JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 7, ISSUE 02, 2020

Prediction of Relations Among Business Networks: A Study

Tanmoy Samanta
Dept. of Economics, Tamralipta Mahavidyalaya
West Bengal-721636, India
tanmoysamanta1975@gmail.com

Abstract

In network theory, the problem of predicting relations between two entities in a network is linked prediction. Examples of communication predictions are the prediction of user relationships on a social network, forecast in a citation network of co-authorship connections, and prognosis of gene-protein interactions on a biological network. The prediction of a link may also have a temporary aspect where the purpose is to preview contacts at time t +1 in view of a snapshot of the links. The prediction of ties is commonly used. Links are typically a subtask to suggest products for users during e-commerce. It can be used for record duplication when curating citation databases. This study is a review of link predictions in social networks, especially in business networks.

Keywords: Link prediction, Social networks, Business networks.

I. Introduction: Social networks have become famous through the development of computer technology and the popularity of computer networks. The cost of generating and disseminating information has fallen significantly due to the convenience of the internet. As a result, the amount of data has exponentially increased, and the internet displays hundreds of millions of websites, emails, blogs, etc., every day. Information can be described as a collection of interrelated objects, and their collection constitutes an enormous network. The network node is the object, and the network edge is the link or relation between the object and the object (Lamari and Slaoui, 2017). Network structure can be described in several complex systems, such as social information and biological systems. The network structure can also describe social networks. The design of the network consists of nodes and borders. The nodes of the network represent entities (i.e., persons) of the system in complex systems such as social systems, and the edges of the network represent interactions among entities. Sun et al. (2017) felt that the 'link' was generally viewed and universal as a relationship. Many connections are actually hidden, and people don't know about these connections so that people will be interested in these unknown connections. The link prediction estimates the likelihood of two unrelated nodes with links based on information on network structure and other links (Sun et al., 2017). The link forecast essentially refers to a link's ranking.

High application value is the link prediction. First, link predictions can be used as "friendly suggestions" as part of a social network. In other words, the friendship that is not reflected in the social network can be predicted according to the existing company of the web so that the social circles can automatically spread, thus significantly reducing the manual time used for tapping potential friends. People need systems to make possible friends expand social processes that also create the task of predicting links. The link was ubiquitous, Jie and Tao (2018) suggested. These connections typically indicate the significance, classification or category of the data (Jie and Tao, 2018). Not every link is usually known. Therefore it is vital to predict the relation between objects. In the future, it is expected over time whether two items will establish a connecting relationship.

A new entry point and research method for social network analysis are provided by the development of link analyses and prediction, and it supports combinations and further development of sociology and computers. Currently, the network structure link analysis method has gradually become a significant field of research for social network analysis, and connection prediction is one of the most critical technologies.

The Social Network is a multi-information provider, including attributes for users, network topology and user content. These data reflect, to a certain extent, users' characteristics or relationships. All forms of social networking studies generally start from analysing these data, and research into predictive links is no exception. The multi-network information carrier usually considers a user node to be the link prediction unit. Measurement of node similarity in connective prediction plays an important role. The study of node similarity includes, in particular, a survey of local topology and attribute resemblance. In previous studies on node similarity, the research on attribute similarity tends to be given more attention. But at the moment, the investigation of network topology similarities is increasingly taken into account. Research on the topology of networks in the fields of sociology and artificial intelligence has earlier been carried out. The original research objective is to perform functions such as information recovery and analysis of network structures. Later, many industries of computer science, including the link prediction, introduced the method of research for measurement of similarity. This study analyses the tools and techniques of link prediction preferable to business networks.

II. Literature Review: Contributions and history

A network is said to be two network mode if two node sets are present, and there are no two nodes connected in the same set (Borgatti 1997). In our case, the referral network is known to be a bipartite graph, since we have general practitioners (GP's) and specialists (SP's). A network that includes nodes of the same set is a one-mode network. A social network of students is an example of the one-mode network. In addition to the other benefits that are discussed in detail in the next section, our connection prediction approach is different from those mentioned in the literature. It can be applied to both network modes. We, therefore, define our contributions to address the two problems mentioned in the following paragraphs.

