THOMSON REUTERS ENDNOTE

[Satpute*, 5(5): May, 2016] ISSN: 2277-9655
Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A SURVEY ON LINK PREDICTION WITH OR WITHOUT TIME AWARE FEATURE IN SOCIAL NETWORK

Prof. Khushboo Satpute*, Prof Pankaj Choudhary

[*] Department of Information Technology MITCOE college of Engineering , India
Department of Information Technology MITAOE Academy of Engineering , India

DOI: 10.5281/zenodo.50967

ABSTRACT

Online network systems have become popular in many social, biological and information system in recent year. Much research has been done in social network, which is increasingly growing and form a large complex network. Social network such as Facebook provides a platform for users to share their interest, photos & post etc with their friend. A social network can be well described by a network graph, in which node represents users and edges between nodes represent some association. In most social network links are dynamics and change over the time in network. To predict an association between two nodes in a graph, which is likely to be occur in near future and is termed as link prediction. Many approach and method have been used for predicting a link in past years, a significant interest of the methods uses local and global structure of the graph to make predictions. In this survey we are highlighting the impact of time in association (collaboration or interaction) & temporal behavior of the link between the nodes in link prediction. In this Survey we summarized link prediction without temporal features and Link prediction with time aware features, particularly the relationship between the time stamps of interactions or links and how link strength affects future link creation

KEYWORDS: Social Network, Link Prediction, Time stamps, Temporal behavior, Global features, Local features.

INTRODUCTION

Link prediction is the problem of predicting link between unconnected nodes in a social network which will be formed next in future. The link prediction problem has raise a lot of interest in recent years, and researchers proposed a variety of techniques that operate on social networks graph. Structural nature of the social network changes regularly over time as the new nodes and edges are added frequently. [2]Online social networking services can provide their users with more accurate service and more precise recommendations or suggestions, and based on measures of similarity in a social network which provides better results for link prediction. Previously work on link prediction has been mostly focused on a static network, where a partial network structure is known and the objective is to predict the hidden links. In such a static network, link occurrence is typically modeled as a one-time event and the primary interest is on the existence of the link. According to [18][19] a social network can be viewed as graph whose nodes represent people or other entities embedded in a social context, and whose edges represent interaction, collaboration or influence between nodes. Given a social network graph $G = < V, E >$ and edge $e = < u, v > \in E$ represent interactions between u and v at some time t(e). Consider a friendship network in which two users may interact in near future, if they have some similar features such as their common interest, age, friends etc.

There are broadly two main approaches for the node similarity measures

1. Local features
2. Global features

1.1 Local features

Local feature considered only the nodes attribute of the network and the global feature uses all the path in the network. [18][19] categorized the local-based similarity measures which are node-dependent such as Common Neighbours, Adamic/Adar index, Jaccard's Coefficient, etc. are used to analyzed the existence of link in network based on proximity of nodes. Some local and global methods are summarized respectively.

 ISSN: 2277-9655
Impact Factor: 3.785

Table 1:Local features based link prediction methods

Method	Description	formulation
Common Neighbor:	A number of common neighbor for node x and y have.	$CN(x,y) = \lvert\Gamma(x) \cap \Gamma(y)\rvert$
Jaccard' coefficient: [26]	Jaccard coefficient is commonly used for information retrieval. It normalizes the size of common neighbor metric.	$JC(x,y) = \frac{\lvert\Gamma(x) \cap \Gamma(y)\rvert}{\lvert\Gamma(x) \cup \Gamma(y)\rvert}$
Admic Adar:	[1] Proposed customized metric and find the similarity between two web pages. For link prediction common neighbor is considered as feature.	$AA(x,y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log \lvert\Gamma(z)\rvert}$
Preferential Attachment	Proposed the probability of vertex in the network. Considering the neighborhood size as feature value, then multiplication can be an aggregation function[4].	$PA(x,y) = \Gamma(x).\Gamma(y)$

1.2 Global features

Global features considers complete graph information, for the existence of link .They provide more efficient link prediction than local features. Global-based approaches i.e. Katz status index, RWR algorithm, Sim-Rank algorithm, etc are path based approach.

