IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 26, 2020, accepted January 9, 2021, date of publication January 25, 2021, date of current version February 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3053995

A Systematic Analysis of Link Prediction in Complex Network

HAJI GUL[1], ADNAN AMIN[1], AWAIS ADNAN[1], AND KAIZHU HUANG[2], (Member, IEEE)
[1]Center for Excellence in Information Technology, Institute of Management Sciences, Peshawar 25000, Pakistan
[2]Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China

Corresponding author: Adnan Amin (adnan.amin@imsciences.edu.pk)

This work was supported in part by the National Natural Science Foundation of China under Grant 61876155, in part by the Jiangsu Science and Technology Programme under Grant BE2020006-4 and Grant BK20181189, and in part by the Key Program Special Fund in XJTLU under Grant KSF-T-06.

ABSTRACT Link mining is an important task in the field of data mining and has numerous applications in informal community. Suppose a real-world complex network, the responsibility of this function is to anticipate those links which are not occurred yet in the given real-world network. Holding the significance of LP, the link mining or expectation job has gotten generous consideration from scientists in differing exercise. In this manner, countless strategies for taking care of this issue have been proposed in the late decades. Various articles of link prediction are accessible, however, these are antiquated as multiples new methodologies introduced. In this paper, give a precise assessment of prevail link mining approaches. The investigation is through, it consists the soonest scoring-based approaches and reaches out to the latest strategies which confide on different link prediction strategies. We additionally order link prediction strategies because of their specialized methodology and discussion about the quality and weaknesses of various techniques. Additionally, we compared and expounded various top link prediction techniques. The experimental results of these techniques, over twelve data-sets are ordered here based on performance, RA, 0.7411 > AA, 0.7285 > PA, 0.7202 > Katz, 0.7141 > CN, 0.6951 > HP, 0.6924 > LHN, 0.6017 > PD, 0.3978.

INDEX TERMS Link prediction, complex networks, prediction and recommendation.

I. INTRODUCTION

In recent years, the real-world complex network become a prominent part of our lives, so it is an emerging challenge for the researcher to study and analyze the data in a given network. Such real-world complex network can be expressed by a graphical shape where vertices or nodes maps to social entities or persons and links correspond to the collaboration or association among nodes, social entities, or persons. The interactions between individual changing continuously, so the deletion and addition of many vertices and edges happen. In the present scenario, we greets a particular issue call as an Edge mining issue. In the intricate network link expectation is a powerful issue, aim of connection expectation issue is to appraise the likelihood of a connection between couple of nodes. Many link prediction applications are involved with discovering of conceivable relationships in huge organizations, disconnected or hidden graph links indicating, social network friend suggestion, in the real-world complex networks [1], [2].

For quite a while, the complex network has been fundamentally studied by mathematicians and network engineers. There was very little weighty improvement in the complex network research territory until the 1960s.

Complex network give a strong abstraction for describing the structure of real-world social technological, biological and social complex network system. While most of the data on real-world complex networks are still imperfect. For a while, social interaction between people may be easy, intentionally unseen, or also we can say unobservable [3], the relationships between cells, species, or genes must be examined by high-priced experiments [4], [5], and the relationships arbitrate by a specific technique exclude all off-platform relationships [6]. The existence of missing link depend on the question of research, spectacularly change scientific conclusions when examined the structure or designing the dynamics of a real-world complex network.

The link prediction problem is essentially considered from three points: (I) Topological (ii) Node (iii) Social theory,

The associate editor coordinating the review of this manuscript and approving it for publication was Tingwen Huang.

of nodes and associations. Some link prediction methods merge the power of node and topological techniques for link mining [7], [8], additionally many others hybrid methods developed for link prediction problem [9].

A larger part of the advancement in the zone of structure-based craving for joins has been made by mathematicians and physicists. The topological method is additionally isolated into three classifications common neighbor, path, and irregular walk based, social hypothesis likewise contains five classes network, ternion, topological hole, tie quality and homophily. A bit of the notable structure-based connection estimation procedures are mutual neighbor, jaccard's, adamicadar, katz and so for [10].

Moreover, the problem of link prediction itself has turned into a standard for analyzing and comparing the design of real-world complex network structure [11], [12]. Many real-world complex networks are relatively scarce, and the amount of unconnected links in a network quadratically grows, such as $O(n)$ for linear growth and $O(n^2)$ for quadratic growth. In the given growth complexity n express the numbers of nodes connected through the edges. To estimate the probability of a link there are many methods introduced by researchers [13].

The graph represents a network which contains edges and nodes and can be used for representing real-world phenomena and provide valuable structure for understanding the relationship between communities. The graph also shows the importance of uncovered links between nodes. A social network represents friendship among individuals and provides information about communities. Computer network represents how digital work is done and help to maintain cybersecurity. In bioinformatics, the most studied network is the protein-protein interaction network which contains vital information about cell organelle interaction and provides information about how to cure the disease efficiently. All this information are in the node and edges but to get this information is a very difficult task because these real-world complex systems are constructed up with a million and billions of nodes and edges. every network is not fully connected and many links are not present but they are of very importance. This analysis of the organization is alluded to as a connection forecast issue. The link prediction is majorly done on static data but as it is time-dependent problem i.e. network change or evaluates with time. This type of network is referred to as a dynamic network. The global topological information can be obtained by the adjacency matrix. In the adjacency matrix, a non-zero value represents a link between two nodes and a zero value represents the missing link i.e., not present. Generally, a link prediction algorithm tries to recover the missing links.

After the abstract in section I, the introduction of the paper was included. Link prediction problem and literature review explained in segment II, while segment III communicates the proposed systematic framework for connection expectation, second last area IV comprise the experimental results and conversations, at long last, segment V holds the conclusion of the work.

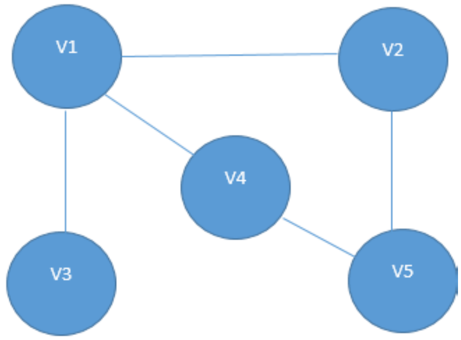

FIGURE 1. Analyzing Complex Network for Link Prediction.

II. LINK PREDICTION PROBLEM AND LITERATURE REVIEW

A. LINK PREDICTION PROBLEM

An extra addressable issue is to recognize the link between two nodes. How new connections are made? Especially the issue communicated here, the prediction of connection among couples of nodes, later on, is a link prediction issue [14]. Link Prediction procedures endeavor to predict the likelihood of a future relationship between two nodes in a given complex network. This can be used in biological frameworks to investigate protein-protein coordinated efforts or to propose a potential friend to a friend in an online social network (Facebook friendship). Because of the huge amount that is accumulated today, there is a prerequisite for versatile approaches to this issue. Analyzing link presently not exist in the data is a critical problem. This is divided into two categories that would have similar arrangements and solutions, the missing link prediction problem, and future link prediction problem. Looking for the missing link that should be in the network but certainly not observed, whether from data estimation errors or unknown information. The link prediction problem centers around those links that may happen after in the future, related to the reviewed real-world complex network.

