Analyzing Link Prediction Techniques in Social Networks

Prof. Akash Brahmbhatt[1] Prof. Hiteh Barot[2] Prof. Shrikant Patel[3]
[1,2,3]Department of Information & Technology
[1,2,3]LDRP-ITR, KSV, Gandhinagar, India

Abstract— In recent years, the advent of social media platforms, such as Facebook, Twitter, and Weibo, has drawn the attention of researchers to the dynamic landscape of social networks. Among the myriad of intriguing issues within this domain, link prediction stands out as one of the most captivating. This paper places a significant emphasis on the prevailing research on link prediction in social networks. Over the past decade, a multitude of studies have delved into the intricacies of link prediction within social networks. The primary objective of this paper is to offer a comprehensive and in-depth review, analysis, and discussion of link prediction in the context of social networks. We aim to summarize the prevailing challenges associated with link prediction, the diverse range of methods employed, and provide a comprehensive survey of the techniques currently in use. Furthermore, this paper explores the typical applications of link prediction within the realm of social networks, shedding light on its significance and real-world implications.

Key words: Social Network, Link Prediction, Data Mining

I. INTRODUCTION

The Link prediction problem is commonly described asa task to predict how possible a link survives between anarbitrary couple of nodes. In other words link prediction isthe problem of recognizing whether a connection exists be- tween two objects or not. A social network can be visualizedas a graph, where nodes represent participants (individuals, organizations) and edges (i.e. links) correspond to interactions between actors. Predicting variations to a social network is known as link prediction problem. With the rapid developmentof internet, communication and cooperation between peoplehave become more convenient.

In recent years, online social networks such as Facebook, Twitter and Weibo, have become an important part of our dailylife and provide us platforms to exchange information with each other. Since the huge amounts of data on social networks has some obvious characteristics such as high quality, big data,semi-structure and direct reaction of real human society, many researchers from different areas or disciplines pay more and more attention to social networks.

As part of the recent surge of research on large, complex networks and their properties, a considerable amount of atten- tion has been devoted to the computational analysis of social networks structures whose nodes represent people or otherentities embedded in a social context, and whose edges rep- resent interaction, collaboration, or influence between entities. Link prediction problem can be posed as a binary classificationproblem that can be solved by employing effective features ina supervised learning framework. Thus, in every paper that I surveyed, link prediction is done using topological features of the social network i.e. existing links, and various classificationalgorithms are applied to classify future link as Yes or No. The link prediction problem has been more properly de- fined as both the recognisation of unnoticed links in a current network or as a time series problem where the task is topredict which links will be present in the network at a period t+1 given the state of a network at period t. As an example, consider a social network of co-authorship among researchers who are close in the network may be more likely to collaboratein the coming time. Link prediction is the merely sub-fieldof social network analysis, which has emphasis on edges between objects. Due to this reason, link prediction turn into more exciting than the traditional data mining areas which emphasis on objects. Link prediction can be used in many regions like recommender systems and criminal investigations.Approaches to link prediction have been anticipated based on various measures for analyzing the proximity of nodes in a network lead to the most accurate link predictions. In link prediction approach all methods assign a link weight score(x, y) two pairs of nodes x and y, based on given proximity measure and contribution graph G. A ranked list in reducing order of score (x, y) is produced. This gives the predicted new links in decreasing order of confidence. The prediction can be evaluated based on real observations on experimental data sets.

II. BACKGROUND

Data mining refers to extracting knowledge from largedata sets. The term data mining should have been more appropriately named as "Knowledge mining from data". The comprehensive goal of data mining is to extract the useful information or knowledge from the stored data. Data Miningis about resolving problems by analyzing data already present in the databases. Data mining tasks can be categorized intotwo categories descriptive and predictive. Descriptive mining tasks focus on general properties of the data in the database. Predictive mining tasks focus on the current data in order to make predictions. The purpose of a data mining determination is normally either to produce a descriptive model or a predictive model. Graphs become important increasingly in modeling composite structures like circuits, chemical compounds, images and social networks. The graph representation is basically used in pattern recognition and machine learning. Graph mining has become a key technique because of the increasing demand on the analysis of huge amounts of structured data in data mining. A Social network consists of a group of people and Links between them. These connections can be any typeof social link that makes a relationship between two people. Liben-Nowell and Kleinberg proposed one of the earliest link prediction models that works explicitly on a social network. Every vertex in the graph represents a person and an edge between two vertices represents the interaction between the persons. Social networks are popular way to mock-up theinteractions among the people in a group or community. Socialnetworks are highly vital in nature. They concentrated mostly on the performance of various graph-based similarity metrics for the link prediction task. They can grow and change as time variations and they can be visualized as graphs, in which a vertex denoted as a person in some group and link represents some form of association between the consequent persons.