Global features based link prediction methods

Method	Description	formulation
Shortest path distance	In social network the formation of links between the two nodes is based on the path distance between them. The shorter the distance the higher the chance that it could happen.	
Katz	[13] Proposed a variant of shortest path distance. It directly sums over all the path exists between a pair of node x and y but it exponentially damped by length to count shortest paths more heavily.	$KATZ(x,y) = \sum_{l=i}^{\infty} \lvert path l_{xy}^{l}\rvert$
Sim Rank	[10] Gives generic metric which recursively defines: two objects are similar if they are referenced by similar object. The simrank score is the fixed point of the given recursive equation	$SimRank(x,y) = \begin{cases} 1 & \text{if } x = y \\ \gamma \frac{\sum_{a \in \Gamma(x)} \sum_{b \in \Gamma(y)} SimRank(a,b)}{\lvert\Gamma(x)\rvert . \lvert\Gamma(y)\rvert} & \text{otherwise} \end{cases}$

SURVEY ON LINK PREDICTION

Existing link prediction techniques predict association or link between two node in future using local and global features whereas temporal prediction is evolved as key concepts over time, suppose we have a social network graph in which the timestamp of interaction for node pairs from time t1 to tn is given, by considering this time in link prediction model we can predict the link at time tn+1. In section 2.1 survey on link prediction techniques without time aware based on basic features has being summarized and section 2.2 deals with the survey on the temporal aspect of link prediction.

2.1 Link prediction Without Time Aware Temporal Features

In [5] presented a learning framework for the heterogeneous and reciprocal link prediction problem.Also addressed the link sign prediction problem in heterogeneous and reciprocal social networks. The experiments shows the method is able to make link predictions and outperforms the use of existing link prediction techniques

 ISSN: 2277-9655
Impact Factor: 3.785

In [1][3] provided friend recommendations, also known as the link prediction problem, by traversing all paths of a bounded length, based on the “algorithmic small world hypothesis”. Experimental results show that our Friend Link algorithm outperforms other approaches in terms of effectiveness and efficiency in all data sets revealing more accurate faster friend recommendations

In [32] considered similarity-based measures, but supplemented the similarity calculations with community information. For many networks ,the inclusion of community information improves the accuracy of similarity-based link prediction methods. Results show that community information often boosts the performance of base metrics.

In [7] Adaptations of popular Link Prediction algorithms for recommending items in large-scale UGCs is shown. Three of the Link Prediction algorithms – Common Neighbors, Adamic/Adar and Katz Measure– performed consistently better than the item-based collaborative filtering technique. Neighborhood-based methods outperform all other algorithms suggesting that users are mostly interested within a small proximity of their tastes in the user-item space. Rooted Page Rank, on the other hand, was very effective in recommending items that are beyond three hops from users in the user-item graph. Results suggest that Link Prediction algorithms are a more scalable and accurate alternative to classical collaborative filtering in the context of UGCs.

In [21] propose new weighted graph proximity measures for link prediction of social networks also showed that link prediction based on graph proximity measures is suitable for open and dynamic online social networks.

In [16] Developed an efficient training algorithm to directly learn the edge strength estimation function based on Supervised Random Walks that naturally combines the information from the network structure with node and edge level attributes and achieved it by using these attributes to guide a random walk on the graph. Formulates a supervised learning task where the goal is to learn a function that assigns strengths to edges in the network such that a random walker is more likely to visit the nodes to which new links will be created in the future. Experiments show that approach outperforms state-of-the art unsupervised approaches as well as approaches that are based on feature extraction.

In [34] Introduced a two-phase method based on the bootstrap probabilistic graph. The first phase generates an implicit social network under the form of a probabilistic graph. The second phase applies probabilistic graph-based measures to produce the final prediction. Also introduced cold start link prediction as the problem of predicting the structure of a social network when the network itself is totally missing while some other information regarding the nodes is available.

