Received 26 July 2024, accepted 19 August 2024, date of publication 27 August 2024, date of current version 10 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3450506

SURVEY

An Overview of Similarity-Based Methods in Predicting Social Network Links: A Comparative Analysis

SACHIN U. BALVIR[1,2], **MUKESH M. RAGHUWANSHI**[2], **AND PURUSHOTTAM D. SHOBHANE**[3]

[1]Yeshwantrao Chavan College of Engineering, Nagpur, Maharashtra 441110, India
[2]S. B. Jain Institute of Technology, Management and Research, Nagpur, Maharashtra 441501, India
[3]Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune 412115, India

Corresponding authors: Sachin U. Balvir (sachin_balvir@yahoo.com) and Purushottam D. Shobhane (purushottam.shobhane@sitnagpur.siu.edu.in)

ABSTRACT Link prediction in the Social Network is most important and an essential part now a days. The continued growth and evolution of this field will lead to new and improved methods for analyzing and understanding social networks. Link prediction is also helpful in various network applications in both academic and real-world contexts. For better understanding of prediction of links in a network graph through the use of different algorithms and information of prediction of missing link between network that all of the clear information is discuss in this paper. This paper presents the study of different types of algorithms which are better informative to understand the connection prediction, in a methodical manner. For this study, the similarity approaches are concentrated with its types of algorithms which are used to forecast the presence of missing links in social networks. This paper addresses the various link prediction approaches considering the structure of the network to reduce uncertainty. Evaluation measures for link prediction and their practical applications are also covered in this work. Lastly, it discusses the difficulties and provides plans for the development of link prediction methods in the future. This discussion may help researchers to choose the proper network structure for predicting the links.

INDEX TERMS Link prediction, social networks, network analysis, similarity based methods.

I. INTRODUCTION

At present Social Network is converting to a crucial part of our life. Due to large amount of used social network, it makes available the different challenging and best platforms and facility with secure system. In social network there are some of the social functions like sharing images, videos, text, and also sharing opinion in the form of comments [1]. In different types of areas, large number of data produced and that data can manage by the system. This information can be seen as a network and is expressed in nodes, connecting edges. Node and edge represent entity and relationship between those entities. It means there is large composite network. Therefore there is a future process of link prediction between that nodes and edges [2]. The prediction of word-to-document relationships is another way to think about link prediction [3], [4].

In social network, the large amount of information is produced by the users, with the help of this information attracted parties know how to examine, determine and even predicting the things regarding a user or grouping of user. There are lots of reasons for extracting the meaningful information from the data such as to make a decision for progress of company or to diminish the future risk. This information facilitates associate to improve the strength and design of numerous applications, such as making recommendations for new connections, predicting the evolution of the network structure over time, and detecting potential fraud or anomalous behavior. On the social networks many users make their personal

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

information public that have actual economic value to the companies [5], [87].

The expansion of internet social networks in recent years has led to an explosion of data on social connections, providing a wealth of information for link prediction research. Link prediction in social networks is a difficult job because of the complexity of the relationships between individuals in a network. Relationships can be influenced by many factors, such as demographics, shared interests, and past interactions. Additionally, the structure of social networks is constantly changing, making it difficult to predict the formation of new connections. Despite these challenges, link prediction has become an active area of research, with numerous methods and techniques developed to tackle the problem.

The majority of user's preferred social networking sites are Facebook. Compared to different social networking sites, it has 2,910 million active users each month. Whatsapp is the third most popular platform with 2,000 million monthly active users. With 346 million active users per month, Twitter is the fifteenth-popular platform. That is all there needs to be said about the significance of social networking and its uses.

According to statista, the top social networks globally in terms of monthly active users (in millions) as of January 2022 are as follows in Table 1:

The current study examines the forecast of links in social networks based on similarity. Despite the fact that this research focuses on social networks; it can also be applied to other kinds of networks. Following is a summary of what this article contributes:

- In this article, we will look at the most recent advancements in the area of similarity-based link prediction.
- We analyze various methods applied on type of link structure also given their complexity.
- To help new researchers come up with fresh research ideas, we will propose future directions for link prediction in social networks.

