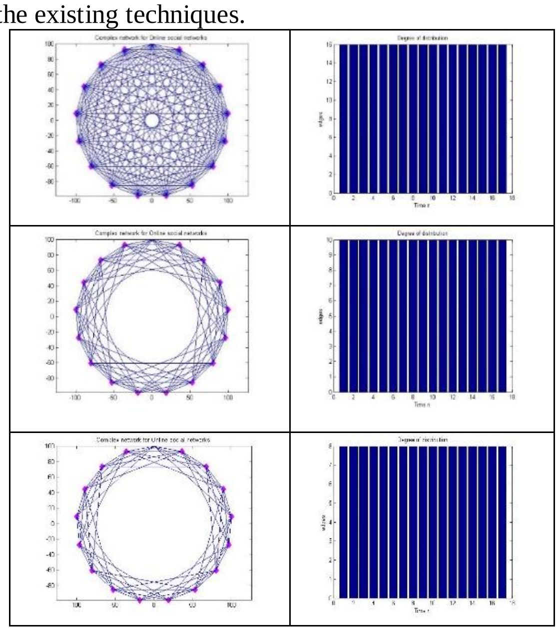

Figure 3: Showing complex networks for online social networks

In this, Figure 3 is showing complex network with higher to lesser number of edges. As number of connections got decrease, the complexity of network also decreases. Thus minimum number of connections leads to a weak complex network. Thus connectivity of these kind of network decreases.

IX. PERFORMANCE EVALUATION

This section offers the cross validation between existing and proposed techniques. Some well-known link prediction performance parameters for fuzzy logic have already been selected to prove that the performance of the proposed algorithm is very much better than the existing methods.

A. Accuracy

Table 1 shows the quantized analysis of the accuracy. The extent to which the consequence of an estimation, figuring, or determination complies with the right esteem or a standard. Exactness points to the closeness of a deliberate worth to a standard or known quality. For instance, if in lab you acquire a weight estimation of 3.2 kg for a given substance, yet the real or known weight is 10 kg, then your estimation is not exact. For this situation, your estimation is not near the known quality.

Accuracy=

$$(number\ of\ true\ positives + number\ of\ true\ negatives)/(number\ of\ true\ positives + false\ postives + false\ negatives + true\ negatives) \quad \ldots\ldots (5)$$

B. Bit Error Rate

The rate at which errors occur in the transmission of digital data the rate at which errors take place in the transmission of computerized information. In computerized transmission, the quantity of bit blunders is the quantity of number bits of an information stream over a correspondence channel that have been adjusted because of noise, obstruction, twisting or bit synchronization mistakes. The bit mistake rate (BER) is the quantity of bit blunders per unit time. The bit mistake proportion (additionally BER) is the quantity of bit blunders separated by the aggregate number of exchanged bits amid a concentrated on time interim. BER is a unit less execution measure, regularly communicated as a rate.

$$BER=\frac{1}{2}erfc\ (\sqrt{Eb/No} \quad \ldots.. (8)$$

Table 1: Shows the values of existing and proposed technique

Parameters	Existing Technique Values	Proposed Technique Values
Accuracy	86.6787	95.5596
Bit Error Rate	13.3213	4.4404

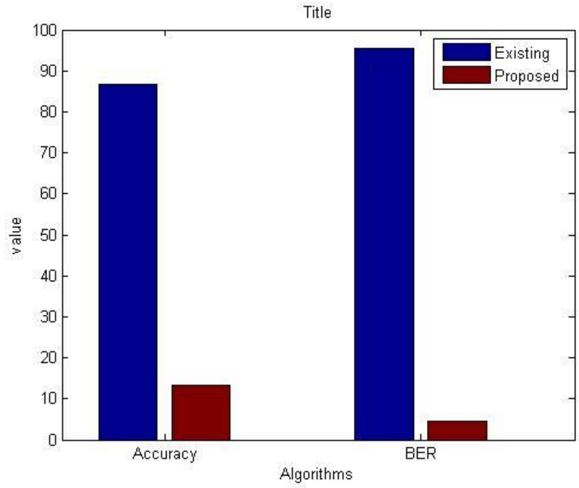

Figure: 4 showing comparison between existing and proposed technique

Figure: 4 showing that the proposed technique has less BER than existing technique and also higher accuracy than earlier technique. Hence the proposed technique is showing better results than earlier algorithms.

X. CONCLUSIONS

Main motive of this paper is to improve link predictions using fuzzy logic. Suitable fuzzy rule will be developed to evaluate the links in more promising manner. In this paper, the comparison is shown between the existing link prediction method using local directed links and the proposed link prediction method using fuzzy logic. The proposed method has proved the better results which are shown by parameters such as accuracy & bit error rate. Because the value of accuracy as high as possible & the value of bit error rate as low as possible in the proposed method.