In [14] developed a supervised learning approach to predict link strength from transactional information. formulating and investigating new task in social network mining: link strength prediction and compare the utility of attribute-based, topological, and transactional features evaluated the approach on public data from the Purdue Facebook network and show that we can accurately predict strong relationships also shown work on the task of predicting link existence. Results reveals overall performance, achieving 87% AUC.

In [30] an improved weighted proximity measure of link prediction is described. An assumption is made for better proximities for predicting the link that is based on both graph proximity measures and the weights of existing links in a social network. The result shows that in a dense social network this method outperforms as compare to previous approaches.

In [22] presents a supervised learning method that build the predictor which is based upon the structural attributes of the underlying network. In the research a co-authorship data is used in which researchers are nodes and link between them is collaboration. This predictor is test on this data and shows an encouraging accuracy.

In [6] a novel local probabilistic graphical method that can scale to large graphs to estimate the joint co-occurrence probability of two nodes. Topological structure of the network is used to identify the central neighbourhood set of two nodes, and then local MRF model is learned which is constrained on non-derivable frequent item sets. When used in combination with other two types of features – topological and semantic features, we find that the resulting classification performance improves.

In [35] the initial efforts to explore the connection between link prediction and graph topology is represented. They exclusively focused on the predictive value of the clustering coefficient measure. . The proposed framework consists of a cycle formation link probability model, a procedure for estimating model parameters based on the generalized clustering coefficients, and model-based link prediction generation.

In [23] a citation prediction system is made using statistical learning that extended inductive logic programming. Their system learnt link prediction patterns from queries to a relational database, including joins, selections and aggregations.

Methods that quantifying the similarity of vertices in networks are consider in [17]. A measure of similarity is proposed based on the concept that two vertices are similar if their immediate neighbors in themselves similar. Which leads to a self-consistent matrix formulation of similarity that can be evaluated iteratively using only knowledge of the adjacency matrix of the network. Test was performed on computer-generated networks their measure is particularly good at discerning similarity between vertices connected by relatively long paths, an area in which more traditional similarity measures such as cosine similarity perform poorly.

In[24] Link prediction of author\document bipartite networks by using clustering is enhanced by the author. Also clustered documents by topic and authors by research community in order to generate new entities that were used in logistic regression of features and relations. System tested the data consisting of an equal number of positive and negative cases. Experimental results shows increase in accuracy over models not using clustering by roughly four percent on average.

[18][19] The predictive power of only proximity metrics are tested which includes common neighbors, the Katz measure and variants of Page Rank. Also found some of these measures had a predictive accuracy of up to 16% (compared to a random prediction's accuracy of less than a percent). Liben-Nowell's doctoral thesis was a chapter on link prediction in social networks revealed his hypothesis, that link prediction could be performed from topology alone.

2.2. Link prediction With Time Aware Temporal Features

According to [15] link strength vary over time with the common neighbor, depends not only on the co-occurrence frequency or number of common neighbors but also on how long the common neighbors have been in contact, for finding such information time stamp of the interaction are useful. A new method time-score is introduced to design this new index which incorporates the effectiveness of common neighbors and their temporality using the following concept. Strength of link become weaker if there is no interaction between nodes for a long time with respect to the current time. Strength of link is represented as a weight in closer proximity of time if two nodes have interacted with their common neighbors than the common neighbors are more effective and the most likely to occur in the future.

In [8] considered bipartite graphs that evolve over time and consider matrix and tensor-based methods for predicting future links. We present a weight-based method for collapsing multiyear data into a single matrix. We show how the well-known Katz method for link prediction can be extended to bipartite graphs and, moreover, approximated in a scalable way using truncated singular value decomposition. Despite the high level of noise, the CP method is able to get an AUC score of 0.845, which is much better than then the "Last Period" method's score of 0.686. We also considered the accuracy in the first 1,000 values returned. The CP-based method is 100% accurate in its top 1,000 scores whereas the "Last Period" method is only 70% accurate.

[20] Provided a way to involve time information which is inspired by the fact that older events are less likely to be relevant to future links than recent ones. For example, author's interests may change over time and thus old publications might be less relevant to his current research area.