III. Feature based link prediction

We can model the link prediction in social network as a observed classification task, where the each of the data point corresponds to a pair of vertices in the social network graph. Here to train the learning model, we can use the link infor- mation from the training interval ([t0, t0']).The predictions of future links in the test interval ([t1,t1']) can be made from this model. Than assume u,v V are two vertices in the graph G(V,E)and the label of the data point¡u,v¿ is y<u,v>. Mention that we can assume the interactions between u and v are symmetric, therefore the pair <u,v>and ¡v,u¿ represent the same data point, so we can represent in this term y<u,v>=y<v,u>. Applying the above labeling for a set of training data points, we can build a classification model. Then that can be predict the unknown labels of a pair of vertices <u,v>.where<u,v> / E in the graph G[t1,t1']. This is a normal binary classification task and any of the popular observed classification tools, such as neural networks, naive Bayes and support vector machines, can be used. But, the main challenge in this approach is to be choose from a set off features for the classification task.

IV. Available framework for link prediction

There are four dissimilar problems given by link prediction are shown in the figure 1 below. The most of the research papers on link prediction spotlight on problem of link existence(whether a new link between two nodes in a social network will exist in the upcoming or not). This is for the reason that the link existence problem can be easily prolonged to the othertwo problems of link load (links have different loads associatedwith them) and link cardinality (more than one link between the same couple of nodes in a social network) and the last problem of link type prediction is a little different which gives unlike roles to association between two objects. Classification whether a link exists or not can be achieved using a variation ofclassification algorithms like decision tree and support vector machine (SVM).Different structures like topological structures, content/semantic material of individual nodes can be used for analyzing the proximity of nodes in a social network.

V. Basic link prediction methods

A. Shortest-path

For ¡x, y¿ A A Eold, define, score(x, y) = (negated) length of shortest path between x and y If there are more thann pairs of nodes tied for the shortest path length, order themat random.

B. Neighborhood-based

Let (x) denote the set of neighbors of x in Gcollab Common neighbors: Based on the idea that links are formed between nodes who share many common neighbors score(x, y) = —(x) (y)—

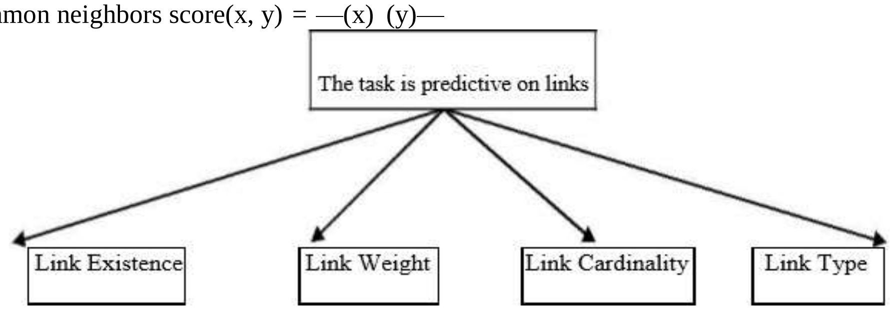

Fig. 1: Differentiation of Link Prediction Tasks

C. Katz measure

Sums over all possible paths between x and y, giving higherweight to shorter paths. x,y is the set of all length-l paths from x to y. Two variants of the Katz measure are considered (a) unweighted: paths(l) x,y = 1 if x and y have collaborated and0 otherwise (b) weighted: paths(l) x,y is the number of times that x and y have collaborated.

D. Hitting and Commute times

Consider a random walk on Gcollab which starts at x and iteratively moves to a neighbor of x chosen uniformly atrandom from (x). The Hitting Time Hx,y from x to y is the expected number of steps it takes for the RW starting at x to reach y. Score(x,y)=-Hx,y The Commute Time Cx,y = Hx,y + Hy,x is the expected number of steps to travel from x to y then back to x. score(x, y) = Cx,y = (Hx,y + Hy,x)

E. Rooted PageRank

The hitting time and commute time measures are sensitive to parts of the graph far away from x and y.The hitting time and commute time measures are sensitive to parts of the graph far away from x and y.Consider instead the random walk on Gcollab that starts at x that has a probability of of returningto x at each step. Rooted PageRank: score(x, y) = stationary distribution weight of y under this scheme