Network analysis and prediction is an important problem field discussed by various disciplines such as informatics and computer biology (Basso et al. 2005). Often its sensitivity and volatility attribute to the dynamic nature of data from gene expression: ties are delicate in that a small disruption in expression values might be the key model or the start of any trend; and because of noise and uncertainty, it is volatile. The difficulty in interpreting gene-gene interactions with different periods is a crucial bottleneck in mining gene expression networks in current research.

They propose a link prediction model in this article to mitigate the bottleneck above and achieve the objective of predicting network interactions between genes. The authors used the yeast time series (DeRisi et al. 1997) microarray data and divided the whole experiment into two parts, based on time intervals and implied regulation networks. In general, our proposed approach offers a social-network model that makes comparative gene-gene interaction analyses easier and efficient network pattern visualisation easier. Predictions are seen by attaching and not-connecting genes to networks.

"Data mining is becoming increasingly popular in healthcare, if not more essential," as quoted in Koh and Tan (2005). Data mining applications to be applied to health care data are various and different. "Prediction models probably include the most common and important applications in data mining" (Koh and Tan 2005). In our article, we concentrate on developing connection prediction techniques which can solve one of the most important problems in a critically important medical system, called a medical referral system.

The method of medical references is the reference of patients to specialist doctors. Canada's health system inspired the model. If a patient experiences a condition, he or she visits a GP. If the GP cannot diagnose a patient's problem, then the GP refers the patient to the specialist SP, known as the patient's type of disease specialist. As the networks of health care supplies expand, it gets harder to find the right specialist and exchange knowledge for patients between suppliers (Clancy et al. 1996; Lee et al. 1983). The explanation of why the referral process has taken a long time to locate the relevant expert is unacceptably lagging behind.

Moreover, often patients are referred to specialists who are busy and cannot thus consider additional references for a particular period. For the availability of their SPs, GPs will not get swift answers. It is even worse to wait weeks and obtain a solution that no additional referrals can be accepted.

In the research community, the model of the social network was given considerable attention. It has succeeded in evaluating numerous applications, such as terrorist networks (Ressler 2006), online society (Crandall et al. 2008) and communities that call on the network (Kianmehr and Alhajj 2009), among other items. They claim that the model is sufficiently strong and can be used for the study of any domain described in a network context. Other research focused on topics related to prediction. The work defined by Liu and Lu (2010), for example, implemented a model of a link-prediction based on the node similarity. This is important in applications where nodes like gender, age, etc. are similar. They suggested two random walk-based similarity indices. In friendship applications where a person can be suggested to be a friend of someone else based on mutual aims, other works apply the relation prediction. The study described in Karahalios and Gilbert (2009) used techniques for forecasting the strength of social media linkages.

Most of the earlier works concentrated to the best of our ability on forecasting future linkages (optimistic link prediction), and no research are mentioned in the literature to predict future links (negative link forecast), but it is imperative to decide how connections are lost. In the framework of the gene expression networks, and in the health care field considered to be two of the essential areas to humans, nobody has yet addressed the issue of connective prediction. Hasan et al. (2006) have done a previous analysis of link prediction as a supervised study activity in the author's field to our research. They also described a variety of main features. With these features, they were capable of proposing a model of classification to predict the potential collaboration of several authors.

In the link prediction called internal links and a weighted projected to predict links in bi-party charts, Allali et al. (2011) introduced a new approach. In specific social networks, this method is useful for predicting connections that will occur in the future, but which connection will not be lost. Our approach to our new relationship prediction model is compared, and more details are given about how this approach works later. The literature describes several methods existing in the mining of gene expression data. The fuzzy factor for financial aspects has been introduced by Lee et al. (2006) and Gao et al. (2009).

A survey on relation forecasting was written by Pobiedina. The forecast method was based on graph approximation. In order to research the issue of relation prediction in unknown social networks, belief function theory has been used. First of all, a new model of the graphical social network was proposed, containing the ambiguity in the relation structure. Then a new approach was proposed for predicting future relations by knowledge fusion of adjacent nodes, based on the resources of the belief feature theory, coupled with the facts and decision of different sources (Jalili et al., 2017). In addition to the weighted variable network in addition to the USAir network, Bai et al. proposed similarity measures by combining resource allocation and local path indexes and six unweighted networks showed that its accuracy is more significant than local path 1; by designing the required weighted versions and tests on a number of weighted networks (Kaur and Singh, 2016). A relation prediction in a binary network was proposed by Gao et al. Initially, the binary network is mapped to a different network called the projection map, and a project map determines the definition of candidate node pairs (CNP).

