Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 6, June 2016, pg.244 – 250

A Survey of Various Link Prediction Algorithms in Complex Networks

Heemakshi Malhi[1], Mini Ahuja[2]

[1]Department of computer science & gndu, India

[2]Department of computer science & gndu, India

[1] mheemakshi@ymail.com; [2] miniahuja06@gmail.com

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. This paper presents a review on various link prediction algorithms. The overall objective is to evaluate various shortcomings in them. This paper ends up with suitable future directions to enhance the link prediction algorithms further

Keywords— Link Prediction, Link Prediction applications, Practices, Link Prediction techniques

I. INTRODUCTION

A network is a collection of entities which can be interconnected with links. With regards to graph principle, the entities are called vertices and the hyperlinks are called edges. In the situation of network principle, a sophisticated network is just a graph (network) with non-trivial a topological characteristic that does not occur in easy systems such as for example lattices or arbitrary graphs but usually occur in graphs of real systems. Complicated network is just a network of varied systems which are interconnected with nodes or paths. Large graphs of true to life are called complicated networks.

Link forecast in complicated systems has attracted raising interest from both bodily and pc technology communities. The formulas may be used to extract lacking data, recognize spurious conversation, and examine network changing mechanisms. Link forecast formulas focus on the contributions from bodily perspectives and approaches. Link forecast in directed network is getting more fascination among many network scientists. Weighed against predicting the living of a link prediction, deciding their path is more difficult. We propose an efficient answer called Regional Guided Route for predicting the link direction. Link forecast is worried with the issue of predicting the (future) living of hyperlinks among various nodes in a network [2]. This makes link prediction forecast different from standard information mining places which target only on objects. Link forecast allows one to understand that people are working together even though their conversation hasn't been straight. The link-prediction problem can be linked to the issue of locating lacking hyperlinks from an observed network: in numerous domains, one constructs a network of relationships centred on observable information and tries to infer extra hyperlinks that perhaps not straight obvious, but are more likely to exist. Link forecast is relevant to wide selection of places like bibliographic domain, molecular biology, offender investigations and recommender systems. Link forecast is the most essential problem that efforts to estimates the living of a link prediction involving the nodes and the characteristics of the nodes. Structure refers to the method by which nodes that prepare the network are interconnected. It reflects the data about network topology. A number of the well-known framework based forecast strategies are Common Neighbour, site rank. The link forecast problem has also been studied from the viewpoint

of the network attribute information. The attribute data refers to description of the options that come with nodes. Such data is difficult to exhibit straight in the network graph. The majority of attribute-based forecast strategies follow a machine learning strategy; that is, they choose classification-based strategies to make predictions.

II. PROBLEM DEFINITION

Let us consider a social network denoted by graph G ($V_{i,}$ $E_{j)}$ where i=1 to n and j=1 to m, n and m are total number of vertices and edges respectively. Also sum|i|<= sum|j|. Also it is assumed that all the nodes are homogenous in nature, thus have similar behaviour.

The prime objective of link prediction is to develop path of E_i's in two or more V_i's such that V_i's are distinct. Since network is dynamic in nature, therefore a link between V_i and V_j may differ every time or stay same.

III. LINK PREDICTION ALGORITHMS

A. *TRADITIONAL LINK PREDICTION*

1) Similarity Measure Based Strategies

As link prediction has its use in real world, due to which a great deal of connection predicting techniques have been proposed. The similarity based techniques are more often used. There is a great deal of strategies to discover the similarity which concentrate on the data of node qualities and system structure. It's hard to get the genuine node qualities on account of security issues, so the techniques depend on comparability of system generally utilized. Liben-Nowell et al. [7] gave the importance of link prediction. Link forecast was actualized to characterize the score (x, y). There are numerous strategies in light of resemblance of node like normal neighbor, preferential expansion. The number of neighbors that nodes x and y have in common.

$$CN(x, y) = \Gamma(x) \cap \Gamma(y) \qquad (1)$$

Where Γ (x) and Γ (y) are the neighbor sets of node x and y.

Another algorithm is also there which is based on system topology called Katz [12]. Katz measure is a weighted amount of how many paths are there in the graph that connect two nodes, where the shorter paths are given more weights [8]. In this paper, modified Katz score which only considers the length of paths up to certain threshold is used as the topological feature. It is expressed as:

$$Katz(x, y) = \sum_{l=1}^{\infty} \beta^{l} |paths^{l}_{x, y}| \qquad (2)$$

Where, $paths^{l}$ x, y is the gathering of all ways with length l joining x and y, and β is the damping component which controls the heaviness of the way. Q is the framework of the system, the closeness lattice can be appeared as:

$$Katz(x, y) = (I - \beta Q)^{-I} \qquad (3)$$

Since the computational unpredictability of Katz is more prominent than the others, so it needs to consider the overall data. Lichtenwalter et al. [13] introduced another unsupervised learning methodology PropFlow. This technique depends on ways. PropFlow resemble Rooted PageRank. In any case, it is a neighborhood method for measuring the spread and is inhumane to noise which is far from source hub.