 ISSN: 2277-9655
Impact Factor: 3.785

[11] Proposed a set of new temporal distance based metrics to quantify and compare the speed (delay) of information. Distance based metrics can be applied effectively to characterize the temporal dynamics of time-varying graphs such as delay, duration and time order of contacts (interactions), compared to the metrics used in the past on static graphs and data diffusion efficiency of social networks from a local and global view. A significantly improvement in the accuracy of link prediction model is obtained by considering the time stamp of previous interaction.

In [31] showed elaborated the probabilistic model to include time awareness and used temporal feature to derive corresponding edge weight which can be used with technique such as Adamic-Adar distance based and rooted Page-Rank based technique and also proposed a testing method for evaluating performance on the basis of abilities to rank nodes from a neighborhood of the selected node.

A representation that encodes temporal data into graphs while fully retaining the temporal information while fully retaining the temporal information of the original data which is called a temporal graph is introduce in [33]. Temporal graph offer a basis for obtaining the inherently dynamic data both for describing and communicating the data itself, as well as for analysing and understanding its properties. Also presented a number of metrics that can be used to study and explore the temporal graphs.

In [28] introduced an efficient learning and inference techniques within the framework is developed by considering a restricted set of temporal relational dependencies. Also represented dynamic relational data with a two phase process, firstly temporal-relational information with kernel smoothing, and secondly moderating attribute dependencies with relational information. Also introduced temporal locality which refers to the occurrence of events in the recent past are more influential than events in the past and temporal recurrence which refers to the regular series of events between two instances which indicate a stronger underlying relationship than an event isolated in time, also they incorporated temporal information into statistical relational models and looks for patterns of temporal locality and temporal recurrence to identify stronger relationships that are more likely to exhibit correlation among the associated attribute values.

In [9] provided a tool called C-Group which allows user for viewing the addition and deletion of nodes (actors) and edges (relationships) over time, its major contribution is its focus on changing group memberships over time. By doing so, users can investigate the context of temporal group memberships for the pair. C-Group provides users with a flexible interface for defining (and redefining) groups interactively, and supports two novel visual representations of the evolving group memberships. This flexibility gives users alternate views that are appropriate for different network sizes and provides users with different insights into the grouping behavior. Temporal behavior of the network is considered so every event in the network has some time association as 1. Time period: an established interval of time, 2. Time point + duration: a starting time point and a time period, 3. Multiple time points: a starting point and an ending point.

According to [25] the approach to the problem of ink prediction given by Liben Nowell [18][19] is limited as it attempts to predict the evolution of a complex entity over time from a snapshot of the previous time step as the network grows increasingly due to this we need to examine more time step than just the previous one. The current state of the social network is given by the traditional social network analysis metric calculated from a snapshot, but for more time step (velocity) the metric is calculated using the history of the changes to a network the temporal statistics are calculated using the history of changes to a network.

In [12] presented a methods for network analysis that explicitly incorporate time and sequence, and are thus well-suited to addressing event data sets also focused on the temporal nature particular temporal link prediction problems, node ranking predicting future event co-participation of entities and rank evolution change in individual rank over time in response to participation in a series of events. Experimental results demonstrated that these approaches can be used to accurately predict organizational structure from event data and to rank likely future co-participations between entities.

ISSN: 2277-9655
Impact Factor: 3.785

In [27] described a method for modelling the relationships that change over time. The historical data are used to develop an understanding for predicting the future interactions. The model can be used to study the behaviour of individual relationships but requires adaptation to model the behaviour of a group of people.