The remainder of this paper is divided into the following sections: We describe the relevant study papers that were used in this paper in Section II. We provided a short overview of the link prediction Problem in Section III. In Section IV, the similarity-based link forecast techniques for local, global, and quasi-local link structures are covered. In section V, we describe the way that how these similarity-based link prediction methods will help in link prediction. The evaluation metrics and applications in real word scenario are discussed in section VI and VII respectively. The upcoming research in this field is described in Section VIII. The summary of the paper is provided in Section IX, and Section X concludes the research.

II. RELETED WORK

Many different approaches to the link prediction problem have been explored in the literature. These approaches can be divided into a number of categories, including similarity-based [6], probabilistic [7] and learning-based [8], and others. Similarity scores between node pairs are computed using structural characteristics of the underlying network in similarity-based methods. These characteristics are simple to calculate because they are directly extracted from the network structure. There are two types of topological properties: 'local' which is derived from a restricted area of network structure and 'global' taken from the whole network. Many link prediction methods are created using this knowledge. Compared to global approaches, which are more complicated and have higher accuracy for prediction, local approaches are simple to calculate.

It's possible that the quantity of shared neighbors isn't always enough to reveal all of the internal similarities between two nodes. A new measure is suggested to address this issue that relies on the similarity of each pair of nodes on the number of shared neighbors and correlation between the nodes' neighborhood vectors. Comparable techniques are less accurate than this one [6]. The SHOPI similarity index, which attempts to stop information leakage by penalizing common neighbors [9]. A novel link-predicting algorithm that combines common neighbor and centrality is proposed [10]. In a different one, proximity to a shared neighbor and distance are combined [11]. A technique using neural networks improved by link prediction models as output data and scale-free networks as input data for training. A greedy link pruning approach is used to address the impact of the neural network's [12].

NodeSim learns the low-dimensional model of a network while capturing similarities among the nodes and the group's structure. The suggested NodeSim random walk effectively explores every part of the region while maintaining the more similar nodes nearby in the context of the node, allowing for the learning of the embedding [13]. The goal of a new similarity measure for link prediction in bipartite networks is to offer a centralized and all-encompassing approach. A combination of criteria depending on neighborhood structure makes up the suggested method [14]. Utilizing friend-based and route-based similarity factors, both of which make use of graph structures, three novel similarities are introduced: degree neighbor similarity (DNS), path neighbor similarity (PNS), and degree path neighbor similarity (DPNS) [15].

Two algorithms, DLP-ILS based on Improved Latent Space and DLP-IRA based on Improved Resource Allocation, are proposed in [15]. In this work, the link prediction processes for a pair of nodes that do not have common neighbors are still conducted in serial instead of in parallel due to the fact that they are based on the nodes adjacency relationship. To increase the algorithm's effectiveness in the initial computation, non-interfering nodes might be discovered for parallel computing.

The article presents AdaSim, a novel framework for link prediction that makes use of attributes derived from random walk-based network embedding. By fine-tuning a parameter depending on the data distribution through supervised learning, the Adaptive Similarity function in AdaSim provides a layer of flexibility and makes the framework robust and

TABLE 1. Monthly active users on social networking sites globally as of January 2022 (https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/).