REFERENCES

[1] Clauset A, Moore C, Newman M E J. Hierarchical structure and the prediction of missing links in networks [J]. Nature, 2008, 453(7191): 98-101.

[2] Lü L, Zhou T. Link prediction in complex networks: A survey [J]. Physica A: Statistical Mechanics and its Applications, 2011, 390(6): 1150-1170.

[3] Huang Z, Li X, Chen H. Link prediction approach to collaborativefiltering[C]//Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries. ACM, 2005: 141-142.

[4] Dong Y, Tang J, Wu S, et al. Link prediction and recommendation across heterogeneous social networks[C]//Data Mining (ICDM), 2012 IEEE 12th International Conference on. IEEE, 2012: 181-190.

[5] Han J. Mining heterogeneous information networks by exploring the power of links[C]//Discovery Science. Springer Berlin Heidelberg, 2009: 13-30.

[6] Davis D, Lichtenwalter R, Chawla N V. Multi-relational link prediction in heterogeneous information networks[C]//Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on. IEEE, 2011: 281-288.

[7] Clauset A, Moore C, Newman M E J. Hierarchical structure and the prediction of missing links in networks [J]. Nature, 2008, 453(7191): 98-101.

[8] Lü L, Zhou T. Link prediction in complex networks: A survey [J]. Physica A: Statistical Mechanics and its Applications, 2011, 390(6): 1150-1170.

[9] Huang Z, Li X, Chen H. Link prediction approach to collaborativefiltering[C]//Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries. ACM, 2005: 141-142.

[10] Dong Y, Tang J, Wu S, et al. Link prediction and recommendation across heterogeneous social networks[C]//Data Mining (ICDM), 2012 IEEE 12th International Conference on. IEEE, 2012: 181-190.

[11] Han J. Mining heterogeneous information networks by exploring the power of links[C]//Discovery Science. Springer Berlin Heidelberg, 2009: 13-30.

[12] Davis D, Lichtenwalter R, Chawla N V. Multi-relational link prediction in heterogeneous information networks[C]//Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on. IEEE, 2011: 281-288.

[13] Liben Nowell D, Kleinberg J. The link prediction problem for social networks [J]. Journal of the American society for information Science and technology, 2007, 58(7): 1019-1031.

[14] Newman M E J. Clustering and preferential attachment in growing networks [J]. Physical Review E, 2001, 64(2): 025102.

[15] Jaccard P. Etude comparative de la distribution florale dans une portion des Alpes et du Jura [M]. Impr. Corbaz, 1901.

[16] Adamic L A, ADAR E. Friends and neighbors on the web [J]. Social networks, 2003, 25(3): 211-230.

[17] Xie Y B, Zhou T, Wang B H. Scale-free networks without growth [J]. Physica A: Statistical Mechanics and its Applications, 2008, 387(7): 1683-1688.

[18] Katz L. A new status index derived from sociometric analysis [J]. Psychometrika, 1953, 18(1): 39-43.

[19] Lichtenwalter R N, Lussier J T, Chawla N V. New perspectives and methods in link prediction[C]//Proceedings of the 16th ACM

[20] SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010: 243-252.

[21] Al Hasan M, Chaoji V, Salem S, et al. Link prediction using supervised learning[C]//SDM'06: Workshop on Link Analysis, Counter-terrorism and Security. 2006.

[22] Benchettara N, Kanawati R, Rouveirol C. Supervised machine learning applied to link prediction in bipartite social networks[C]//Advances in Social Networks Analysis and Mining(ASONAM), 2010 International Conference on. IEEE, 2010: 326-330.

[23] Wang C, Satuluri V, Parthasarathy S. Local probabilistic models for link prediction[C]//Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on. IEEE, 2007: 322-331.

[24] Kashima H, Abe N. A parameterized probabilistic model of network evolution for supervised link prediction[C]//Data Mining, 2006. ICDM'06. Sixth International Conference on. IEEE, 2006: 340-349.

[25] Dunlavy D M, Kolda T G, Acar E. Temporal link prediction using matrix and tensor factorizations [J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2011, 5(2): 10.

[26] Huang Z, Li X, Chen H. Link prediction approach to collaborative filtering[C]//Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries. ACM, 2005: 141-142.

[27] Silva Soares P R, Bastos Cavalcante Prudencio R. Time Series Based Link Prediction[C]//Neural Networks (IJCNN), The 2012 International Joint Conference on. IEEE, 2012: 1-7.

[28] Soares P R S, Prudêncio R B C. Proximity measures for link prediction based on temporal events [J]. Expert Systems with Applications, 2013, 40(16): 6652-6660.

[29] Bliss C A, Frank M R, Danforth C M, et al. An Evolutionary Algorithm Approach to Link Prediction in Dynamic Social Networks [J]. arXiv preprint arXiv:1304.6257, 2013.