Sett et al. (2018) assumed that machine learning was an artificial intelligence science. The main object of research was artificial intelligence, immensely how to improve the output in experiential learning for specific algorithms (Sett et al., 2018; Bell et al. 2018). The analysis of computer algorithms can automatically be improved by experience (Bell et al.), was a clarification of some concepts in data theory (Tang et al., 2016).

Social networks were analysed through the mining of objects, connections or social networks. In most instances, the relation itself was the key. However, not every relationship was observed at some point (Hofman et al., 2017). The social network interaction was an essential element of data collection and a subset of social network analysis (Li et al., 2017).

The RSM index of link prediction is a new kind of method to predict the relationship in uncertain environments. The authors, Mahapatra et al. (2020), claimed that the relationship predictions are based on the positive attitude of the familiar neighbours. If the common neighbours are not friendly, their prediction score will be meager.

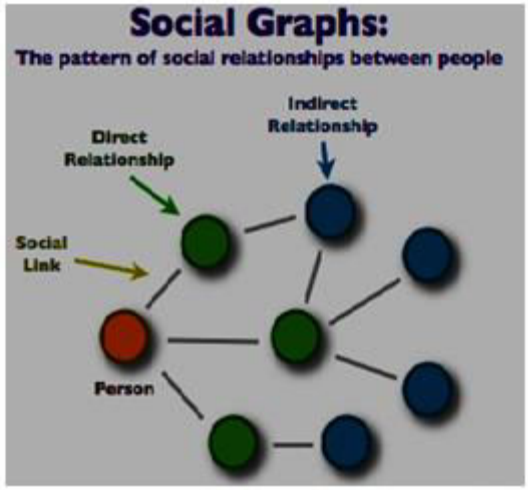

Figure 1: A social graph

III. Link Predictions in Business Networks:

A corporate network is a dynamic network of businesses working together to achieve such goals. Based on their market position, the business networks follow these goals, which are strategic and operational. There are two types of business networks — alliances and aggregations of companies — that enable SMEs to become more competitive and creative.

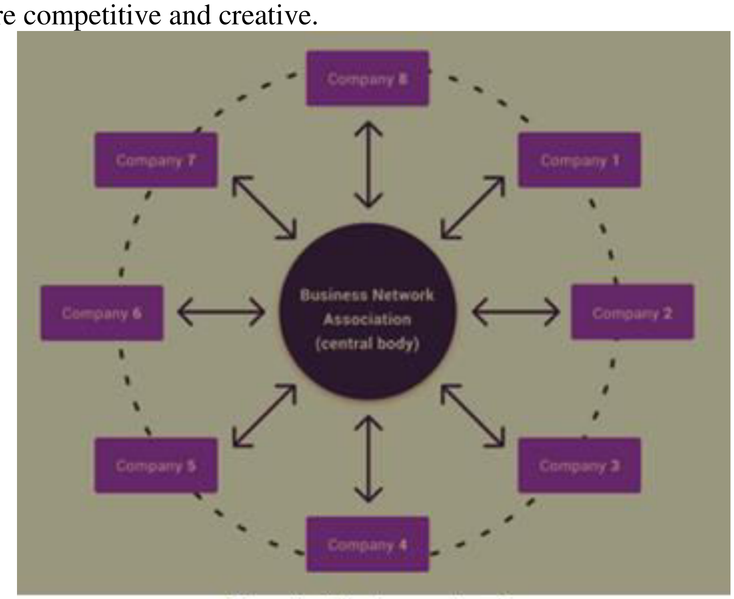

Figure 2: A business network

Thus, relationship predictions between two business workers are not common compared to link prediction in social networks. The familiar neighbours /familiar customers are not effective in business networks. It can be said that friendly neighbours within self-company employees are enough to link predictions. When the links are predicted between two persons from two separate companies, the degree/weight of the person matters. The following picture demonstrates the concept. In Figure 1, the area of circles indicates the weightage of the company and/ or weightage of the persons. There are three circles. Two circles are big in size. That means that these companies/persons have high degrees. Thus, their relationship predictions will have a high score (see Figure 4).