CONCLUSION

Social interaction among the people raised popularity of social networking site over the years. Social network which allow users to share their content with others and by making friendship can be well described by a network graph, in which node represents users and edges between nodes represents their interaction. Link prediction is used in many applications such as facebook, twitter, criminal investigation, information retrieval etc. to predict the future association among the links. In this survey Link Prediction is presented in two categories. Firstly based on the feature for predicting the link uses structural properties of graph such as node degree. Some researcher used supervised machine learning technique and probabilistic model for predicting the link using these feature.Secondly categorizing the time aware based temporal feature used by the researcher for link prediction. The previous method of link prediction are less accurate as the exponential and increasing growth of social service make them dynamic and by considering the single snapshot of network is not sufficient, multiple snapshot are needed which are considered over particular time intervals, also the complexity of network require more time for computational process. Since the single article cannot be a complete review of the research done in the mentioned area because of the wide variety of link prediction application in different research field so only survey based on local, global feature and time aware based temporal feature is used for link prediction is summarized here. In this paper, the contributions of research work done in recent years, in each method were summarized and existing research challenges are also defined. It is hoped that this survey can serve as a useful guide for the researchers interested in temporal based feature attribute that can be used in the recent research work of link prediction and provides a better accurate results as compare to the previous work done. In future work, we will trying to use time aware feature for link prediction framework with some real world data of social network having dynamic nature which change quickly.In [29] future work in link prediction where the integration of Ant colony optimization improves the output.

ACKNOWLEDGEMENTS

Prof . Khushboo Satpute

She received the Masters of Technology in IT and Bachelors in CSE ,from RGTU bhopal,Madhya Pradesh, Country India. Currently she is working as Assistant Professor in MIT College of Engineering, Pune. She has Publish many research paper in data mining, and currently working on optimization techniques.

Prof Pankaj Choudhary

He is currently working as Assistant Professor in MITAOE PUNE. He received the Masters of Technology in IT ,SOIT UIT RGTU.He has published my paper in Social network,Link predictions. His Area of interest is Social network,Data mining,Database,Data Classification.

REFERENCES

[1] Adamic, L.A., & Adar, E. Friends and neighbors on the Web. Social Networks, 25(3), 211–230, 2003.
[2] Alexis Papadimitriou, Panagiotis Symeonidis, Yannis Manolopoulos. Fast and Accurate Link Prediction in Social Networking Systems In:Journal of systems and software 2012.
[3] Alexis Papadimitriou Panagiotis Symeonidis, Yannis Manolopoulos.Scalable Link Prediction in Social Networks based on Local Graph Characteristics In:Information Technology: New Generations (ITNG), 2012 Ninth International Conference PP. 738 - 743 16-18 April 2012
[4] Barabási, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. Evolution of the Social -+network of scientific collaboration. Physica A, 311(3–4), 590–614, 2002.
[5] Cai, Xiongcai; Bain, Michael; Krzywicki, Alfred Reciprocal and Heterogeneous Link Prediction in Social Networks In Advances in Knowledge Discovery and Data Mining (2012) 7302: 193-204 , January 01, 2012
[6] Chao, W., Satuluri, V., Parthasarathy, S.: Local probabilistic models for link prediction.In: Proceedings of the 7th IEEE International Conference on Data Mining. pp. 322–331 (2007)