Assume a network without direction $G = (V, E)$, where the set of nodes/vertices denoted by V and set of links/edges by E. A network might be undirected or directed. Right now will think about the undirected network, where self-edges and multiple edges are avoided. A universal set consisting of all possible links denoted by $\frac{V(V-1)}{2}$, where V expresses the total number of vertices in a given network. If a link exists between x and y, it can be express by $e = (x, y)$, at that point the reaction worth will be 1 otherwise 0 if the link does not exist between pair of nods. Link prediction aims to identify the link between nodes based on previous knowledge.

Suppose a given complex network in figure 1 at time t, in which we want to analyze the future link at time $t + 1$. Let us consider the co-authorship network of researchers, we can predict all possible links that will be forming in the future. In this given network there are total four nodes. Suppose V_2 is the node that needs to create a link with another node in the given network, here node V_2 is the source nodes and the predicted node will be the destination nodes. In figure 2 V_2 has been predicted a link with V_4 based on common neighbor link predictor and denoted by a red dot line. V_2 also can create a link with V_3 but based on common neighbor there is only one common neighbor, so that must not be the predicted link.

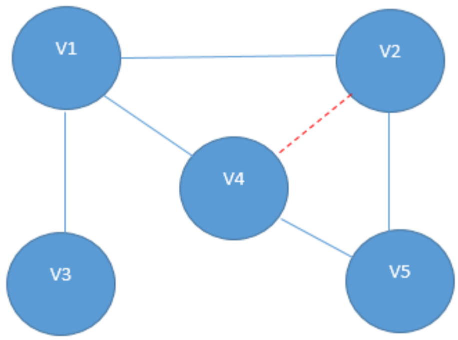

FIGURE 2. Predicted a Link Based on Common Neighbor.

B. APPLICATION OF LINK PREDICTION

Aside from the function of connection expectation as a fundamental question, it could be associated with many applications of real word Complex network, the latest survey about link prediction techniques and applications [15].

Spam Mail Detection: In this situation, LP indices can analyze anomalous emails for traffic reporting in various communication channels [16] based on the graph theory approach.

Social Network Privacy Control: LP can be used to hide the important information of users such as phone numbers, photos, etc. from other unreliable users [17].

Expert Detection: LP deals with predicting the expert in a given field. For example, a network that deals with corresponds to the links among different researchers, in this situation we need to predict experts from the network. In this scenario, LP can also apply in co-authorship networks for prediction or analyzing the domain expert [18]. A link prediction method is utilized by [19] and represents that how candidates can be ranked for high-level government posts. The most relevant field of expert detection is customer behavior prediction explained in [20], [21].

Recommended Systems: This is an information filtering system that recommends and analyze new items for the clients dependent on the past inclination or rating to indistinguishable items. We can apply LP to improve the quality of the recommendations systems [22], [23].

Disease Prediction: link prediction method developed by [24], which is used to identify or anticipate the beginning of infections utilizing the current well-being status of a patient.

Influence Detection: A method proposed by [25], that can be used to assess high impact clients in a cooperation organization. In this methodology, the user or node adds and removes iterative based on the link prediction procedure. Many fields contain link prediction applications. Related prediction framework introduce for students cognitive skill's measurement [26], [27], [28].

C. RELATED WORK

Recently many researchers incorporated time data in network inserting to make it fit for holding onto dynamic development of the mind boggling network. A method introduced that used to design the dynamic changes of complex network shape [29]. To learn the temporal transitions in a given complex network a model created by [30], deep auto-encoder based frame introduced which aims to generate highly non-linear embedding [31]. A hybrid model based on graph convolution network developed and its aim to memorize complex network structure and temporal characteristics [32]. Although, such techniques mostly require the whole graph for input which results int low efficiency for the large complex network graph.

From the literature, social complex network information is incomplete since just a small a piece of social data can be assembled from informal organization stages. Second, informal organizations are exceptionally powerful. Along these lines, predicting the absent or imperceptibly interfaces in current interpersonal organizations and as of late included or eradicated joins later on, informal communities are significant not only for understanding the advancement of informal communities while likewise for the completion of structure of the current social complex system. This issue is customarily known as the Link expectation problem. Link prediction has numerous applications, It tends to be seen that an ever-increasing number of works focus on the link mining in informal communities, especially in the past five years, there are numerous of papers related to this issue every year. As such, the investigation design regarding this matter is developing [33]. Many research papers are also available in which existing algorithms are combined and estimate the probability of link prediction, such as combining the power of similarity and connectivity of nodes [34]. Link prediction in complex networks [10], a survey In which detailed included that an undeniable downside of the most extreme probability framework are that, it is exceptionally monotonous (tedious), while the probabilistic model will upgrade a collected objective ability to set up a model which can best fit the watched or studied information. Most of the features-based prediction methods follow the machine learning approach, and mostly use grouping based style to make forecasts. Broadly strategies utilized Decision Tree, Support Vector Machine (SVM), and Naive Bayes [35], [36].

In [36]–[38] creators announced that the exhibition of connection forecast improves when AI approaches are utilized. In most cases utilizing extra information about the networks are not constantly accessible. While many methods need just require the fundamental structure of complex network data-set and in this way many researchers exclude machine learning methods in their experiments [39].

Predicting unobserved link further divided into two categories (i) Missing Link Prediction and (ii) Future Link prediction. The links already created in the network and due to some issues these links are removed, these are missing links. While those links already not observed and to be created in the future, there are future links.

It is unimaginable that a solitary technique performs well over all the data-sets. Since these data-sets have a place with various domain and contain various topological features such as degree, neighbor, path length, and label for supervised techniques, and so forth.

Many link mining techniques have been presented by the scientists. Every strategy has various functionalities to anticipate links in an unpredictable system. For example, path, common neighbor, degree, supervised, unsupervised, and social hypothesis based link prediction techniques. Underneath we have included weaknesses of existing link prediction techniques.

Many link mining techniques have been presented by the scientists. Every strategy has various functionalities to anticipate links in an unpredictable system. For example, path, common neighbor, degree, supervised, unsupervised, and social hypothesis based link prediction techniques. Underneath we have included weaknesses of existing link prediction techniques.

Supervised Methods: Because of the inaccessibility of label for preparation/training, in some cases, it decreases the accuracy of prediction [40].

Unsupervised Methods: When weighted extended, and otherwise modified, unsupervised link predictors are still domain-specific and inflexible.

Topological Based Methods: Topological based method mostly predict link based on node features and information. But all the node (user in a social network) information mostly publicly not available and it can disturb prediction accuracy.