2) Classification Model Based Strategies

Link prediction issue can be considered as a grouping based issue. The classification model is flexible than others models when compared. The components which are extracted from the comparability based method can be used in grouping models for connection forecast.

Hasan et al. [14] showed that that SVM is desirable over others .He presented that the limit set depends on the particular application. He made research on different results of varied estimations, including SVM, decision trees, K-nearest neighbor, etc.

3) Strategies Based on Probabilistic Model

The probabilistic model has higher accuracy, can find the relationship of network structure. It uses the complete network information .But cannot be used for large networks because computational complexity is limited. Kashima et al. [17] showed parameter based probabilistic model of network evolution and for predicting link among nodes he derived the learning algorithm.

Wang et al. [16] demonstrated that to show the nearby neighbourhood which contains two nodes, he introduced a neighbourhood probabilistic model for connection forecast which depended on undirected chart model. He made the neighbourhood show to diminish the expense for huge systems. It included finding the node and to decide the structure of graphical model.

B. HETEROGENEOUS LINK PREDICTION

Research has been made on homogenous networks, but only few of them shows the work on heterogeneous networks for inferring the problem in link prediction.

1) Normal Link Prediction

For handling the link prediction in heterogeneous network there are two possible ways:(1)make sure that all links are treated equally.(2)to study various types of links and to find relationship between them [4].

Davis et al. [6] showed that triad census has higher computational complexity and presented the unsupervised method MRLP. It proved that MRLP is better than other old-fashioned link prediction methods. As the unsupervised method was not flexible so Davis introduced the new multi-relational supervised method. But in this various types of links are not treated equally.

Yang et al. [23] introduced a new model which is more efficient method in link prediction. In this model Yang considered the correlation between various types of links keeping in mind the end goal to diminish the computational complexity. The model was named MRIP. The style of MRIP is reliant on two components [23]: (1) if a link i is given then the influence propagation can take place from both links i and other types of links also. (2) The probabilities of link j is dependent on the correlation between link i and link j. The impact score is as follows:

Flow (x, y, i) = score (x). β. $\frac{\text{Weight}(x,y,i)}{\text{degree}(x,i)}$ +

$$\text{score}(x)\ \beta \sum_{j \neq i}^{k} \left[\sigma(i,j).\frac{\text{weight}(x,y,i)}{\text{degree}(x,i)}\right] /(|E(x,y| - 1) \qquad (4)$$

Where x and y are hubs, β is the Katz variable, score (x) is the likelihood of a connection between the source hub and hub x, and E (x, y) −1 is the quantity of connection structures between hub x and y with the exception of sort i.

Aggarwal et al. [24] also worked on heterogeneous network for dynamic link prediction. In social networks important information cannot be obtained if privacy issues are not taken into account so, Kuo et al. [25] made a novel unsupervised stage to assess the perspective holder in a unsupervised technique to predict perspective holder in the heterogeneous system.

2) Moderator Link Prediction

Social networks have become popular in today's world. Many websites have been developed and people are having multiple numbers of accounts in those websites. A single person can have multiple accounts at the same time. Kong et al. [26] made his work on anchor links on multiple networks. When the area of the anchor link is known then the extracted the heterogeneous features from multiple networks and gave a solution called MNA (multi-anchoring network).

3) Link Prediction for New Clients

The regular users have the information but the new users do not have much information which leads to the difficulty in predicting the link of new users. Zhang et al. [27] introduced a new method SCAN-PS, which use data from the data providing network for link prediction in the perspective network. The outcomes demonstrate that this technique performs superior than other link forecast strategies for new clients.

C. TEMPORAL LINK PREDICTION

In today's world, the vast majority of interpersonal organizations are progressive and variable, which bring more difficulties for the link prediction.

1) Matrix and Tensor Decomposition Based Strategies

Daniel et al. [18] presented work on two connection forecast methodologies, system based and tensor-based, for bipartite charts. The co-creation system is truly a typical bipartite system, where in reality the types of hubs are specialists and meetings. They also demonstrated how Katz strategy is broad to bipartite charts, and used the truncated SVD to devise an adaptable technique which was then utilized to compute a "truncated" Katz score. The drawback of the tensor-based methodology is that the computational expense is high. They also showed the usefulness of natural three-dimensional structure of temporal link.

