# A Survey on Link Prediction using Temporal Approach

Andhare Vipul[#1], Joshi Apurv[#2], Mahajan Prathamesh[#3], Panzade Amol[#4]

[#]*Department of Information Technology, MIT Academy of Engineering, Alandi (D), Pune, India*

[1]vipul.andhare@gmail.com

[2]aapurvjoshi@gmail.com

[3]prathamesh.mhj@gmail.com

[4]panzadeamol992@gmail.com

*Abstract*— **In recent year link prediction considered as main area of research in social network. A link is predicated based on structural and temporal aspect of the social network. Feature vector used for predicting the links between the nodes of social graph. Many researchers proposed feature vector based on structural data of social graph. As the evaluation in internet web technology social network grows exponentially there for predicting accurate link structural data of social graph is not sufficient some temporal data of social graph such as no. of interaction is used to proposed temporal feature vector by many author. A survey on temporal feature vector is present in this paper.**

***Keywords***— **Link prediction, feature vector, collaborative networks, structural and temporal link prediction, global and local similarity indices.**

## I. Introduction

Social networks are those collaborating networks which are formed with the interaction between autonomous entities and which leads to achieve common compatible goal and the strength of interaction among the nodes is used to predict the future links. It is possible to represent social, biological and information systems using such networks where nodes are represented by entities of system and edges represents collaboration among them [1]. Further these edges can be adorned with consideration of added information like time stamps and weights. Efforts taken on estimation of possibility of presence of link between two nodes depending on observations of known links and their features are known as link prediction.

Link prediction can be done by two types: (1) Structural, where a portion of graph is considered for observation so that remaining possible collaborations between nodes can be predicted and (2) Temporal, in which graph is observed at several time steps with the goal of predicting graph state at next time step [2]. As exploiting information of pattern of interaction is possible with different approaches of temporal link prediction, it may be considered as more significant [3].

Both of link prediction types have practical applications which made it as important task in network science however the challenges of complexities emerging in extensive real world networks can be handled more optimistically [4]. Social network users are recommended for possible interactions, applications like criminal investigations, Researchers can find other individuals who are working in same domain.

In next section we have discussed concept of feature vectors, both structural and temporal types along with established methodologies for link prediction.

## II. Feature vector

When to consider a huge dataset of social network, to predict the links which may or may not form in future, it can represented as nodes and edges of graph and various graph attributes like common neighbour, shortest path etc. for every pair of node can be calculated.

Various studies proposed feature vectors on the basis of structural data of social network but since the data of networks is increasing very rapidly these methods are considered to be inefficient to handle the dynamics of the network. Thus, some researchers proposed methods to form temporal feature vectors using data of network at time t. In the temporal methods, the current dataset is used which may vary with respect to time such as no. of interactions, degree of nodes.

## III. Feature vectors based on structural attributes of graph

In structural types where nodes ($n_i$, $n_j$) represent individuals, edges ($e_{ij}$) represent relationship among them. These node and edges can be represented as a graph. For every pair of node we can calculate various graph attribute (viz. shorted path and common neighbour) which when associated with the probability of appearing links in future, a feature vector is generated.

*1) Local Similarity Indices:* Local feature measured only the nodes attribute of the network and the global feature uses all the paths in the network. Linyuan Lu and Tao Zhou [1] categorized the locally-based similarity indices which are node-dependent such as Common Neighbours, Adamic/Adar index, Jaccard's Coefficient, etc. are used to analyse the existence of a link in a network based on proximity of nodes. As shown in table [Table I].

*2) Global Similarity Indices:* Global features consider complete graph information, for predicting the existence of a link. They provide more efficient link prediction than local features. The global-based approaches i.e. KATZ status index, RWR algorithm, Sim-Rank algorithm, etc. are path based approach. As shown in table [Table II].

## IV. Feature vectors based on temporal attributes of graph

The majority work in link prediction is performed by considering various states of the network to predict the new links, but not considering the time as a constraint. While in time aware link prediction, time is considered as one factor and other attributes are also applied resulting in construction of a time weighted network.