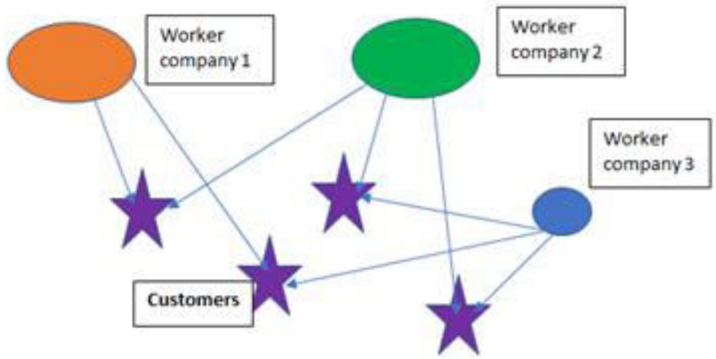

Figure 3: An inter-company network (The area of circles indicates the weightage of the company and/ or weightage of the persons)

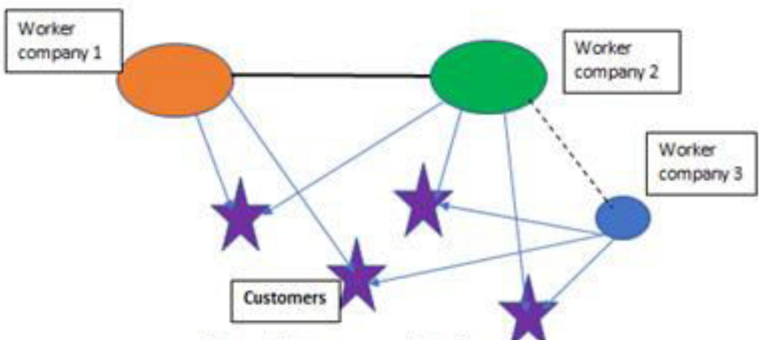

Figure 4: Inter-company relationship prediction

IV. Experimental information

This section describes the experiment in the year 2009 version provided by Carnegie Mellon University contains a selected Enron email server containing 280 folders with 20,581 files total. The email and received addresses, the CC and Bcc carbon copy, the date, the time, subject and the body and other email details are included in each file. The information was processed, and 161 active Yonghua were selected, and five relations selected following.

Adjacency: if two objects are related to send-receive, they will be connected. The sending relationship: when two objects send mails to the same object, the link is there. The relationship between co-recipients: when two objects receive emails from the same object, they have the connection. Relationships co-copying, two things together have the relation when copying the email to the same thing. Job relationship: when two objects are in the same place, they have a connection.

For the Enroll-extracted application, the email list procedure is as follows: for every email, the following fields have been extracted, for example, information, a user sent, recipient and CC user, and the user name extracted from the middle name have been removed. A piece of information was registered in the sending connection list from the sending user to the receiving user (when there were multiple receiving users, a plurality of pieces of information would be generated in the sending relationship list). In the CC list, a message was registered from the user sent to the user.

The resulting data have been saved in each part of the tensor for each relation. The 20, 40 and 60% proportions have been chosen. The value of the observed relation mode is 0. The tensor size was N to N to N to K. N was the number of objects, and K the number of connections. Patterns of N to 1⁄4M were obtained. 20% M, 40% M and 60% M have been discarded as test results, respectively. The process chosen was a random procedure, and the relation mode was selected as 0 in the weight matrix, with the other marked 1.

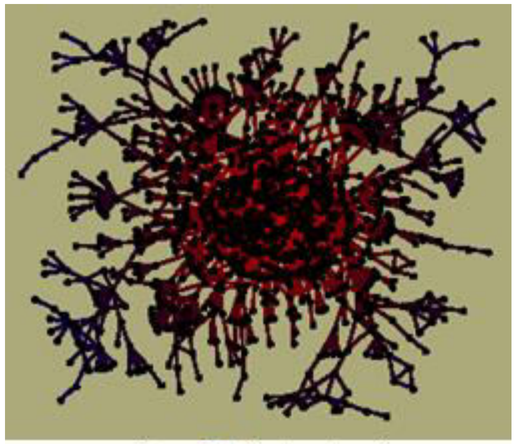

Figure 5: A cluster network

The consequence of the estimated value was obtained following the tensor: factorisation. Every result was compared with the real values. By measuring the AUC value for the ROC curve, the ROC curve of the prediction model was obtained and evaluated. Compared to SVD this experiment. In order to get an approximation result, every piece of the tensor (i.e. the adjacency matrix) was calculated by the SVD process. The real value was also compared to the ROC curve and its AUC value.