VI. LINK PREDICTION PROBLEMS

There are many works focused on solving special linkprediction problems, which can be divided into six categories: temporal link prediction, active/unactive link prediction, link prediction in bipartite networks, link prediction in heteroge- neous networks, unfollow or disappearing link prediction, and link prediction scalability.*Temporal Link Prediction*

In recent years, the research on link prediction has evolvedover various aspects. One is to consider the time in the model, which can be named as temporal link prediction [53, 52] .A social network with time can be organized as a third-order tensor, or multi-dimensional array. A tensor Z of size M N T can be dened as Z(i,j,t) =(1 if vertex i links to vertex j at timet ,0 otherwise.It can answer specic questions such as "Whois most likely to publish at ICDM next year". Given social network for times 1 through T, it needs to predict the links at time T + 1.Richard et al. investigated the links of the graph and its topological features that have been evolving over time may also be useful to predict future links. Their work lies in the observation that a few graph features that can capture the dynamics of the graph evolution and provide information for predicting future links. The main idea is to learn over timethe evolution of well-chosen local features (at the level of the vertices) of the graph, and then uses the predicted value of these features on the next time period to discover the missing links. Jahanbakhsh et al. exploited time-spatial properties of contact graphs as well as the popularity and social informationof mobile nodes to propose a method for reconstructing the missing parts of contact graphs in mobile social networks.

A. Link Prediction in Heterogeneous Networks

Most of the existing link prediction works focus on homo- geneous networks, in which only one type of nodes or links ex-ists. However, many social networks contain dierent link typesand dierent kinds of nodes, which may have dierent typologiesor link formation mechanisms and inuence each other.Thekey component of the MRLP is an appropriate weighting scheme for dierent edge type combinations. The weights are determined by counting the occurrence of each unique triad census with three nodes. The triad census also provides the probability of each structure, which further translates to the probability that a partial triad is closed by respective edge type.To predict the relationship building time between twoobjects, namely, whether or when a relationship between two objects will be built, target relation and topological featuresare encoded in a meta-path, then a generalized linear model based supervised framework is used to model the relationship building time.

B. Link Prediction with Active and Unactive Links

Munasinghe and Ichise gave an assumption that if a node pair interacts recently, then the link between them becomes active. The time stamp of the last interaction is a vital information in deciding the activeness of a link. Hence, the most recent time stamps of the interactions between nodes is used in link prediction computations. T Flow based on the PropFlow metric is proposed.The experiments on several real world datasets show that younger links are more informative than older ones in predicting the formation of new links. Sinceolder links become less useful, it might be appropriate to remove them when studying network evolution.

C. Link Prediction in Bipartite Networks

Many social networks are bipartite networks such as the user-product networks in e-commerce area. However, the link prediction problem is usually dened on unipartite graphs, where common link prediction methods make several assump-tions: (1) triangle closing: new edges tend to form triangles; (2)clustering: nodes tend to form well-connected clusters in the graph.It shows that both time attenuation and diversion delay play key roles in link prediction in a user-object network.

D. Link Prediction for Unfollow or Disappearing Links

The formation and dissolution of link are two fundamental processes of link change and evolution in dynamic networks. Links in social networks could be appeared or disappeared. Forexample, user A in Twitter breaks the relationship with anotheruser B. In social networks, we call this behavior "unfollow". To the best of our knowledge, numerous eorts have been madein studying link formation for predicting new links in future, but only few attentions are paid to link dissolution, namely, predicting the disappearing link in future.

E. Link Prediction Scalability

The scalability and eectiveness are both important for massive real world social networks. Sarkar et al. proposeda nonparametric link prediction for dynamic networks in which their model can accommodate regions with very dierentevolution proles, otherwise impossible by the link prediction metric or heuristic. It also enables learning based on bothtopological as well as other externally available features. Theyalso adapted the locality sensitive hashing algorithm to solve the scalability for link prediction in large networks and long time sequences.

VII. LINK PREDICTION APPLICATIONS

In social networks, link prediction can be used for various applications; here we will address some typical applications, such as recommendation in social networks, network comple- tion, and social ties prediction.

A. Recommendation in Social Network

Recommending partners, friends, followees and followers is a typical application for link prediction. Wu et al. proposed anovel interactive learning framework to formulate the problemof recommending patent partners into a factor graph model. Wu and Dong et al. also developed a transfer-based factor graph model that combines them with network structure information for link recommendation across heterogeneous social networks.