Time stamps are built for each link formation and computing scores at different past times. Time score is used to capture the important aspects of time stamps. Effectiveness of time score varies for different network datasets and different time measures. If the two common neighbor nodes have interacted in a specific period of time,

TABLE I
METHODS OF LOCAL SIMILARITY

| Method Name | Description | Formula |
|---|---|---|
| Common neighbor | Common neighbor is the method for which we can find the common node between predicate nodes | $s^{cn}_{xy}$= \|□(x) ∩ □(y)\| |
| Adamic – Adar[11] | This index refines the easy including of common neighbors by conveying the less-connected neighbors more weight. | $AA(x,y) = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{log\|\Gamma(z)\|}$ |
| Jaccard's coefficient | It is define as the size of fork of the neighbor Of two nodes [12]. | $JC(x,y) = \frac{\|\Gamma(x)\ \cap\ \Gamma(y)\|}{\|\Gamma(x)\ \cup\ \Gamma(y)\|}$ |

TABLE III
METHODS OF GLOBAL SIMILARITY

| **Method Name** | **Description** | **Formula** |
|---|---|---|
| Katz Index [12] | This index is based on resemble of all paths, which directly Summation over all possible paths. | $KATZ(x,y) = \sum_{l=i}^{\infty} \|path^{l}_{xy}\|$ |
| Shortest path distance[13] | In social network the arrangement of links between the two nodes is based on the path distance between them. | --- |
| Sim Rank [13] | If two objects are referenced by similar objects, then they are said to be similar. This is defined by simrank generic metrics | $SimRank(x,y) = \begin{cases} 1 & if\ x \\ \gamma \frac{\sum_{a\in\Gamma(x)}\sum_{b\in\Gamma(y)} Sim}{\|\Gamma(x)\|.\|\Gamma(} & \end{cases}$ |

then the time stamp for their link has higher value and the links related to them which are not linked have higher probability of

TABLE IIII
METHODS USING TEMPORAL ATTRIBUTES

| Method Name | Description | Formula |
|---|---|---|
| Vector Auto Regressive (VAR) Method[14] | It is **"baseline"** method, which generalizes the univariate auto-regressive (AR) model to multiple time-series. | ~Xt ≈ At−1 · ~Xt−1 + ... + At−T · ~Xt−T |
| The Lasso Granger Method | The Lasso algorithm for linear regression is an incremental. Algorithm that embodies a method of variable selection using the L1-penalty term. | w = arg min 1 n X (~x,y)∈S \|~w · ~x − y\|2 + λ\|\|~w\|\|1 |
| Time-aware Maximum Entropy[23] | This time-aware method works by allowing to forget old events, thus minimizing noise which they might initiate. | wj ← lj ( $\frac{wj}{maxk\ wk}$ ) |
| Time Score[5] | The vital aspects of time stamps of interactions and the temporality of the link strengths. **Time score** which is basically assigned a weight for future possible links. | $\sum_{c\cdot}$ $\frac{Hm\beta k}{\|t1-t2\|+1}$ S |

linking in future than the common neighbors having links with lower time stamps.

In a dynamic social collaborative network, the formation of new links indicate the interaction among nodes has started. The strength/weakness of the link is dependent on the frequency of interaction.

Time varying social networks can be used to model groups whose dynamics change over time. Individuals, represented by nodes, may enter or exit the network, while interactions, represented by links, may strengthen or weaken.

Traditional collaborative network analysis techniques such as co authorship model and social network model which often collect the data from a huge network statically. But in the temporal network, we consider the data on time basis. The weights given to the links and the ranks given to nodes change by the period of time. The feature vectors are formed temporarily and with context of time the prediction is done [14].

The user data available in social network consists of relationships which will be represented as a link between them. Here the temporal link prediction is considered on time, i.e. the dynamics of the network graph changes according to time. In other words we can say that the association or links between the node pair becomes weak when they are not interacting with the each other for a long time with respect to current time. The weight of links is given by the time they have interacted. If two users have interacted with their common neighbours, then the probability of them being linked in future is high. In the temporal link prediction, we analyse the static graphs of the network taken at different time intervals. In fig 1 we have taken a snapshot of a network at some instance. The nodes are the users of the network and the date of last interaction of the nodes is given on the edges. The node *Eva* interacted with the neighbours Alies and John, so the link between Alies and John is formed. Similarly, link between Eva and Bob is formed using the temporal behaviour of the network. In the fig 2, the snapshot of the static graph of the network after formation of links is shown.