Social Network Site	Company	Active monthly Users (in millions)	Usage
Facebook	Meta Platforms	2,910	Social Networking, Media Sharing
Youtube	Alphabet Inc.	2,562	Media Sharing
WhatsApp	Meta Platforms	2,000	Instant Messaging
Instagram	Meta Platforms	1,478	Social Networking, Media Sharing
Weixin/WeChat	Tencent	1,263	Instant Messaging
TikTok	ByteDance	1,000	Media Sharing
Facebook Messenger	Meta Platforms	988	Instant Messaging
Douyin	ByteDance	600	Media Sharing
QQ	Tencent	574	Instant Messaging
SinaWeibo	Sina Corp	573	Micro blogging
Kuaishou	Kuaishou	573	Social Networking, Media Sharing
SnapChat	Snap, Inc.	557	Social Networking, Media Sharing
Telegram	Telegram	550	Instant Messaging
Pinterest	Pinterest	444	Media Sharing
Twitter	Twitter, Inc.	436	Micro blogging
Reddit	Reddit	430	Social News
Quora	Quora	300	Social Networking,

flexible in many network scenarios [16]. In situations when the network topology is tree-like, the research tackles the problem of link prediction and offers a comparison analysis to identify the best method. The paper illustrates the efficacy of the suggested approach in precisely predicting linkages inside tree-like network architectures through experimental validation and performance evaluation [17]. The research addresses the shortcomings of conventional algorithms by introducing a novel local link prediction technique created especially for long-line and closed-circle networks. Through testing on real-world networks such as a metropolitan water distribution network and a sexual contact network, it shows that the suggested method is effective, beating conventional local and global algorithms [18].

The use of direct links in link prediction algorithms was suggested by the author. Demonstrates how employing direct ties yields better prediction results than indirect links in dynamic networks. Overall, the results showed how crucial it is to take into account direct links, network topology, and time when doing link prediction tasks [19]. In order to forecast link directions in directed networks, the research study suggests a technique called Link Direction for Link Prediction (LDLP). LDLP provides a more thorough understanding of network dynamics and behaviors by integrating link direction into the prediction process, especially in situations where knowing the direction of connections is crucial. The efficiency of LDLP in capturing link directionality and enhancing prediction accuracy is demonstrated [20]. The hormonal imbalance dataset has been exposed to a novel fuzzy data transformation technique by the author. By using machine learning on fuzzy modified biological datasets, a wider range of diagnoses can be achieved, including a third class that suggests PCOS. In the end, this would alert the patient to the need for preventive measures to lessen the chance that the illness may recur [21].

The author defined community dynamicity, an actor-level metric that takes changes in neighborhood infrastructure, community engagement, and cliquishness into account to reflect the temporal evolution of community awareness [22]. The research shows how well community-level knowledge may be used to sophisticated machine learning methods to gain a deeper comprehension of the dynamics of collaboration and citation in scientific literature. The results of this study imply that sophisticated machine learning methods can be used to detect hidden patterns in scientific data, giving academics, organizations, and decision-makers insightful information [23]. The research presents a DSLP, the model incorporates a number of community structures and topology variables into a probability model [24].

The author presented a method for predicting links in complex networks by utilizing average centrality metrics, including clustering coefficient, betweenness, closeness, and degree centralities. When compared to conventional similarity metrics, the technique showed an average 24% improvement in AUROC [25]. LEHMN (learning embedding based on hyper-motif of the network) model seeks to capture subtle similarities between nodes that conventional models could overlook. The effectiveness and superiority over cutting-edge methods are further validated by numerical simulations, underscoring the need of implementing such creative frameworks in complex network analysis [26]. Three new iterations of the 3-hop path, quasi-local path, and global path of the Common Neighbor and CCPA were proposed, considering varying path lengths. This improved the prediction accuracy in complicated networks by introducing four new link prediction algorithms based on community detection data [27].

The research suggests a unique method for link prediction in multiplex networks that makes use of GNN. The Study presents as approach based on calculating inter-layer similarities and aggregating adjacency matrices to address the difficulties brought on by the intricate and multi-layered nature of multiplex networks [28]. The another study presents Graph Conversion Capsule Link (GCCL), a unique link prediction technique that combines GNNs and CapsNets to solves the link prediction task as a graph classification issue [29]. The paper's novel PS2 architecture and experimental findings highlight the significance of tailored sub-graph selection in improving link prediction tasks using GNNs. The inductive process of choosing sub-graphs without access to inference edges during training and the exponential growth in sub-graph selection space for various edges are challenges that the PS2 framework tackles [30]. NMFLP can be used with a variety of networks, regardless of whether or not they include node semantic properties. In networks when node characteristics are present, it employs these attributes directly to forecast links; in the absence of attributes, it substitutes the network's structure for the attributes [31].