Path-Based Methods: Shortest path fever to prediction link between those co-authors/nodes with distance 2. While it has been proved in [14] that, many pairs who collaborate are at a distance greater than 2 and new edges form between pairs at distance 3 or greater.

Link prediction consists of two main steps, features extraction and prediction, as a result of these features inductions, are made the overall engineering for a connection estimation framework. Right now, will just concentrate on predicting the occurrence of links. We enumerated diff rent link prediction techniques and also used for experiments.

D. TOPOLOGICAL BASED METHOD FOR LP

In the literature, many topological information-based techniques present for interface forecast issue by the analyst. For connection analysis dependent on topological features. An outline of the topological based link prediction is given in [14]. Topological-based methods are further divided into three categories. (i) Neighbor based, (ii) Path-based, and (iii) Random walk based link prediction. In this section, we explained commonly used topological base link prediction techniques.

Preferential Attachment (PA) [41]: Preferential Attachment link prediction method based on a very common framework. This method simply considers the number of neighbors between two nodes and decide the probability of an unobserved link. According to Preferential attachment (PA), two nodes with more neighbors i.e. higher degree nodes are more likely to be linked with each other. It can be represented by the formula given below.

$$PA(x, y) = |\Gamma(x) \cdot \Gamma(y)| \tag{1}$$

Common Neighbors (CN) [42]: Common neighbor is the most widely used method for link prediction problems, due to the reason of its simplicity, it can be computed very easily. The common neighbor can be defined as if a node z has direct relationship with other nodes i.e. x and y then the node z is a common neighbor of x and y. This method is very similar to the friend of a friend approach (Facebook friend of a friend). In link prediction, common neighbor follow the strategy that, if there are more common neighbors between a node suppose x and node y, so it will maximize the probability of link prediction between these nodes (node x and y). Mathematically it can write as

$$CN(x, y) = |\Gamma(x) \cap \Gamma(y)| \tag{2}$$

Jaccard Coefficient (JC) [43]: Jaccard coefficient considers the strength of Common neighbors in such a way it is proportion of the Common neighbor to all the neighbors of node x and y. The JC follows a strategy that the probability of link prediction will be grater between two nodes if their Common neighbors have great a proportion to their total neighbors. This method can also define in simple words as, intersection over the union set of nodes x and y. The Jaccard coefficient can be written as

$$JC(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \tag{3}$$

Salton Cosine Similarity (CS): The Salton Cosine Similarity is a cosine metric used to analyze the similarity between two nodes. The CS can be defined as, the proportionality of common neighbor to the square root of the total number of neighbors node x and y. If the proportionality result is high then also the link prediction probability will be high between vertices. Also we can say, the Salton cosine similitude is the proportion between normal neighbors and square root to the preferential attachment. It can be written as

$$CS(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{\sqrt{|\Gamma(x) \cdot \Gamma(y)|}} \tag{4}$$

Adamic and Adar(AA) [44]: Adamic-Adar is another link prediction method that is a variant of Common neighbor. It was first introduced by Adamic and Adar. This method is globally used due to its link prediction high accuracy.

Liben-Nowell and Kleingberg found that the Adamic Adar prediction accuracy is better than most of the other state-of-the-art prediction algorithms.

AA follows the method that, two nodes x and y have less number of their Common neighbor i.e. low degree nodes has a high probability of link prediction between them. A New person in a group may become the most popular and create links with multiples persons/nodes. AA mathematical formula can be write as

$$AA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log |\Gamma(z)|} \tag{5}$$

Resource allocation (RA) [45]: This strategy proposed by zhou et al which follows the principle of physics for link prediction. The resource allocation is a similar concept as to the adamic adar, because it also punishes the higher degree nodes and take the decision on lower degree nodes. Due to this, the result of both RA and AA are very close in prediction for smaller average degree networks but for higher degree networks the RA performs well as compare to AA. RA not only consider the common neighbors it also supposes neighbors of the common neighbor. The mathematical formula cab be write as

$$AA(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{|\Gamma(z)|} \tag{6}$$

Hub Depression (HD): HD work somehow similar to hub promoted link prediction method but the main difference is that, it considers the higher degree node from the pair of nodes such as x and y. Mathematically it can be express as

$$HD(x, Y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{max(|\Gamma(x)|, |\Gamma(y)|)} \tag{7}$$

Katz [46]: Katz is a path-dependent grid that assesses all the path among a couple of vertices (node x and y) and give more weight to the path with the smallest/ the shortest path length. Katz formula can be write as

$$Katz(x, y) = \sum_{l=1}^{\infty} \beta^l . |Path^l_{x,y}| = \beta A + \beta^2 A^2 + \beta^3 A^3 + \cdots \tag{8}$$

Hub Promoted (HP) [47]: Hub Promoted is also a link prediction method. HP works based on topological information and it can be defined as the ratio between common neighbors and a lower degree of nodes form with a pair of nodes. The mathematical representation of Hub Promoted index as given below.

$$HP(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{min(|\Gamma(x)|, |\Gamma(y)|)} \tag{9}$$

E. NODE BASED METHODS FOR LINK PREDICTION

Node similarity methods predict links based on the feature and similarity of nodes, the more similarity between pair of nodes mean that the likelihood probability will also be more of a link between them [48]. In an interpersonal organization, a vertex contain various credits as some of which we illustrated here name, age, gender, email, region, etc. Based on this information, similarity can be computed between pair of nodes and compute the probability of links existence.

F. SOCIAL THEORY-BASED METHODS FOR LINK PREDICTION

For the link prediction problem, there are recently growing attentions towards social theory such as strong and weak ties, structure balance, community, and triadic closure. Link prediction performance can be improved by using additional social interaction information with topological or node-based link predictor. In [49], authors have merged the topology method with social theory additional community information such as user behavior and interest to improve the accuracy and time complexity of link prediction problem in a given real-world complex network.

Based on social networks' user suggestions, Social media contains important and huge information, while in unstructured format to automatically investigate these data a novel method proposed in [50]. A related model proposed by [51] that automatically play-up the crucial field in an image, while the traditional methods work based on boundary nodes. These traditional methods are also not able to work automatically. Noisy conditions can crate huge damage for tolerate of listening loss, to remove audio noise, a perfect method introduced by [52]. Many algorithms proposed in the recent years for biomedical field. For the behavior of bio-inspiration and algorithmic a novel framework developed by [53]. These algorithms follow different mechanisms, such as neural network ect. Neural network is one of the important methods that provide solutions for various problems in many fields. This network contains different layers, the importance of neural network explained in [54], [55].

III. PROPOSED SYSTEMIC ANALYTICAL MODEL

In this portion, the analytical model has been expressed and also its role in selecting the best, appropriate link prediction method is demonstrated. Since the shape and construction of each real-world complex network is different from other, many link prediction methods are introduced by the researcher. The purpose of this systematic analytical model is to give a stage that perceives the most natural, normal, and modern techniques.

This systematic analytical model classifies all these methods consistently and logically. In this model, with the presentation of rating criteria, the comparison and analysis of categorization and classification have been accomplished.