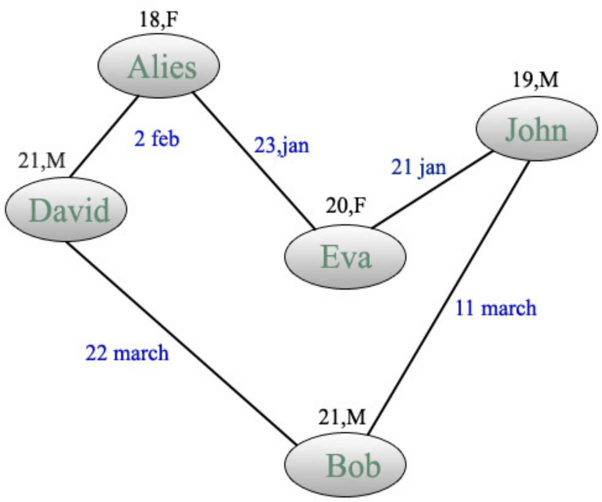


Fig. 1 Graph snapshot before link prediction.

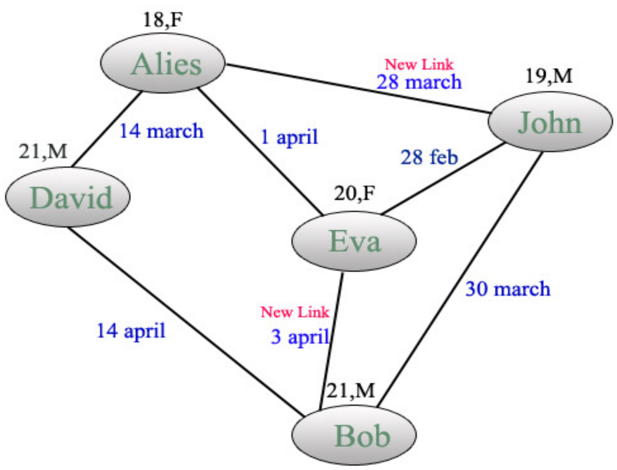


Fig. 2 Graph snapshot after link formation.

## V. Survey

Lankeshwara Munasinghe and Ryutaro Ichise proposed new technique of time aware index for link prediction in social network [5]. Time stamp of interaction is the main part for the evaluation of link. The new idea which focuses on the temporal link strength and time stamp and examine the temporal type of prediction on the two data sets is proposed.

In [6] XiongcaiCai, Michael Bain, Alfred Krzywicki, Wayne Wobcke , Yang Sok Kim, Paul Compton, and Ashesh Mahidadia consider the heterogeneous and reciprocal link prediction problem and proposed a framework to address prediction of the sign of a link in heterogeneous and reciprocal networks. Link prediction is defined as the inference of new interactions among the members of a given social network.

Sucheta Soundarajan, John E. Hopcroft presented the various methods for link prediction [7] and shown enhanced local similarity metrics often outperform their associated base metrics. Though no single metric was best for every network, they showed that one can perform cross-validation or compare metric values for existing pairs of connected and un-connected nodes to determine whether a particular method is likely to succeed on a given network. A related problem is that of determining which base local similarity metric is best for a particular network.

In [8] new temporal distance metrics to quantify and compare the speed (delay) of information diffusion processes taking into account the evolution of a network from a local and global view. Analyzing how varying the window size aspects the temporal metrics. It is inversely proportional to the temporal path length, temporal efficiency decreases as the window size increases.

In the paper [9], the authors defined a non-parametric link prediction algorithm using the sequences of graph snapshots at different time intervals. They implemented link prediction using local-sensitive hashing. . Their main contribution was as follows: 1) Nonparametric problem formulation, first ever non parametric formula for link prediction. 2) Consistency of the estimator 3) Fast implementation via LSH (locality-sensitive hashing) & 4) Empirical improvements over previous methods.

Daniel M. Dunlavy and Tamara G. Kolda considered bipartite graphs that evolve over time and consider matrix and tensor-based methods for predicting future links [10] and presented a weight-based method for collapsing multiyear data into a single matrix. They examine how the well-known Katz method for link prediction can be extended to bipartite

Jeyanthi Narasimhanand Lawrence Holder given two step solution for link prediction in dynamic networks. Feature construction approach with combination of domain and topological attributes and edge selection learner are the two phases proposed. The problem of rank deficient Time Feature matrix is not covered in article [16].