The article presents GCCDC as a workable metric for assessing complicated networks and deriving meaningful conclusions from empirical data, greatly expanding the body of knowledge on network analysis. GCCDC or GD has a positive correlation with Betweenness Centrality (BC) in huge network analysis, suggesting that it can be used to gather important nodes in a network [32]. In social network analysis, algorithmic fairness is addressed by the FairSNA framework, which focuses on a topic that has received less attention: reducing structural bias and inequality in large-scale social networks. The NodeSim random walk technique, which incorporates community information into the model, is presented by the authors. It improves link prediction accuracy by examining a node's intra- and inter-community neighborhoods based on structural similarities [33]. In order to forecast graph links, the research presents a novel approach termed Conditional Diffusion-based Multi-level Negative Sampling (DMNS). In order to produce negative nodes in several degrees of varied hardness and reconcile them for efficient graph link prediction, DMNS takes advantage of the Markov chain aspect of diffusion models [34].

The paper tackles the challenging subject of link prediction in social networks by utilizing machine learning and Node2Vec to uncover hidden relationships and predict future connections. Along with investigating the relationship between community evolution and factors such as the underlying social network structure, the study focuses on optimizing node network neighborhoods through the use of Node2Vec node mapping [35]. To improve link prediction performance in dynamic networks, the proposed framework presents a novel feature set that incorporates quasi-local, global, local, and community information-based characteristics. In this, the author uses four different machine learning models NN, XGB, LDA, and RFC to evaluate the suggested COMMLP technique in conjunction with three cutting-edge algorithms [36].

III. LINK PREDICTION PROBLEM

In the social network, the problem of predicting the links is understood as a potential or realizable link between edges or nodes [37]. Given a snapshot of a graph 'G', the link prediction problem predicts the edge or link which can be formed between two nodes in near future in a network. A social network can be view as a graph G (V, E), where set of nodes represented by V and connections between these nodes represented by E. Consider an example that in a below (figure 1) snapshot of graph taken at time t1, shows the five nodes namely Sachin, Arvind, Nitin, Santosh and Amol. The connection between them indicates the friendship.

The architecture of social networks from a time period t1 to t2 is shown in Figure 1. At time t1, black color line indicates the friendship exists between two nodes and dashed blue color line indicates that the friendship may occur in future. The objective of the link prediction problem to find out which two nodes which are unconnected friends at time 't1' will become connected friends in near future. It may happen that all unconnected friend becomes connected or some of them may be connected. A snapshot is taken after some time 't2'. This graph shows that out of five unconnected links only one links is connected in future i.e. at time t2, represented by blue color line between nodes Sachin and Amol. Computing or predicting such type of links is not an effortless assignment. Numerous similarity based approaches available for link prediction are clear formed on environment to calculate it, as local, global & quasi local structure [38], [39].

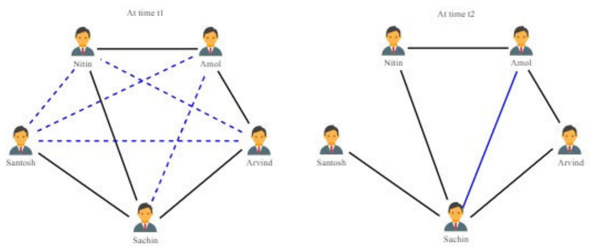


FIGURE 1. Graphs at time intervals t1 and t2.