In this paper, we investigated a huge range of link prediction problem methods over different real-world complex networks. Hence, the structure and shape of the model are assembled based on some points we enumerated below. The purpose of this analysis was to explain the various sorts of strategies and recognize their adequacy and fairness.

- Data-set selection and Preparation, Training, and Testing Portions Separation.

TABLE 1. Data-Sets Information.

Datasets	N	E	MaxD	Avg D	Δ
Karate	33	78	17	4.5882	0.1477
Router	3722	6258	103	2.4927	0.00090371
Women Social Events	18	89	17	9.8889	0.5817
Us Power Grid	4939	6594	19	2.6691	0.00054074
Yeast	1458	1948	56	2.6722	0.0018
Face Book	2699	2981	769	2.0644	0.00081874
Sampson	18	189	28	21	1.2353
Contiguous USA	49	107	8	4.3673	0.0910
Kangaroo	14	91	15	10.7059	1.00
Train Bombing	63	243	29	7.5938	0.1244
Misc Co-Occurrence Character	74	254	36	6.598	0.0940
Dolphin	62	159	12	5.1290	0.0841

Note: In the given all the short terms expended here, N used for Numbers of Nodes or Vertices, E express Numbers of Edges or Links, MaxD represent the Maximum Degree of the data-set while Avg D shows the Average Degree of the Data-set. Δ = Density of the Data-set

- Apply Different Link Prediction Algorithms.
- Find and compare the Predicted Results of Each Method of Link Prediction.

A. DATA-SETS SELECTION AND PROCESSING

To start the process of link prediction firs of all, the adjacency square matrix has been created equal to the max size (max of max number in the data-set) of a real-world complex network. The adjacency matrix expresses a network in the form of matrix contains just 0 and 1, where 0 used that there is no link between the pair of nodes while 1 implies that the link exists between the pair of nodes in the network and mathematically can be represented by $e = (x, y)$. The production of evaluation is also in the form a matrix, in which the likeness score of nodes pair express by the component of network data-set. To find the performance of a prediction matrix provided by link prediction method we used AUC (area under the curve) evaluation matrix over different data-sets.

All the network data-sets used in the experiments are real-world complex networks. The Complex network is a network, where study the properties and characteristics of a real-world network. To perform different operations, the complex is then expressed in the graph. All these network data-sets are downloaded from .[1]

Data-Set Selection Criteria: This is much important to select data-set very carefully. Therefore, some basic and essential requirements for data-sets selection are, the data set must contain nodes interactions, associations, or some other type of activity that allows identifying the connections between nodes. Based on these criteria some data-sets are selected and its original information is shown in table 1, where N represents the numbers of nodes and E represents the numbers of edges. Maximum Degree is the number of connected links with a node. These all are undirected networks so incoming and outgoing links are not allowed here. The fifth column represents the average degree of the node where the last column is the density column, which represents the density of real-world complex networks. Δ can be calculated by the formula given below, where E use for numbers of edges and N for numbers of nodes in a real word network data set while Δ used for density.

$$\Delta = \frac{2(E)}{N(N-1)} \tag{10}$$

Data-sets: After data-sets selection, the following main features are analyzed, all these are unweighted and undirected real-world complex networks. Below we have briefly explained network data-sets used in experiments.

Karate Network: This real word undirected network contains humans in a karate club, where node identifies by a person and an edge between two persons shows that there was a relation between them.

In figure 3 we included an original Karate network, that shows how the members are interconnected to each other. A single member is a node and link is the relation, association between two members.

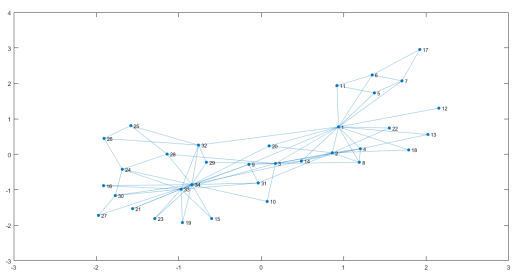

FIGURE 3. Karate: Real-World Complex Network.

Router Network: Internet network in which the router denoted by a node and the association represents link or edge.

Women Social Events: An undirected network which contain 18 nodes and 89 edges. In this network nodes expressed by women and edges express by the event attendance of women.

Us-Power Grid Network: A network contains information about the power grid of the western state United States of America. Every hub speaks to a generator, transformer, or substation while an edge speaks to a force gracefully line.

[1]http://konect.uni-koblenz.de/networks/

Yeast Network: This is a commercial leaving agent consist of yeast cells that represent nodes and the interaction between them represent by link or edge.

Face Book Network: A real-world network Facebook that representing people (users) of the social network. Nodes represent people and edges represent friendships or links. This is one of the most popular real-world complex networks, where the aim of link prediction to suggest future a friend.

Sampson Network: This is a human social network and node represents a monk and an edge represent by two monks that the left monk rated the right monk.

Contiguous USA Network: Total 48 contiguous states are there in the Columbia district of the United States of America. All the states are included except Alaska and Hawaii, they are not connected with other states by land. The link is denoted by the border between the two states while the states are denoted by nodes or vertices. We downloaded the updated data-set from the repository link included in the above section.

Kangaroo Animal Network: This is an animal real word network. It contains the interactions between free-ranging gray kangaroos. In this network, the node represents a kangaroo and an edge show the interaction between two kangaroos. The link or edge values represent the total count of interactions.

Train Bombing Terrorist Contact Network: The network contains suspected terrorist contacts involved in the train bombing of Madrid. A node identifies the fear monger and the edge between two psychological oppressors shows that there was contact between two fear mongers. This real-world network includes friendship and co-participating in the training camps or previous attacks.

Misc Network: This undirected real-world network contains co-events of characters in Victor Hugo's tale. A hub recognizes a character and an edge between two hubs shows that these two characters are in a similar part of the book. In this organization, each connection weight shows how regularly such a co-appearance happened.

Dolphin Network: A real word network representing the dolphins' social interaction. The node represents a dolphin and the edge expresses their interaction.

B. EVALUATION CRITERIA

AUC (Area Under the Curve) [56] express the accuracy and performance of a technique used for link prediction problem, which is appeared by the numerical equation given as

$$AUC = \frac{n' + 0.5n''}{n} \tag{11}$$

where n express to the quantity of created links set from unconnected links set and missing links set. $\acute{n}$ Express the number of times in which the missing links accomplished a higher score than undetected links. $\acute{\acute{n}}$ express to the occasions in which both absent and detached links produce an equivalent score. The AUC steady estimation of 0.5 is utilized when the score is produced by independent and the same distribution.

C. EXPERIMENT SETUP

In the methodology see figure 7, first of all, we imported the original data-set of a real-world complex network. As included that the direction and weight of all these real-world complex networks have been ignored here. Pre-Processing is the second step in methodology, if some important data-sets contain direction and weights these will be removed here in the pre-processing step, also it will check either the data-set complete or incomplete.

