LINK PREDICTION IN SOCIAL NETWORKS

Arockia Panimalar.S[1], Sruthi.K[2], Rakshitha.K.R[3], Soundarya.S[4]

[1]*Assistant Professor, Department of BCA & M.Sc SS, Sri Krishna Arts and Science College, Tamilnadu*
[2,3,4] *III BCA Students, Department of BCA & M.Sc SS, Sri Krishna Arts and Science College, Tamilnadu*

Abstract: *It is important to analyze the evolution of social networks for efficient working of the networks. Link prediction is mainly used to predict missing links, new or dismissing links in current networks of the social networks. In this paper, a methodical category for link prediction techniques and problems is presented. Many works on link prediction have been done in early days. The objective of this paper is to extensively analyze and discuss the link prediction content in social networks. Achievements and future challenges are also discussed.*

Keywords: Social Network, Link Prediction, Dynamic Network, Similarity Metric, Learning Model.

1. INTRODUCTION

Stanley Milgram, the one to describe modern social network and the procedures how they are connected typically. In the social network, a single person is considered as the node and the link node is considered as "connections" that establishes the relationship between people.

There are a lot of examples social network services such as:

- MySpac
- Facebook;
- LiveJournal etc.

Why are we more interested in Facebook?

Facebook is doubling in size once every six months by 100,000 users per day. More than 60 percent of users are outside of college age. Mark Zuckerberg accredited the power of Facebook to the "social visual representation" the network of connections and relationships between people on the service.

Why is it hard to predict links in social networks?

Social networks are eminently dynamic, dispersed and have combined structure, this is the reason its outcome is difficult to predict. Moreover, because of early-stage fame, it is possible to estimate the popularity at the more recent stage. The task of detailed prediction the presence of links or connections in a concern is on the one hand important task and on the other hand is highly imposing. So, can we anticipate larger on the social network? Since the links from the network, their subsistence and description reflect social actions of personage, the research on them can be beneficial at the quantitative and qualitative estimation of human relationships in the age of information humankind. Moreover, link prediction is associable to a wide variety of areas like bibliographic authority, molecular biology, criminal investigations and recommender systems.

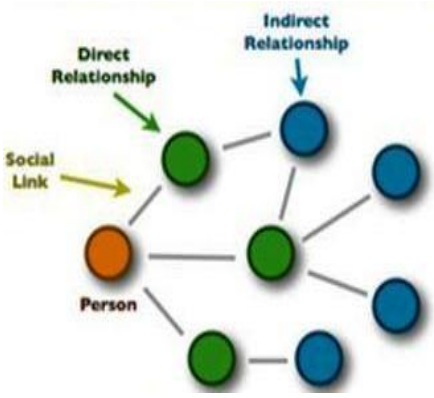

Fig 1: Social Graph Representation

2. RELATED STUDIES

In the study of sociology, criminology, and intelligence the Link analysis and social network analysis (SNA) tasks are considered as because of using graph-theoretic representations. In one side, most of the current data-mining methods assume that the data is from a single relational table.

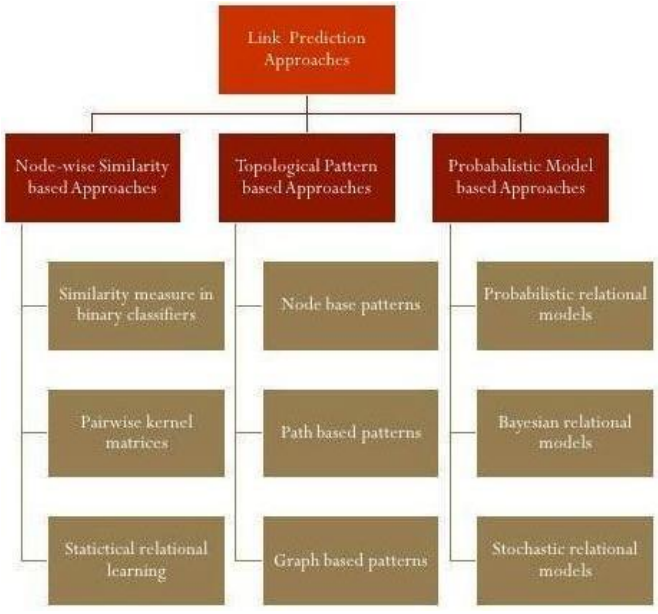

Fig 2: Different Approaches to Link Prediction Task

The most widely examined methods for convincing relational patterns are those in inductive logic programming (ILP). ILP concerns the initiation of Horn-clause rules in first-order logic from data in first-order logic. There are several definite approaches for dealing with link prediction task in social networks: supervised vs. unsupervised are established on learning a binary classifier that will anticipate whether a link exists between a given pair of nodes or not.