Complex network structures of evolving networks where nodes and links are added and removed over time. A precise challenge is the description and explanation of dynamics network, with a key test being the prediction of short and long term changes. In case of short-term link prediction, present methods calculate the neighborhood metrics that associate with the form of a link in the next observation period. Recent search recommended that the integration of topological features and node attributes can improve link prediction. [1] In this paper, primary focus is on the link prediction problem and local information and use similarity indices to distinguish the probability of future interactions is also focused. They considered the two main classes of similarity indices: topological-based and node attribute.

Link Prediction techniques often focus on global properties or local properties. Abir De, Niloy Ganguly, Soumen Chakrabarti have described a discriminative Link Prediction algorithm that exploits two new indicators, First, a co-clustering algorithm provides community level link density estimates, Second, links in the instant neighbourhood of the link to be predicted are not interpreted at face value[11]. The evaluation of the new predictor using five different data sets that are standard in the literature is done and new two-level learning algorithm for link prediction is described. In the future the combination of the signals with supervised personalized PageRank can be done. Another opportunity to replace item-wise or pair wise losses with list-wise losses suitably aggregated over query samples.

Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti1, Albert-László Barabási find that the similarity between two those movements strongly correlates with their closeness in the social network and explore how the predictive power hidden in such correlations can be broken to address a challenging problem: which new links will develop in a social network. Mobility procedures on your own predictive power, similar to traditional network-based measures. In this work, they follow the trajectories and communication. Patterns of approximately 6 Million users over three months, by using CDR data from an anonymous country, aiming to measure for any pair of users [12]. They improvement in link prediction tasks by thoughtfully assimilation mobility and network measures.

In the online social networks, the new links are formed among the users which have some friends in common. There are two types of approaches to predict the links viz. local based and global based. Normally the depth considered for local similarity index is maximum 2, but in [18], the authors have suggested to use depth more than 2. The method was more efficient for the network as compared to the global approach. They designed their algorithm ‘FriendLink’ for signed and unsigned networks. Although the interactions like commenting and tagging are not considered in their work.

Umang Sharan and Jennifer Neville extend relational techniques to temporally-evolving domains and outline a representational framework that is capable of modeling both temporal and relational dependencies in the data and develop efficient learning and inference techniques within the framework by considering a restricted set of temporal- relational dependencies and using parameter-tying methods to generalize across relationships and entities. By comparing the new models to the competing static relational methods on three real-world datasets and show that the temporal-relational models consistently outperform the relational models that ignore temporal information achieving significant reductions in error ranging from 15% to 70%.[9].

Andrew Arnold, Yan Liu and Naoki Abe proposed a systematic evaluation of the relative performance of a host of related methods of temporal causal modelling based on Granger causality and graphical modelling. As shown in [20] the definition of Granger Causality is a notion of causality that is highly relevant to the present context of temporal causal modelling called “Granger Causality”. The present paper proposed that upto what extent temporal information present in time series data can assist in the modelling and understanding of the causal structures between time-persistent features, rather than temporal variables.

As Ben Taskar, Ming-Fai Wong, Pieter Abbeel and Daphne Koller shown that the use of a probabilistic model over link graphs allows us to represent and exploit interesting subgraph patterns in the link graph, these methods have been applied to the problem of predicting or classifying a single link at a time [21].

## VI. Conclusions

In this survey paper, we have reviewed several current papers on link prediction, temporal graphs and time aware link prediction. The link prediction problems can be solved using various algorithms present but one thing noticed is that the problem which can be solved by one algorithm need not be solved up to that extent by other algorithm. The time aware link prediction is done on the temporary data sets and the process of link prediction is different than the static link prediction done on the network.

## Acknowledgment

We would like to express our special thanks of gratitude to our guide Mr. Pankaj Choudhary and our HOD Prof. S.M. Bhagat who gave us the golden opportunity to do this survey paper on the topic *“Link Prediction using Temporal Approach”* which also helped us in doing a lot of research and we are also thankful to those who have directly or indirectly guided us in this work.