To create a zero square matrix, first the maximum number of data-set will be computed and then the matrix will be created equal to the size of maximum number in the data-set. This matrix will be converted into an adjacency matrix that expresses the relationship of nodes. An adjacency matrix is also the representation of a graph, this matrix is break into a training E^T set and testing/prob set E^P. All the algorithms will be train and test on training and prob sets. This is also must be analyzed that $E^T \cap E^P = 0$.

AUC (area under the curve) is utilized to discover the exactness of every calculation. Finally, compared all the results and select the best performer of link prediction problem. Step by step explanation elaborated below.

To make the data-sets adequate for experiments, there are some steps to be performed on the real-world network data-sets. We selected undirected and most popular real-world network data-sets. As included earlier that, direction and weight of all the networks are ignored in this paper.

Adjacency Matrix: In this step, the adjacency matrix will be created from the data-set. All original real-world network data-sets contain only two columns that identify the connectivity of nodes. While the adjacency matrix contains a binary value 0 and 1, the value 0 represents a negative prediction sign mean there is no connection between the pair of hubs, while 1 is the positive prediction sign and it expresses that, there is a link between the pair of nodes. A similar way has been used for all methods while experimenting.

Data-sets Sample: Each data-set as shown in table 1 randomly divided into 90% and 10%. The 90% data separated from the data-sets for training purpose while 10% data selected from the data-set for the testing purpose during the experiment.

Data-sets Splitting: We removed 10% links from each data-set shown in table 1 and created a connected graph from the remaining 90% data-set. This is the most important point while splitting the data-sets that the order and connectivity must be maintained of the remaining 90% data-set. Once the data-set is split into training and testing then referred both of these to the prediction methods for link prediction.

IV. RESULT AND DISCUSSION

This systematic analysis paper, where we have compared eight link predictors over the twelve most popular data-sets. All these data-sets belong to different domains such as social, animal, co-authorship and human interaction, etc.

TABLE 2. Numerical comparison results.

Data-sets	AA	PA	Katz	PD	RA	CN	HP	LHN
Karate	0.7113	0.72898	0.7125	0.41414	**0.7476**	0.6913	0.7019	0.5986
Router	0.7169	0.7153	0.7245	0.4177	**0.7428**	0.6997	0.7093	0.5846
Women Social Events	**0.7332**	0.7148	0.7194	0.3974	0.7303	0.6914	0.6757	0.599
US Power	0.7325	0.7092	0.6974	0.3784	**0.7351**	0.6984	0.6887	0.6147
Yeast	**0.7486**	0.7207	0.7308	0.3978	0.7286	.7126	0.6927	0.5837
Face Book	0.7206	0.735	0.7302	0.3879	**0.7586**	0.7046	0.7107	0.6071
Sampson	0.7253	0.7097	0.7359	0.3949	**0.7381**	0.675	0.6957	0.6176
Continuous	0.7354	0.7108	0.7039	0.4217	**0.754**	0.694	0.6827	0.6148
Kangaroo	0.7328	0.717	0.7051	0.3925	**0.7482**	0.679	0.6841	0.6096
Train bombing	0.7272	0.7304	0.697	0.408	**0.7354**	0.6918	0.6807	0.605
Misc Co-occurrences characters	0.7321	0.7225	0.6941	0.3773	**0.7454**	0.7052	0.6862	0.604
Dolphin	0.7266	0.7278	0.7185	0.3855	**0.7293**	0.6978	0.7008	0.5822

* Note: In the given table short terms stands for, AA Adami-Adar, PA Preferential Attachment, Katz, PD Path Dependent, RA Resource Allocation, CN Common Neighbor, HP Hub Promoted and LHN stand for Leicht Holme Newman link prediction index.

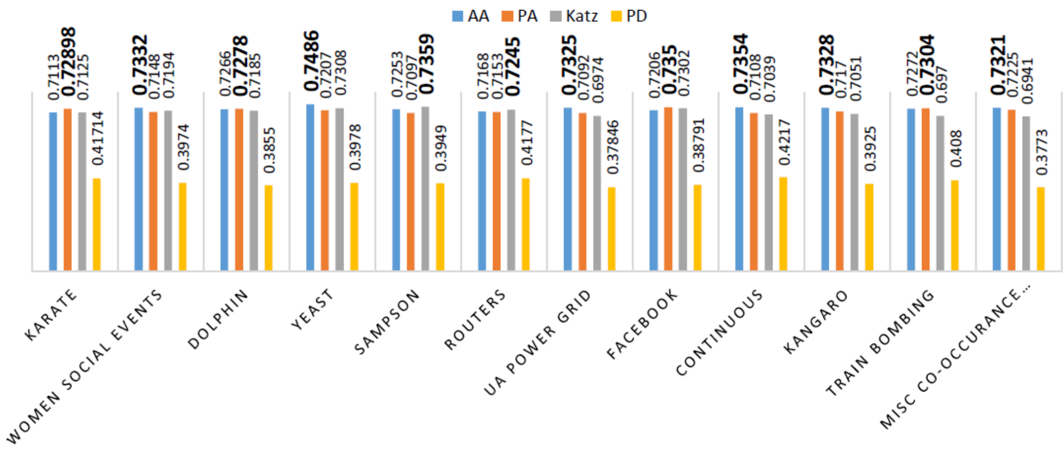

FIGURE 4. Visualization of Comparison Results I.

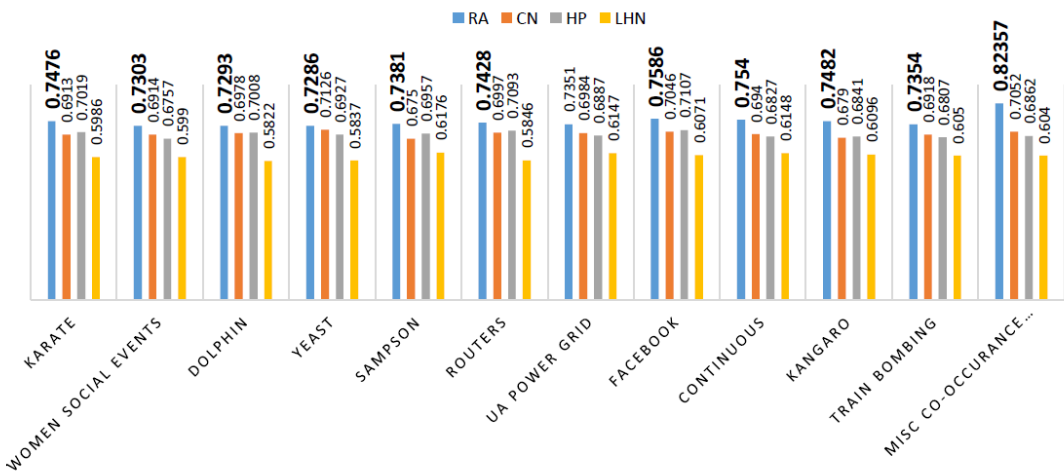

FIGURE 5. Visualization of Comparison Results II.