2) Time Series Model Based Strategies

Huang et al. [19] presented a hybrid prediction algorithm which composes of a set of algorithms under the time series and static graph representation model .Huang used the static graph to examine the link prediction algorithm and exhibited it utilizing time arrangement model. But problem with the model was that it become unmanageable if we take large number of links. Limitation of this model is that it do not include the temporal correlation among various links. It considers only dynamic evolution of network.

Paulo et al. [20] instead of using the time factor used topological matrices so as to remove the limitation of old-fashioned strategies. They clarified advancement of system's utilizing time series issues. But the time arrangement model has a

restricted application, on the grounds that the strategy must be connected to the circumstance where there are abundant time arrangements of occurrence of link. Firstly, matrices which are similar with time of the system are computed utilizing fundamental comparability measure calculation, for example, PA, CN, and so forth. Then, to predict the future values a set of well-known statistical forecasting models, such as moving average random walk, etc. were used. Hence calculated values will be used in supervised and unsupervised learning models font.

IV. LITERATURE SURVEY

A. Liao et al. In that paper, they add a brand new url prediction process based on Pearson correlation. The method is effective in calculating likeness based on large buy paths and is more robust to noise than conventional indices based on common neighbour .Further we found that this technique can't outperform the normal neighbour and its alternative in url prediction.

Ryan N. et al. In that paper, they include the thought about a vertex collocation page (VCP) for the target of topological url assessment and expectation. VCPs offer almost add up to data about the encompassing neighborhood configuration of stuck vertex sets. The VCP system offers another product for area powers to grasp the primary development components inside their systems and to break down url arrangement components in the privilege sociological, natural, physical.

Hisashi Kashima et al. proposed a Link Propagation as a fresh out of the box new semi-directed learning process for url forecast issues, wherever the occupation is dependably to foresee so far not known ranges of the framework plan by utilizing assistant information, for example, for instance hub similitudes. Since the proposed procedure may finish lacking zones of tensors, it is suitable to multi-social spaces, permitting individuals to manage numerous types of hyperlinks all the while.

Darcy Davis et al. include a story probabilistically weighted augmentation of the Adamic/Adar figure for heterogeneous information systems, which we use to show the conceivable awesome things about assorted confirmation, particularly in situations when homogeneous connections are to a great degree meagre. We likewise introduce some basic flaws of routine unsupervised url expectation.

Xiaojie Wang et al. In that paper, they propose a successful option called Local Directed Way to foresee url bearing. With the expansion of a supplementary ground hub to the framework, we tackle the information lessening issue in slim framework, which makes our procedure compelling and strong. In that paper, we offer consideration regarding foreseeing url bearing in guided systems and propose LDP technique. With the option of a supplementary ground hub and numbering what number of guided ways, we extend the LP way to deal with guided cases and add a semi neighborhood list to foresee the course of connection.

Evan Wei et al. In that paper, they coordinate the thin connected survey right into a structured speech and summarize the recent study operates on the hyperlink prediction task. We separate the existing url prediction techniques into three courses: the node-wise likeness based techniques tries to discover an appropriate distance rating for 2 objects; the topological design based techniques give attention to exploiting both local or worldwide techniques that may effectively describe the system; probabilistic design based techniques try to understand a tight design that may abstract the social network.

Panagiotis Symeonidis et al. In that paper, they include the thought about a vertex collocation page (VCP) for the target of topological url assessment and expectation. VCPs offer almost add up to data about the encompassing nearby plan of stuck vertex sets. The VCP technique offers another product for space powers to grasp the principle development components inside their systems and to dissect url arrangement components in the privilege sociological, natural, physical, and other connection.

Panagiotis Symeonidis et al. proposed a Link Propagation as a fresh out of the plastic new semi-administered learning process for url expectation issues, wherever the employment is dependably to anticipate up 'til now not known regions of the framework plan by utilizing assistant information, for example, for instance hub similitudes. Since the proposed procedure may finish lacking ranges of tensors, it is suitable to multi-social areas, permitting individuals to manage different types of hyperlinks all the while.

Buket Kayaet al. include a story probabilistically weighted expansion of the Adamic/Adar figure for heterogeneous information systems, which we use to exhibit the conceivable awesome things about assorted proof, particularly in situations when homogeneous connections are to a great degree scanty. We likewise display some straightforward deficiencies of traditional unsupervised url expectation.

Catherine et al. in this paper, they propose a compelling option called Local Directed Way to foresee url course. With the expansion of a supplementary ground hub to the framework, we tackle the information lessening issue in slender framework, which makes our procedure successful and powerful. In that paper, we offer consideration regarding foreseeing url bearing in guided systems and propose LDP strategy. With the option of a supplementary ground hub and tallying what number of guided ways, we extend the LP way to deal with guided occasions and add a semi nearby record to foresee the bearing of connection.