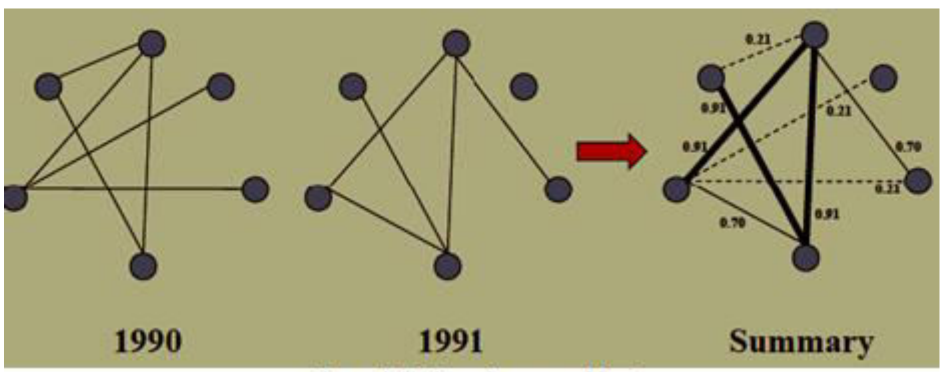

Figure 6: Link prediction medthod

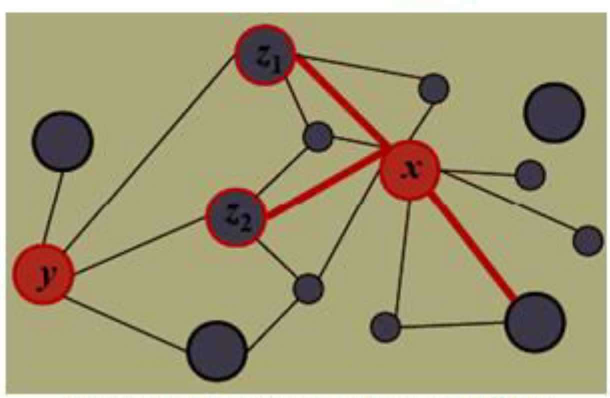

Figure 7: Link prediction using common neighbour

Secondly, a certain proportion of values as test results, set to zero, have been selected.
Thirdly, the ranking size and number of iterations have been selected for the decomposition of tensor factorisation or operation of SVD.
Fourthly, the results measured were used to determine finally whether the two items were connected to each other.
A total of 30 factorisation iterations were chosen as the maximum number. Three estimates of ten, twenty and thirty were selected for Tensor rank. In order to test the model's efficiency, the AUC value of the ROC curve was used. The ROC curve was the function curve of the recipient, also known as the sensitivity curve. This name reflects the

ISSN- 2394-5125 VOL 7, ISSUE 02, 2020

same susceptibility of the points on the curve. They were all answers to the same signal stimulus, but they achieved only the results under varying criteria. A graph with the incorrect positive probability of being the horizontal axis and the likelihood of hitting the vertical axis is the receiver operating characteristic curve. The curve was drawn by different results produced under specific stimulating conditions by different evaluation parameters. The abscissa showed the incorrect positive rate and the order showed the true positive rate. The estimated value of the relationship could be derived when the unobserved data is estimated. This study could obtain a truly positive and false-positive result rate by setting different closed values and then drawing a ROC curve. The wider the area under the curve, the more discrimination and the greater the model's accuracy. The AUC value of the ROC curve was the area between the curve and the horizontal axis.

V. Review of test results

First, if the data is not observed, the 20% value is chosen as the ROC result. Second, the value of 40 per cent is chosen as the result ROC for unnoticed data. Third, if the data are not observed, the value of 60% is chosen as the ROC curve result. From the above information it can be seen that the link prediction method based on connection mode is stronger than SVD, regardless of the 20, 40, and 60 percent ratio (when the rank is 10). Therefore the link method based on link mode is diversified information mining in consideration of the influence between multiple relations. The conditions for connection prediction have been further limited and the accuracy of connection predictions improved.

80 percent are also chosen for the selected data ratio in the Enron email dataset. However without taking into account the influence of various relationships, the liaison prediction method based on the liaison mode and the SVD method did not achieve good results. The test results are not stable, showing that previous social network expertise is inadequate to obtain additional prediction information if the data is very scarce. The sense of the relation forecast is not great in this situation.