In table 2 there are twelve data-sets at the position of the first column while the first row contains the name of link prediction methods. Table 2 shows that the Resource Allocation (RA) performed very well as compared to other predictors and secondly, the Adamic-Adar performed well while the Katz link prediction method also not bad. The numerical and visual results of each algorithm on each data-set were included after the average of 100-time execution results. Each algorithm we execute 100 times on each data-set and then find the average result of the algorithm which shows the actual and powerful link predictor. In this paper, as we included that Resource Allocation (RA) > Adamic-Adar (AA) > Katz > Preferential Attachment (PA) is the ranking of link prediction methods. The complete evaluations are included in table2 which shows numerical results of the accuracy of all algorithms.

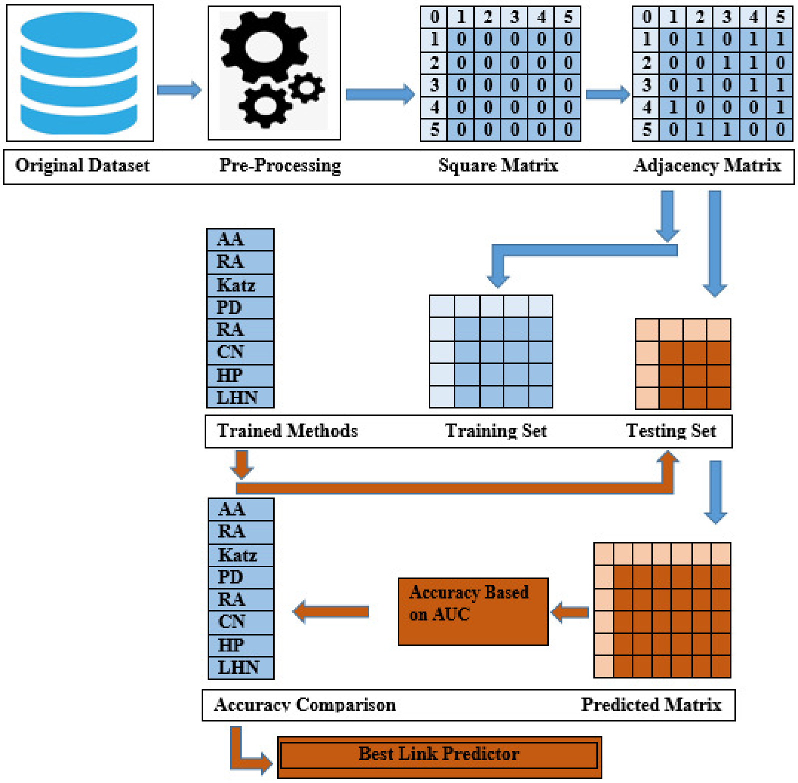

FIGURE 6. Methodology Diagram.

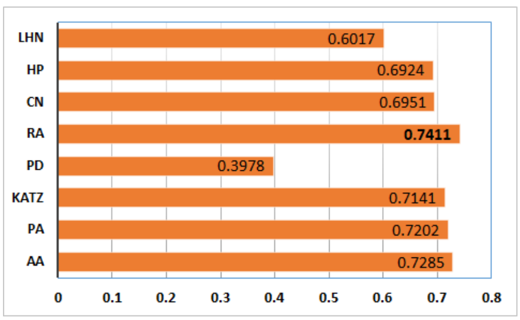

FIGURE 7. Average results.

Figures 4 and 5, there are totals of twelve data-sets and eight algorithms of link prediction, each group is the combination of four link predictors. Here we have compared figure 4 and figure 5, above (figure 4) group of four link predictors and below (figure 5) group of four link predictors. The goal is to choose the best one in all these eight predictors. As shown that resource allocation performed well as compared to other link prediction methods. So resource allocation will be a good choice to use for recommendation and other systems.

In the given figure 7 average results summary included, it clearly shows that the resource allocation performed well as compared to other link prediction state-of-the-art algorithms. In a few cases, the performance of the resource allocation is not good while mostly good, this depends on the structured and topological feature of complex network data-set.

A. FINDINGS AND RECOMMENDATIONS

As we realize that link mining has a crucial job and also contains applications in various domains, such as face-book: friend suggestion, kangaroo animal organization: collaboration expectation, human remote organization: to analyze future remote contact, and women's social organization: foresee future functions participation, and so on. But the prediction of links in these networks is very costly and time expensive. This is moreover important to realize which link mining will perform well over different data-sets. Through this systematic analysis model, we have anticipated links in various networks and indicated every technique's performance on various data-sets. The proposed model is helpful to analyze the link mining method over various field data-sets.

Despite the fact that the proposed systematic analysis model can be reached out to additional complex networks, for example, a weighted or directed system. Additionally, we have applied our model to static real-world complex networks. It would likewise be fascinating to investigate the role of our model on dynamic networks.

V. CONCLUSION

In a complex network link mining is a prominent issue and very useful in examining and understanding complex network structure and organization. The link prediction systematic analysis was spurred by the prerequisite to comprehend the possible favorable position of connection forecast procedures contrasted with one another. This paper granted a scientific system for interface expectation in complex organizations and explain that as of now there are different sorts of difficulties and techniques. This orderly examination comprises an auxiliary point of view and the segments are, data-sets selection and preparation, pre-processing, testing, and training data splitting, creation of an evaluation matrix, and selecting the best predictor. Finally, we included the comparison results in a numerical and visual format. All the present and significant works cultivated and afterward explored expectation, in view of the introduced investigation measures. The model could be used to oversee future investigations in complex organizations.

The work describes here, can be expended in the future for large data-sets. Many real-world complex network fields contain large data-sets where link prediction can be applied, such as twitter, ppi, etc. It would also be very good and interesting to apply over directed, and weighted Graph/Network.

REFERENCES

[1] J. Wu, H. Yang, Y. Ren, and X. R. Li, ''A two-stage algorithm for network reconstruction,'' *Appl. Soft Comput.*, vol. 70, pp. 751–763, Sep. 2018.

[2] J. Wu, N. Dang, and Y. Jiao, ''Reconstruction of networks from one-step data by matching positions,'' *Phys. A, Stat. Mech. Appl.*, vol. 497, pp. 118–125, May 2018.

[3] M. Fire, L. Tenenboim-Chekina, R. Puzis, O. Lesser, L. Rokach, and Y. Elovici, ''Computationally efficient link prediction in a variety of social networks,'' *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 1, pp. 1–25, Dec. 2013.

[4] M. Nagarajan *et al.*, ''Predicting future scientific discoveries based on a networked analysis of the past literature,'' in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 2019–2028.

[5] L. Lü and T. Zhou, ''Link prediction in complex networks: A survey,'' *Phys. A, Stat. Mech. Appl.*, vol. 390, no. 6, pp. 1150–1170, Mar. 2011.