B. Reciprocal Relationship Prediction

In social networks, a two-way (also called reciprocal) relationship, usually developed from a one-way (parasocial) relationship, represents a more trustful relationship between people. Understanding the formation of two-way relationships can provide us insights into the micro-level dynamics of the social network, such as the underlying community structure and users's influence on each other. Hopcroft et al. studied the extent to which the formation of a two-way relationship canbe predicted in a dynamic social network.

C. Network Completion Problem

Usually, the collected social network data is incomplete with nodes and edges missing. Since that only a part of the network can be observed or collected, it needs to infer the unobserved part of the network. This is the network completionproblem, wherein, given a network with missing nodes and edges, one has to complete the missing part. Tang et al. used the Expectation Maximization (EM) framework to model the social network completion problem, where the observed partof the network is used to t a model of network structure, and then estimates the missing part of the network using the model,re-estimate the parameters and so on.

D. Finding Experts and Collaborations in Academic Social Network

Academic social networks contain massive amounts of experts in various disciplines and it is difficult for the individual researcher to decide which experts will match his own expertise best. Pavlov and Ichise propose a method for building link predictors in academic social networks, where nodes can represent researchers and links represent collaborations. It usesa supervised learning method for building link predictors fromstructural attributes of the underlying network.

E. Social Tie Prediction

When a social network dynamically changes, the social tieswould change over time. The social-tie strengths are dierent one another even though they are in the same group. Zhangand Dantu investigated the evolution of person-to-person social relationships, quantify and predict social tie strengths based oncall-detail records of mobile phones.

VIII. CONCLUSION

The link prediction is not a new problem in link mining and analysis. New link prediction techniques, problems and applications are emerging quickly in recent years.it is based on well-organized social network. Link prediction is concerned with the problem of predicting the (future) existence of links among nodes in a social network or graph. . This paper attempts to systematically summarize all typical works on the link prediction in social networks. Link prediction problems is proposed. Link prediction problems and applications are also presented.. Link prediction techniques can provide a very efficient way for discovering useful knowledge from existing information. Most of the link prediction information is provided in this paper.

REFERENCES

[1] HERMAN YULIANSYAH, ZULAIHA ALI OTHMAN, et al." Taxonomy of Link Prediction for Social Network Analysis: A Review", IEEE Access, VOLUME 8, 2020, pp. 183470- 183487
[2] Ajay Kumar Singh Kushwah1 and Amit Kumar Manjhvar2,"International Journal of Grid and Distributed Computing", Vol. 9, No. 2 (2016), pp.43-50
[3] WANG Peng1,3 , XU BaoWen1,2,3 , WU YuRong1 ZHOU XiaoYu1 January 2015, Vol. 58 011101:1–011101:38 doi: 0.1007/s11432-014- 5237-y
[4] David Liben-Nowell and John Kleingerg,January 8, 2004
[5] M.S. Graduate Student,State University of New York ,May 15, 2008
[6] [6] Qiu B, Ivanova K, Yen J, et al. Behavior evolution and event-driven growth dynamics in social networks. In: Proceedings of IEEE Second International Conference on Social Computing (SocialCom), Boston, USA, 2010. 217– 224
[7] D. Sharma, U. Sharma, and Sunil Kumar Khatri, "An Experimental Comparison of the Link Prediction Techniques in Social Networks ", vol. 4.no. 1, February (2014).
[8] Lada A. Adamic and Eytan Adar. Friends and neighbors on the web.Social Networks, 25(3):211–230, July 2003
[9] Local Probabilistic Models for Link Prediction – C. Wang, V. Satuluri –ICDM 2007
[10] Link Prediction using Supervised Learning - M. Hasan, V. Chaoji, S.Salem, M. Zaki
[11] Discovering Missing Links in Wikipedia - Sisay Fissaha, Adafre Maarten de Rijke – LinkKDD 2005
[12] Label and Link Prediction in Relational Data – B. Taskar, P. Abbeel, M. Wong, D. Koller
[13] Tylenda T, Angelova R, Bedathur S. Towards time-aware link prediction in evolving social networks. In: Proceedings of the Third Workshop on Social Network Mining and Analysis, Paris, France, 2009.

[14] H. J. Patel, R. Prajapati, Prof. M.Panchal, Dr. M. J. Patel,"A Survey of Graph Pattern Mining Algorithm and Techniques" vol. 2, Issue 1, ISSN 2319 – 4847, January (2013).
[15] N.. Jain, V. Srivastava," Data Mining Techniques" vol. 02 Issue: 11, Nov (2013).