To identify whether a link exists or not can be carried out using various supervised learning or classification algorithms like the decision tree, K-nearest neighbors or support vector machines (SVM). Now let the details on link prediction task and how they are categorized can be considered where we are given two existences in a network, and we should assign if there any link exists.

Categorization based on features of entities: entity attributes and relational features (indirect relations). The difference between the Directed Graphical Models and Undirected Graphical Models such as Bayesian networks and PRMs admit easily captures the dependence of link individuality on attributes and constraint probabilistic dependency graph be a directed acyclic.

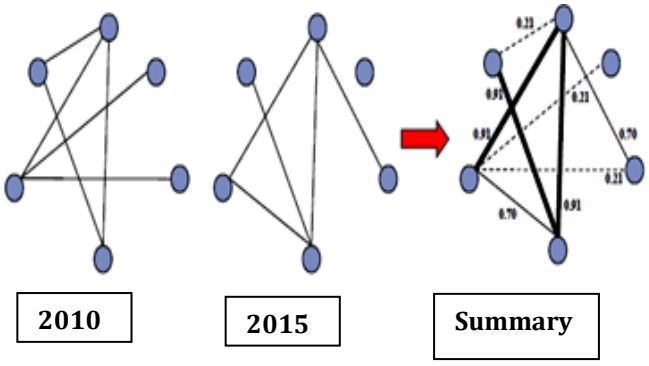

Fig 3: Application of Probabilistic Models to Link Prediction Task

In order to resolve the main issue let's consider existed approaches that can be helpful:

•Exploring relational structure, clustering.
•Using links to predict classes/attributes of existence predicting link types based on known individual classes.
•Predicting links based on location in high-dimensional space.
•Ranking potential links using a single graph-based feature.

In addition, link discovery problem was deeply researched at Kansas State University:

-Considered the problems of predicting, distinguishing annotating friends' relations in social networks by exercising feature constructing approach.

-Genetic programming-based symbolic regression method was proposed which is used for the construction of the relational features of link analysis task in the social network domain.

3. THEORETICAL BACKGROUND

A. Centralization

Centralization is the difference between numbers of links for each node divided by the maximum possible sum of differences.

B. Adjacency

Adjacency in the degree of an individual is near all other individuals in a network (directly/indirectly).

C. Closeness

Closeness is the contrary of the sum of the shortest distances between each distinct and every other person in the network.

D. Path Length Measure

Path Length Measure represents the interval between pairs of nodes in the network.

E. Clustering Coefficient

A degree of the likelihood that two associates of a node are associated them.

F. Cohesion

It is the degree to which actors are connected straightly to each other by cohesive bonds.

G. Degree

The count of the number of ties of a single actor with other actors in the network is the degree ("geodesic distance").

H. Local Bridge

An edge is a local bridge if its endpoints share no ordinary neighbors. "Degree/Proximity/Status", are all its measures.

I. Radiality

The Degree of an individual's network connection to the network and provides required information and authority.

J. Structural Cohesion

The minimum number of joiners who are eliminated from a category would disconnect the group.

K. Structural Equivalence

Refers to the nodes have a common set of affiliation to other nodes.

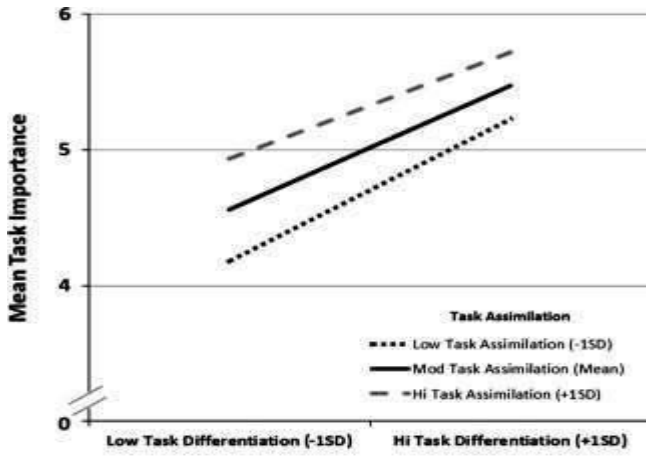

Fig 4: Differentiation of Link Prediction Tasks

4. LINK PREDICTION PROBLEM

In social networks, link prediction problem and the network adjustment prediction problem are considered as the same issue. There are many types of familiar link prediction methods that are based on:

- Node information (redistribution point)
- Structural information (visual perception)

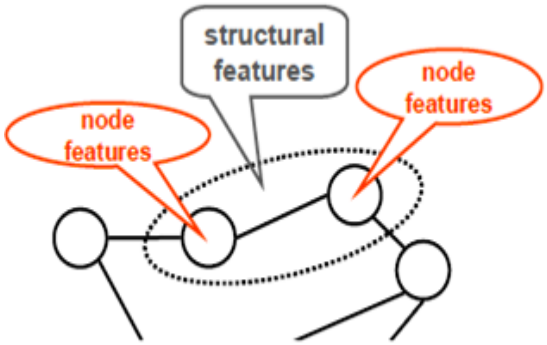

Fig 5: Difference between Nodes Attributed Structural Features

Using visual representation structure the social network connections among data are represented. In the graphical representation, each node symbolizes a data and a link represents a relation between two data. In addition, each node can also have an affiliated vector-structured data in the network model. In contrasting to machine learning standard tasks settings: data is represented as tables, where rows symbolize observations and columns represent features/attributes.

The link prediction problem is usually defined as:

Given a set of data instances $V = vii=1n$,

which is adapted in the form of a social network $G = (V,E)$, where E is the set of observed links, then the task to calculate how likely an unobserved link $eij \notin E$ exists between an inconsistent pair of nodes vi,vj in the data network. The mathematical description of a problem

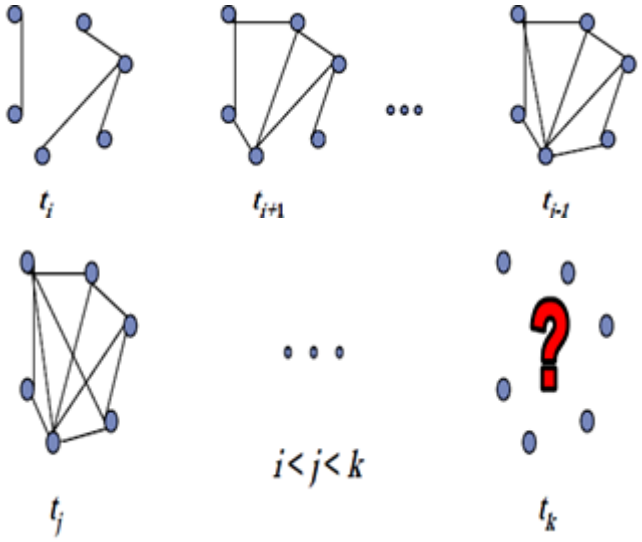

Fig 6: Visual Representation of Link Prediction Task

All the procedures assign an associate weight $score(x,y)$ to pairs of nodes x,y, based on the input graph, and then exhibit a ranked list in descending order of $score(x,y)$. They can be viewed as computing a range of closeness or "similarity" between nodes x and y. contradicted length of the shortest path between x and y. Every node that shares one neighbor will be linked.

5. CONCLUSION

The links that are not noticed are identified in social networks using link prediction concept is explained in this paper along with that the concept of the social networks is defined in the graphical representation as well as in mathematical representation with deep categorization and approaches for predicting unobserved links are described. J48, OneR, IB1, Logistic, NaiveBayes are the categorized task of application in the metabolism of link prediction task. Mining tasks in network-structured data were analyzed. The test was planned based on crawling technique with the application of free open-source SQL full-text search engine Sphinx. The proposed training corpus is the Facebook social network website. The database for collective data was represented in SQL format. The visualization tools for social networks graph representation were explained. Social networks operate in the different level and hence it plays a vital role in deciding the way the problems should be fixed and o run the organization efficiently and successfully.

6. REFERENCES

[1] L.A. Adamic, E. Adar: Friends and Neighbors on the Web. Social Networks, 25 (3) 2003, 211-230

[2] Hanneman, R. and Riddle, M. (2005). Introduction to social network methods. Riverside, CA: University of California, Riverside

[3] Samuel R. Smith "Postmodernism is dead, now what? Distributed culture and the rise of the network age"

[4] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. Physic A: Statistical Mechanics and its Applications, 311(3-4):590–614, August 2002

[5] J. Golbeck: Computing and Applying Trust in Web-Based Social Networks. Ph.D. Thesis, University of Maryland, College Park