Feng Xiea et al. propose a story and unified method to cope with this lack, modelling the relational dualities applying complex numbers. In experiments with the Movie Lens dataset and the proposed Complex Illustration based Link Prediction strategy (CORLP) achieves significant efficiency in precision and coverage. Additionally, the outcome disclose

many new findings. First, efficiency is increased, when the consumer and piece degrees are taken in to account. Second, piece amount represents a more important role than the consumer amount in the ultimate recommendation.

Linyuan Lüa et al. summaries recent development about url prediction methods, focusing on the contributions from physical perspectives and approaches, like the arbitrary go based methods and the maximum likelihood methods. We also present three normal purposes: reconstruction of systems, evaluation of network developing process and classification of partially branded networks. Ultimately, we present some purposes in future.

Jingyi Reduction et al. proposed a straightforward link prediction strategy, which fully examine the community framework data of the networks. Firstly, the community framework of the many systems below various answers is extracted. Then, a straightforward frequency statistical product is placed on assess how often that a pair of nodes is divided into the exact same neighborhood below various resolutions. Ultimately, the likelihood of the lacking hyperlinks is calculated.

Xuzhen Zhua et al. propose a therefore named powerful way catalogue (EP) to leverage powerful influence of endpoints and solid connection in similarity calculation. In investigation of routes based similarity, we see that the powerful influence of endpoints and solid connection in making routes which attributes more similarity between two unconnected endpoints, resulting in a more exact url prediction.

YuXiao Zhua et al. 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.

V. COMPARISON TABLE

This section includes the various existing techniques and their benefits.

TABLE I
COMPARSION TABLE

Name of author	Title of the paper	Technique	Benefits
Kashima	Link propagation: A fast semi-supervised learning algorithm for link prediction	Link Propagation as a new semi-supervised learning	the task is to predict unknown parts of the network structure by using auxiliary information such as node similarities
Davis	Supervised methods for multi-relational link prediction	supervised learning approaches for link prediction in multi-relational networks	supervised learning approaches for relationship (link) prediction in multi-relational networks
Lichtenwalter	Vertex collocation profiles: subgraph counting for link analysis and prediction	vertex collocation profile	The same resolution that gives VCP its analytical power also enables it to perform well when used in supervised models to discriminate potential new links.
Wang	Predicting link directions using local directed path	Local Directed Path to predict link direction	The method can deal with large-scale networks in a reasonable time.
Symeonidis	Transitive node similarity for link prediction in social networks with positive and negative links	Online social networks	To detect the overall path structure in a network, being computationally prohibitive for huge-size social networks.
Kaya	Supervised link prediction in symptom networks with evolving case	link prediction method	To identify the connections between parameters, building the evolving structure of medical data network with respect to patients' ages.
Bliss	An evolutionary algorithm approach to link prediction in dynamic social networks	Covariance Matrix Adaptation Evolution Strategy	The incorporation of topological features and node attributes can improve link prediction.

VI. CONCLUSIONS

The Link prediction in directed network is attracting more interest among different networks. Compared with predicting the existence of a link to determine its direction is more difficult. It proposed efficient solution named Local directed path to predict the direction of the link. By adding an extra ground node to the network, one can solve the problem of data loss due to sparsity in networks. Majority of existing methods has considered the bivalent values only. The use of fuzzy logic is ignored by the existing researchers.

In near future, in order to improve link predicting using fuzzy logic suitable fuzzy rule will be developed to evaluate the links in more promising manner be created to assess the connections in all the more encouraging way.

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] 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.

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

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

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

[11] 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.

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

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

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] Yang Y, Chawla N, Sun Y, et al. Predicting Links in Multi-relational and Heterogeneous Networks[C]//Data Mining (ICDM), 2012 IEEE
[25] 12th International Conference on. IEEE, 2012: 755-764.
[26] Aggarwal C C, Xie Y, Yu P S. A framework for dynamic link prediction in heterogeneous networks [J]. Statistical Analysis and Data Mining, 2013.
[27] Kuo T T, Yan R, Huang Y Y, et al. Unsupervised link prediction using aggregative statistics on heterogeneous social networks[C]//Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013: 775-783.
[28] Kong X, Zhang J, Yu P S. Inferring anchor links across multiple heterogeneous social networks[C]//Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, 2013: 179-188.
[29] Zhang J, Kong X, Yu P S. Predicting Social Links for New Users across Align Heterogeneous Social Networks [J]. arXiv preprint arXiv:1310.3492, 2013.