ISSN: 2277-9655
Impact Factor: 3.785
[7] **Chiluka, N., Andrade, N.** and **Pouwelse, J.** Link Prediction Approach to Recommendations in Large-Scale User-Generated Content Systems. In Proceedings of the 33rd European Conference on Information Retrieval. April 2011
[8] Daniel M. Dunlavy and Tamara G. Kolda ,Temporal Link Prediction Using Matrix and Tensor Factorizations ACM Transactions on Knowledge Discovery from Data (TKDD) Volume 5 Issue 2,, 2011.
[9] Hyunmo KangKang H, Getoor L, Singh L (2007) Visual analysis of dynamic group membership in temporal social networks. SIGKDD Explor Newsl 9(2):13–21.
[10] Jeh G., &Widom, J. SimRank: A measure of structural-context similarity. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining New York: ACM Press. (pp. 271–279), 2002.
[11] J. Tang, M. Musolesi, C. Mascolo, and V. Latora. Temporal distance metrics for social network analysis. In Proceedings of WOSN '09, 2009.
[12] Joshua O'Madadhain, Jon Hutchins, and Padhraic Smyth. Prediction and ranking algorithms for event-based network data. SIGKDD Explor. Newsl., 7(2):23{30, 2005.
[13] Katz, L. A new status index derived from sociometric analysis. Psychometrika, 18(1), 39–43, 1953.
[14] Kahanda, I. and Neville, J. 2009. Using transactional information to predict link strength in online social networks. In Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM) 2009.
[15] L Munasinghe, R Ichise. Time Aware Index for Link Prediction in Social Networks - Data Warehousing and Knowledge Discovery, Springer, 2011
[16] Lars Backstrom, Jure Leskovec, Supervised random walks: predicting and recommending links in social networks, 4th ACM international conference on Web search and data mining, Hong Kong, China. 2011,
[17] Leicht E A, Holme P, Newman M E J. Vertex similarity in networks. Phys. Rev. E, 2005
[18] Liben-Nowell, D., Kleinberg, J., The link prediction problem for social networks. In: Proceedings 12th International Conference on Information and Knowledge Management, CIKM 2003.
[19] Liben-Nowell, David, and Kleinberg, Jon. The Link Prediction Problem for Social Networks. Journal of the American Society for Information Science and Technology, 58(7):1019-1031, 2007.
[20] Lin yuan L¨u, Tao Zhou. “Link Prediction in Complex Networks: A Survey”, in Physic a A: Statistical Mechanics and its Applications Volume 390, Issue 6, 15 March 2011, Pages 1150–1170, 2011.
[21] Mohammad Al Hasan, Mohammed J. Zaki A Survey of Link Prediction in Social Networks Social Network Data Analytics2011, pp 243-275
[22] Pavlov, M., Ichise, R.: Finding experts by link prediction in coauthorship networks. In: Proceedings of the 2nd International Workshop on Finding Experts on the Web with Semantics (2007).
[23] Popescul, A & Ungar, LH 2003, 'Structural Logistic Regression for Link Analysis', Proceedings of KDD Workshop on Multi-Relational Data Mining, 2003, viewed 12 June 2006.
[24] Popescul, A and Ungar, LH 2004, 'Cluster-based Concept Invention for Statistical Relational Learning', Proceedings of Conference Knowledge Discovery and Data Mining (KDD-2004), 22-25 August 2004.
[25] Richard Jeremy, Edwin Cooke. "Link prediction and Link detection in sequences of large social network using temporal and local metrics", 2006.
[26] Salton, G., & McGill, M.J. Introduction to modern information retrieval. New York: McGraw-Hill, 1983.
[27] Sarkar and Moore (2005) Sarkar P, Moore AW (2005) Dynamic social network analysis using latent space models. SIGKDD Explor Newsl 7(2):31–40.
[28] Sharan Umang, Jennifer Neville. Temporal-Relational Classifiers for Prediction in Evolving Domains. In: Proceedings Eighth IEEE International Conference ICDM '08. On Data Mining , 2008.
[29] Sherkat, E., Rahgozar, M., Asadpour, M. (2015). Structural link prediction based on ant colony approach in social networks. Physica A: Statistical Mechanics and its Applications, vol. 419, p. 80-94
[30] Tsuyoshi Murata and Sakiko Moriyasu, ‘Link prediction based on structural properties of online social networks’, New Generation Comput., 26(3), 245–257, (2008).
[31] Tylenda, T., Angelova, R., Bedathur, S., Towards time-aware link prediction in evolving Social networks. In: Proceedings 3rd Workshop on Social Network Mining and Analysis (SNA-KDD’ 2009), pp. 9:1–9:10, 2009.
[32] Using Community Information to Improve the Precision of Link Prediction Methods Sucheta Soundarajan ACM WWW 2012 Companion, April 16–20, 2012, Lyon, France.
[33] V.Kostakos. Temporal Graphs. Physica A,388(6):1007{1023, Mar. 2009.

ISSN: 2277-9655
Impact Factor: 3.785

[34] V. Leroy, B. Cambazoglu and F. Bonchi. Cold start link prediction. KDD, 2010.

[35] Zan Huang. Link Prediction Based on Graph Topology: The Predictive Value of Generalized Clustering Coefficient. In Workshop on Link Analysis Dynamics and Static of Large Networks (LinkKDD'06), in conjunction with KDD-2006, 2006