[6] G. C. Kane, M. Alavi, G. Labianca, and S. P. Borgatti, ''What's different about social media networks? A framework and research agenda,'' *MIS Quart.*, vol. 38, no. 1, pp. 275–304, 2014.

[7] I. Ahmad, M. U. Akhtar, S. Noor, and A. Shahnaz, ''Missing link prediction using common neighbor and centrality based parameterized algorithm,'' *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, Dec. 2020.

[8] X. Wu, J. Wu, Y. Li, and Q. Zhang, ''Link prediction of time-evolving network based on node ranking,'' *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105740.

[9] H. Ghorbanzadeh, A. Sheikhahmadi, M. Jalili, and S. Sulaimany, ''A hybrid method of link prediction in directed graphs,'' *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113896.

[10] L. Lü and T. Zhou, ''Link prediction in complex networks: A survey,'' *Phys. A, Stat. Mech. Appl.*, vol. 390, no. 6, pp. 1150–1170, Mar. 2011.

[11] A. Ghasemian, H. Hosseinmardi, and A. Clauset, ''Evaluating overfit and underfit in models of network community structure,'' *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 9, pp. 1722–1735, Sep. 2020.

[12] T. Vallès-Català, T. P. Peixoto, M. Sales-Pardo, and R. Guimerà, ''Consistencies and inconsistencies between model selection and link prediction in networks,'' *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 97, no. 6, Jun. 2018, Art. no. 062316.

[13] V. Martínez, F. Berzal, and J.-C. Cubero, ''A survey of link prediction in complex networks,'' *ACM Comput. Surveys*, vol. 49, no. 4, pp. 1–33, Feb. 2017.

[14] D. Liben-Nowell and J. Kleinberg, ''The link-prediction problem for social networks,'' *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007.

[15] A. Kumar, S. S. Singh, K. Singh, and B. Biswas, ''Link prediction techniques, applications, and performance: A survey,'' *Phys. A, Stat. Mech. Appl.*, vol. 553, Sep. 2020, Art. no. 124289.

[16] Z. Huang and D. D. Zeng, ''A link prediction approach to anomalous email detection,'' in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, vol. 2, Oct. 2006, pp. 1131–1136.

[17] S. Al-Oufi, H.-N. Kim, and A. El Saddik, ''Controlling privacy with trust-aware link prediction in online social networks,'' in *Proc. 3rd Int. Conf. Internet Multimedia Comput. Service (ICIMCS)*, 2011, pp. 86–89.

[18] M. Pavlov and R. Ichise, ''Finding experts by link prediction in co-authorship networks,'' *FEWS*, vol. 290, pp. 42–55, Nov. 2007.

[19] J.-S. Liu and K.-C. Ning, ''Applying link prediction to ranking candidates for high-level government post,'' in *Proc. Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2011, pp. 145–152.

[20] A. Amin, S. Anwar, A. Adnan, M. Nawaz, N. Howard, J. Qadir, A. Hawalah, and A. Hussain, ''Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study,'' *IEEE Access*, vol. 4, pp. 7940–7957, 2016.

[21] A. Amin, F. Al-Obeidat, B. Shah, M. A. Tae, C. Khan, H. U. R. Durrani, and S. Anwar, ''Just-in-time customer churn prediction in the telecommunication sector,'' *J. Supercomput.*, vol. 76, no. 6, pp. 3924–3948, Jun. 2020.

[22] I. Esslimani, A. Brun, and A. Boyer, ''Densifying a behavioral recommender system by social networks link prediction methods,'' *Social Netw. Anal. Mining*, vol. 1, no. 3, pp. 159–172, 2011.

[23] Z. Huang, X. Li, and H. Chen, ''Link prediction approach to collaborative filtering,'' in *Proc. 5th ACM/IEEE-CS Joint Conf. Digit. Libraries (JCDL)*, 2005, pp. 141–142.

[24] F. Folino and C. Pizzuti, ''Link prediction approaches for disease networks,'' in *Proc. Int. Conf. Inf. Technol. Bio Med. Informat.* Springer, 2012, pp. 99–108.

[25] E. Perez-Cervantes, J. P. Mena-Chalco, M. C. F. de Oliveira, and R. M. Cesar, ''Using link prediction to estimate the collaborative influence of researchers,'' in *Proc. IEEE 9th Int. Conf. e-Sci.*, Oct. 2013, pp. 293–300.

[26] S. Ahmad, K. Li, A. Amin, M. S. Anwar, and W. Khan, ''A multilayer prediction approach for the student cognitive skills measurement,'' *IEEE Access*, vol. 6, pp. 57470–57484, 2018.

[27] G.-L. Huang, K. Deng, and J. He, ''Cognitive traffic anomaly prediction from GPS trajectories using visible outlier indexes and meshed spatiotemporal neighborhoods,'' *Cognit. Comput.*, vol. 12, no. 5, pp. 967–978, Sep. 2020.

[28] S. Ahmad, K. Li, A. Amin, and S. Khan, ''A novel technique for the evaluation of posterior probabilities of student cognitive skills,'' *IEEE Access*, vol. 6, pp. 53153–53167, 2018.

[29] L. Zhou, Y. Yang, X. Ren, F. Wu, and Y. Zhuang, ''Dynamic network embedding by modeling triadic closure process,'' in *Proc. 22nd AAAI Conf. Artif. Intell., (AAAI), 30th Innov. Appl. Artif. Intell. (IAAI), 8th AAAI Symp. Educ. Adv. Artif. Intell. (EAAI-18)*, New Orleans, LO, USA, Feb. 2018, S. A. McIlraith and K. Q. Weinberger, Eds. AAAI Press, 2018, pp. 571–578.

[30] P. Goyal, S. R. Chhetri, and A. Canedo, ''dyngraph2vec: Capturing network dynamics using dynamic graph representation learning,'' *CoRR*, vol. abs/1809.02657, 2018.

[31] P. Goyal, N. Kamra, X. He, and Y. Liu, ''Dyngem: Deep embedding method for dynamic graphs,'' *CoRR*, vol. abs/1805.11273, 2018.

[32] J. Chen, X. Xu, Y. Wu, and H. Zheng, ''GC-LSTM: Graph convolution embedded LSTM for dynamic link prediction,'' *CoRR*, vol. abs/1812.04206, 2018.

[33] P. Wang, B. Xu, Y. Wu, and X. Zhou, ''Link prediction in social networks: The state-of-the-art,'' *Sci. China Inf. Sci.*, vol. 58, no. 1, pp. 1–38, Jan. 2015.

[34] I. Güneş, C. S. Gündüz-Öğüdücü, and Z. Çataltepe, ''Link prediction using time series of neighborhood-based node similarity scores,'' *Data Mining Knowl. Discovery*, vol. 30, no. 1, pp. 147–180, 2016.

[35] O. J. Mengshoel, R. Desai, A. Chen, and B. Tran, ''Will we connect again? Machine learning for link prediction in mobile social networks,'' in *Proc. 11th Workshop Mining Learn. Graphs*, 2013.