## References

[1] Linyuan Lu and Tao Zhou. “*Link Prediction in Complex Networks: A Survey.*” Physica A: Statistical Mechanics and its Applications 390, no. 6 Elsevier Science 2010.

[2] Aditya Krishna Menon and Charles Elkan. “*Link Prediction via Matrix Factorization”*. In Machine Learning and Knowledge Discovery in Databases, pp. 437-452. Springer Berlin Heidelberg, 2011.

[3] Conrad Lee, Bobo Nick, Ulrik Brandes and Pádraig Cunningham.: “*Link Prediction with Social Vector Clock”* In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 784-792. ACM 2013.

[4] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki “*Link Prediction using Supervised Learning“*In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 784-792. ACM Rensselaer Polytechnic Institute, Troy. 2006

[5] Lankeshwara Munasinghe and Ryutaro Ichise “*Time Aware Index for Link Prediction in Social Networks”*In Data Warehousing and Knowledge Discovery, pp. 342-353. Springer Berlin Heidelberg,.2011

[6] Xiongcai Cai, Michael Bain, Alfred Krzywicki, Wayne Wobcke , Yang Sok Kim, Paul Compton, and Ashesh Mahidadia “*Reciprocal and Heterogeneous Link Prediction in Social Networks.”*In Advances in Knowledge Discovery and Data Mining, pp. 193-204. Springer Berlin Heidelberg *2012*

[7] Sucheta Soundarajan, John E. Hopcroft “*Using Community Information to Improve the Precision of Link Prediction Methods.”*In Proceedings of the 21st international conference companion on World Wide Web, pp. 607-608. ACM, 2012.

[8] V.Kostakos. “*Temporal Graphs*.” Physica A, 388 (6):1007{1023} 2009.

[9] Lars Backstrom, Jure Leskovec: “*Supervised Random Walks Predicting and Recommending Links in Social Networks”*.In Proceedings of the fourth ACM international conference on Web search and data mining, pp. 635-644. ACM,2011

[10] Daniel M. Dunlavy and Tamara G. Kolda “*Temporal Link Prediction Using Matrix and Tensor Factorizations”* ACM Transactions on Knowledge Discovery from Data (TKDD) 5, no. 2,2011

[11] Catherine A. Bliss, Morgan R. Frank, Christopher M. Danforth, Peter Sheridan Dodds:”*An Evolutionary Algorithm Approach to Link Prediction in Dynamic Social Networks” Journal of Computational Science 5.5* 2014.

[12] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti1, Albert-László Barabási:”*Human Mobility, Social Ties, and Link Prediction” Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM,* 2011.

[13] Pankaj Choudhary, Nishchol Mishra, Sanjeev Sharma “*A Survey on Link Prediction with or without Time aware feature in Social Network”*.

[14] Andrew Arnold, Yan Liu, and Naoki Abe:”*Temporal Causal Modeling with Graphical Granger Methods” Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2007.*

[15] Joshua O’Madadhain, Jon Hutchins:”*Prediction and Ranking Algorithms for Event-Based Network Data*” ACM SIGKDD Explorations Newsletter 7.2 (2005).

[16] Jeyanthi Narasimhan and Lawrence *Holder:” Feature Engineering for Supervised Link Prediction on Dynamic Social Networks*” arXiv preprint arXiv: 1410.1783 (2014).

[17] Abir De, Niloy Ganguly, Soumen Chakrabarti:” *Discriminative Link Prediction using Local Links,Node Features and Community*

*Structure" Data Mining (ICDM), 2013 IEEE 13th International Conference on. IEEE, 2013.*

[18] Alexis Papadimitriou, Panagiotis Symeonidis, Yannis Manolopoulos:"*Fast and Accurate Link Prediction in Social Networking Systems* "Journal of Systems and Software 85.9 (2012).

[19] Umang Sharan,Jennifer Neville *"Temporal-Relational Classifiers for Prediction in Evolving Domains"* Eight IEEE International Conferrence 2008.

[20] Andrew Arnold, Yan Liu, Naoki Abe:" Temporal Causal Modeling with Graphical Granger Methods" Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2007.

[21] Ben Taskar, Ming-Fai Wong, Pieter Abbeel, and Daphne Koller:"Link Prediction in Relational Data" MIT press, 2007.