The following assumptions are drawn for various values of R. The AUC value of R 1⁄4 20 for the CP-ALS algorithm will be the highest when a ratio of 20 per cent is chosen as test results. For the CP-ALS algorithm, the AUC value is the highest when R 1/2 10 was selected for a 40 and 60 percent ratio. The explanation is if the 20% ratio is chosen as the trial data, as the information on links found in the data is very rich, the properties and relationships of the objects are relatively high. R is selected in a larger ranking, plus features. This increases the precision of the relation prediction. For 40 and 60% of the data, however more link values are hidden by the ratio of

Stanley Milgram [Was94] was the individual building the current social network theory. One entity is the network node, while the edges which link nodes and are also called "connections" "links" correspond to the relationships of persons as shown in Fig. 1. Many examples of social network services are available, like:

- MySpace; - MySpace;
- Facebook; - Facebook;
- Twitter, etc.
- Youtube, etc.

Why do we want Facebook more? Due to its scale, 100,000 users a day are doubling once every six months. More than 60% of users are out of school age. The social graph, the network of connections and ties between the service people, was attributed to Zuckerberg by the Facebook force. "This is why Facebook works," he said.

Social networks are highly competitive, sparse and collective, so their result is hard to predict. In addition, popularity can be calculated later on due to early-stage popularity.

The role of predicting the existence of links or links in a domain is, on the other hand, an exciting challenge. Can we expect more about social networking, then?

As the network's ties, its maintenance and its quality reflect individuals' social behaviour, research can help quantitatively and qualitatively evaluate human relations in the age of information society. In addition, relation prediction is applicable to a number of fields, including bibliographic fields and molecular biology. In inductive logic programming (ILP) [Lav94], the most frequently discussed methods for inducing relationship patterns. ILP concerns the induction from the first-order logic of data induction of horn-clause rules (i.e. logic programs).

There are several simple approaches in social networks for dealing with relation prediction: supervised vs unmonitored. Directed graphical models versus undirected graphical models (for example, Markov Networks) as Bayesian networks and PRMs (Get02) can easily capture the dependence of the presence of links on attributes.

The problem of prediction is equal to the question of prediction of network structures, especially where social networks are concerned. There are many well-known methods for predicting connections based on:

· Towards node information;
· Taking into account structural details.

VI. Findings

The role was discussed with the project for predicting previously overlooked connections in social networks. Both social graphic representation and social network definition have been established.

The corresponding work has been examined in the research field. There have been similarities between the current approaches: supervised versus the unmonitored representation of single-table data as a vector vs. relational mining etc. The use of a range of inducements, like J48, OneR, IB1, Logistic, NaiveBayes as well as other approaches available for resolving the given task was digested in the case of the interpretation relation prediction task as a classification task. PRMs, SVM, GP, BN.

Network-structured data mining tasks have been analysed. There were a deep classification and explanation of measures for relation prediction approaches in mathematical representation in the Social Networks.

The experiment was designed using the free, open-source SQL full-text search engine Sphinx technique for crawling. The proposed corpus of training is the social network platform of Facebook. SQL format was displayed for the database of the selected data. Diagrams have been identified on visualisation software for social networks.

Research in a variety of scientific fields has shown that social networks function on several levels from families to nations, playing a key role in deciding how conflicts are resolved, organisations are managed and how people are efficient.

VII. Conclusions and vision for the future

This work is regarded as the first model to consider the prediction of negative relations in the field of interaction prediction. We demonstrated how our model could predict the links from which the network can be removed. This can be seen as the first step in developing a forecasting model for the potential associations between doctors in the field of healthcare and the network of gene expression; reducing reference traffic can be very beneficial. Moreover, the co-expression values of genes have been predicted successfully by our proposed connection prediction model within the network. In comparison with the internal relation approach, it is compared our PLP approach and concluded that our policy yielded better interpretations of reference data. It is applying the same model to other areas, including terrorist networks, financial data and workers networks to demonstrate the relevance and applicability of the proposed model to different domains. In order to see if this will impact the decision on the relation prediction, future directions will rely upon the inclusion of patient data. It is assumed it leads to more reliable forecasts that show enhanced decision-making processes like patient data. In future, the technique will be applied for large business network data. This will capture the real scenario of relationship predictions in social networks.