[36] Z. Liu, Q.-M. Zhang, L. Lü, and T. Zhou, ''Link prediction in complex networks: A local Naïve Bayes model,'' *EPL (Europhys. Lett.)*, vol. 96, no. 4, p. 48007, 2011.

[37] K. Yu, W. Chu, S. Yu, V. Tresp, and Z. Xu, ''Stochastic relational models for discriminative link prediction,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1553–1560.

[38] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, ''Link prediction using supervised learning,'' in *Proc. Workshop Link Anal., Counter-Terrorism Secur. (SDM)*, vol. 30, 2006, pp. 798–805.

[39] F. Gao, K. Musial, C. Cooper, and S. Tsoka, ''Link prediction methods and their accuracy for different social networks and network metrics,'' *Sci. Program.*, vol. 2015, Jun. 2015, Art. no. 172879.

[40] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New perspectives and methods in link prediction," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2010, pp. 243–252.
[41] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the social network of scientific collaborations," *Phys. A, Stat. Mech. Appl.*, vol. 311, nos. 3–4, pp. 590–614, Aug. 2002.
[42] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 2, Jul. 2001, Art. no. 025102.
[43] P. Jaccard, "Étude comparative de la distribution florale dans une portion des alpes et des jura," *Bull Soc. Vaudoise Sci. Nat.*, vol. 37, no. 142, pp. 547–579, 1901.
[44] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social Netw.*, vol. 25, no. 3, pp. 211–230, 2003.
[45] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," *Eur. Phys. J. B*, vol. 71, no. 4, pp. 623–630, Oct. 2009.
[46] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.
[47] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–1555, Aug. 2002.
[48] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Effects of user similarity in social media," in *Proc. 5th ACM Int. Conf. Web Search Data Mining (WSDM)*, 2012, pp. 703–712.
[49] J. Valverde-Rebaza and A. de Andrade Lopes, "Exploiting behaviors of communities of twitter users for link prediction," *Social Netw. Anal. Mining*, vol. 3, no. 4, pp. 1063–1074, 2013.
[50] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, "A hybrid persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks," *Neurocomputing*, vol. 380, pp. 1–10, Mar. 2020.
[51] F. Jiang, B. Kong, J. Li, K. Dashtipour, and M. Gogate, "Robust visual saliency optimization based on bidirectional Markov chains," *Cogn. Comput.*, vol. 13, pp. 69–80, May 2021.
[52] M. Gogate, K. Dashtipour, A. Adeel, and A. Hussain, "CochleaNet: A robust language-independent audio-visual model for speech enhancement," 2019, *arXiv:1909.10407*. [Online]. Available: http://arxiv.org/abs/1909.10407
[53] D. Molina, J. Poyatos, J. D. Ser, S. García, A. Hussain, and F. Herrera, "Comprehensive taxonomies of Nature- and bio-inspired optimization: Inspiration versus algorithmic behavior, critical analysis and recommendations," 2020, *arXiv:2002.08136*. [Online]. Available: http://arxiv.org/abs/2002.08136
[54] S. Scardapane, M. Scarpiniti, E. Baccarelli, and A. Uncini, "Why should we add early exits to neural networks?" 2020, *arXiv:2004.12814*. [Online]. Available: http://arxiv.org/abs/2004.12814
[55] X. W. Jiang, T. H. Yan, J. J. Zhu, B. He, W. H. Li, H. P. Du, and S. S. Sun, "Densely connected deep extreme learning machine algorithm," *Cognit. Comput.*, vol. 12, no. 5, pp. 979–990, Sep. 2020.
[56] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve.," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982.

HAJI GUL received the B.Sc. degree in computer science, mathematics, and physics from the University of Peshawar, Pakistan, the bachelor of Education from Allama Iqbal Open University, Islamabad, Pakistan, and the M.S. degree in computer science and the M.Sc. degree in computer science from the Institute of Management Sciences, Hayatabad Peshawar, Pakistan. His M.S.C.S. thesis title was Link Prediction Using Node Information on Local Path. His M.Sc. project was Internet download application. Further, he excelled in a Diploma in information technology from 2014 to 2015 with information technology courses along with environmental Awareness and education from 2007 to 2008. He is currently a Lecturer with the Centre for Excellence in Information Technology, City University of Science and Information Technology, Peshawar, Pakistan. He expertise includes graph analysis, graph clustering, link prediction, data mining, and data analysis.

ADNAN AMIN received the M.Sc. degree in computer science from the University of Peshawar, Pakistan, in 2008, and the M.S. degree (Hons.) in computer science from IMSciences, Peshawar, Pakistan, in 2015, where he is currently pursuing the Ph.D. degree in data mining/machine learning with the Centre of Excellence in Information Technology. Previously, he was working for Maxwell Stamp PLC, London, as an International Consultant for curricular and academic development activities for the School of ICT, National Institute of Management and Administration, University of Jyvaskyla, Finland, under the World Bank funded project (P102573). He is currently a Lecturer with the Centre of Excellence in Information Technology, Institute of Management Sciences, (IMSciences). He is also a leading Expert in data mining, machine learning, and data science. He is supervising many M.S. research students in the area of data mining and machine learning. He has authored more than 22 research articles, including more than 12 journal articles. His research interests include cross-disciplinary and mostly applied industry-oriented including Churn prediction, Prudent-based expert systems, customer analytics and target marketing, active and adaptive learning, real-time classification and segmentation, cyber security, rough set theory, ripple down learning, oversampling, cognitive skills, and big data analytics. He has been the Track Chair of WorldCist'21 workshop (LCBUDAMLT) and a Program Committee Member in numerous conferences, such as WorldCist-15, WorldCist-16, and SDIWC and also active reviewers for more than 23 reputed journals. He received the Gold Medal for his M.S. degree from IMSciences. He has conducted and led software development projects and collaborative scientific research projects with academia and industry.

AWAIS ADNAN received the M.Sc. degree in computer sciences and the M.B.A. degree from the University of Peshawar, the M.S. degree in computer software engineering from the National University of Science and Technology Islamabad, and the Ph.D. degree from the Institute of Management Sciences Peshawar, where his area of research was content aware video system. He is currently working as an Assistant Professor and the Director Office of the Research, Innovation, and Commercialization, IM Sciences. His research interests include artificial intelligence, data engineering, data visualization, and multimedia systems.

KAIZHU HUANG (Member, IEEE) received the Ph.D. degree from the Chinese University of Hong Kong (CUHK), in 2004. He worked with the Fujitsu Research Centre, CUHK, University of Bristol, the National Laboratory of Pattern Recognition, Chinese Academy of Sciences, from 2004 to 2012. He is currently a Professor with the Department of Intelligent Science, Xi'an Jiaotong-Liverpool University (XJTLU), China. He also acts as the Associate Dean of research with the School of Advanced Technology, XJTLU, and is also the Founding Director of the Suzhou Municipal Key Laboratory of Cognitive Computation and Applied Technology. His research interests include machine learning, neural information processing, and pattern recognition.