References

Allali O, Magnien C, Latapy M (2011) Link prediction in bipartite graphs using internal links and weighted projection. In: 2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS), pp 936–941.

Anderson JG, Jay SJ (1985) Computers and clinical judgement: the role of physician networks. Soc Sci Med 20(10):969–979

Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A (2005) Reverse engineering of regulatory networks in human B cells. Nat Genet 37(4):382–390

Borgatti S (1997) Network analysis of 2-mode data. Soc Netw 19(3):243–269

Cao L, Tay F (2003) Support vector machine with adaptive parameters in financial time series forecasting. IEEE Trans Neural Netw 14(6):1506–1518

Clancy CM, Lanier D, Grady ML (1996) Research at the interface of primary and specialty care. AHCPR 96-0034

Clauset A, Moore C, Newman MEJ (2008) Hierarchical structure and the prediction of missing links in networks. Nature 453(7191):98–101

Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S (2008) Feedback effects between similarity and social influence in online communities. In: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, pp 160–168

DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680–686

Dunlavy DM, Kolda TG, Acar E (2011) Temporal link prediction using matrix and tensor factorisations. ACM Trans Knowl Discov Data (TKDD) 5(2), Art no 10

Gao S, Alhajj R, Rokne JG (2009) Modeling knowledge discovery in financial forecasting. In: Proceedings of the

ISSN- 2394-5125 VOL 7, ISSUE 02, 2020

IEEE International Conference on Information Reuse and Integration, pp 41–46
Getoor L, Diehl CP (2005) Link mining: a survey. SIGKDD Explor Newsl 7(2):3–12
Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the 27th international conference on Human factors in computing systems, CHI '09. ACM, New York, pp 211–220
Hasan MA, Chaoji V, Salem S, Zaki M (2006) Link prediction using supervised learning. In: Proceedings of SDM 06 workshop on link analysis, counterterrorism and security, pp 936–941
Kianmehr K, Alhajj R (2009) Calling communities analysis and identification using machine learning techniques. Expert Syst Appl 36(3, Part 2):6218–6226
Koh HC, Tan G (2005) Data mining applications in healthcare. J Healthc Inf Manag 19(2):64–72
Lee CHL, Liu A, Chen WS (2006) Pattern discovery of fuzzy time series for financial prediction. IEEE Trans Knowl Data Eng 18(5):613–625
Lee T, Pappius EM, Goldman L (1983) Impact of inter-physician communication on the effectiveness of medical consultations. Am J Med 74(1):106–112
Liu W, Lu L (2010) Link prediction based on local random walk
Naji G, Nagi M, Elsheikh AM, Gao S, Kianmehr K, Özyer T, Demetrick D, Alhajj R, Rokne J, Ridley M (2011) Effectiveness of social networks for studying biological agents and identifying cancer biomarkers. In: Counterterrorism and open source intelligence, pp 285–313
Nowell DL, Kleinberg J (2003) The link prediction problem for social networks. In: Proceedings of the twelfth international conference on information and knowledge management, CIKM '03. ACM, New York, pp 556–559
Ressler S (2006) Social network analysis as an approach to combat terrorism: past, present, and future research. Homel Secur Aff 2(2).
Tylenda T, Angelova R, Bedathur S (2009) Towards time-aware link prediction in evolving social networks. In: Proceedings of the 3rd workshop on social network mining and analysis, SNA-KDD '09. ACM, New York, pp 1–10
S. Samanta, and M. Pal (2016), Fuzzy k-competition graphs and p-competition fuzzy graphs, Fuzzy Information and Engineering 5 (2), 191-204.
K. Das, S. Samanta, and M. Pal (2018), A study on centrality measures in social networks: a survey Social network analysis and mining 8 (1), 13.
R. Mahapatra, S. Samanta, M. Pal, and Q. Xin (2020), Link Prediction in Social Networks by Neutrosophic Graph International Journal of Computational Intelligence Systems 13 (1), 1699-1713.
S Samanta, and M Pal (2015), Fuzzy planar graphs, IEEE Transactions on Fuzzy Systems 23 (6), 1936-1942.
S Samanta, M Akram, and M Pal (2015), m-Step fuzzy competition graphs, Journal of Applied Mathematics and Computing 47 (1-2), 461-472.