Link prediction have a discussion about forecasting the chance that two nodes will connect which are not yet been associated in the network from side to side known nodes and structural information [3], [12]. In the undirected graph (network) shows G (V, E) where V and E symbolize sets of node and link, correspondingly. Link prediction method requires providing a keep score (S) to estimate the likelihood of an active link between two disconnected nodes. For example, S_{XY}^{CN} representing the score of probable links among two network nodes x and y calculate by the Common Neighbors (CN) model. All links are not generate in the native network, which are sort in lessening order according to their scores, and the link at the peak are mainly probable to be present [12].

Link prediction is the process of determining how likely it is that two nodes in a network will connect or create a connection. The prediction score, which is often referred to as the "link prediction score," measures the probability. Several similarity metrics or techniques that evaluate the probability or strength of a connection between two nodes can be used to compute this score.

Example Calculation:

Using the Common Neighbors approach as an example, let's forecast links:

- **Step 1**: Identify the neighbors of nodes x and y:
  - neighbors of node x i.e. $\Gamma(x) = \{a, b, c\}$
  - neighbors of node x i.e. $\Gamma(y) = \{b, c, d\}$
- **Step 2**: Find the common neighbors:
  - $\Gamma(x) \cap \Gamma(y) = \{b, c\}$
- **Step 3**: Count the common neighbors:
  - Score for node x and y i.e. $S(x, y) = |\Gamma(u) \cap \Gamma(v)| = 2$

Based on the Common Neighbors technique, nodes x and y have two common neighbors, as indicated by their score S of 2. Similar to that, we can use this method to compute scores for every pair of nodes in the network. We can determine which node pairings have the highest scores and the best likelihood of connecting in the future by ranking these pairs of nodes according to their scores. Certain link prediction techniques, such as Common Neighbors (CN), assign a score to each pair of nodes and then rank the scores downward to determine which connections are most likely. Alternative techniques, such as the Adamic/Adar index (AA), identify an index for every pair of nodes and also arrange these indices in a descending order. Alternative methods allocate a rank to every pair of nodes, wherein the ranks are pre-arranged in a descending sequence. It is anticipated that the node pair with the greatest score, index, or rank will have the highest likelihood of connecting in the future.

All possible links in a network can have their scores computed and compared by using different link prediction techniques. Accurate forecasts of future connections are made possible by these scores, which aid in ranking the relationships according to their chance of creation. We have described the similarity measure-based link prediction techniques in Section IV.

IV. LINK PREDICTION METHODS

In link prediction, there is already defined number of methods which are used for predicting the link. This work is solely concerned with similarity-based techniques to link prediction. Depending on the context in which the likeness score or index is computed, similarity-based techniques are further divided into three types [37]:

- Similarity-based methods on local link structure
- Similarity-based methods on global link structure
- Similarity-based methods on Quasi-Local link structure

A. SIMILARITY–BASED LINK PREDICTION METHODS ON LOCAL LINK STRUCTURE

To establish a forecast between the nodes of a link, the similarity-based on local techniques utilize node information. A basic calculation is used in the local similarity-based strategy and it calculate with a more rapidly solution [40]. There are several methods which already exist. Few methods are discussed below:

1) COMMON NEIGHBORS (CN)

Common neighbors receive the numeral of common neighbors as a keep score to come to a decision an association or not a relation can be produce involving two nodes. A node's likelihood of linking in the future increases with the number of similar neighbor's it has. For example, such a nodes X, Y will connect in future or not will be decided by the score of common neighbor. The common neighbors' set is defined as [12] and [41],

$$S^{CN}{}_{XY=\lceil(X)\cap(Y)\rceil} \tag{1}$$

2) JACCARD COEFFICIENT (JC)

A method for comparing two items is the Jaccard Coefficient, usually referred to as the JC index. Typically, it is represented as (X, Y), where X and Y stand for two network nodes [42]. And is defined as [38],

$$S_{XY}^{JC} = \frac{|(X) \cap (Y)|}{|(X) \cup (Y)|} \tag{2}$$

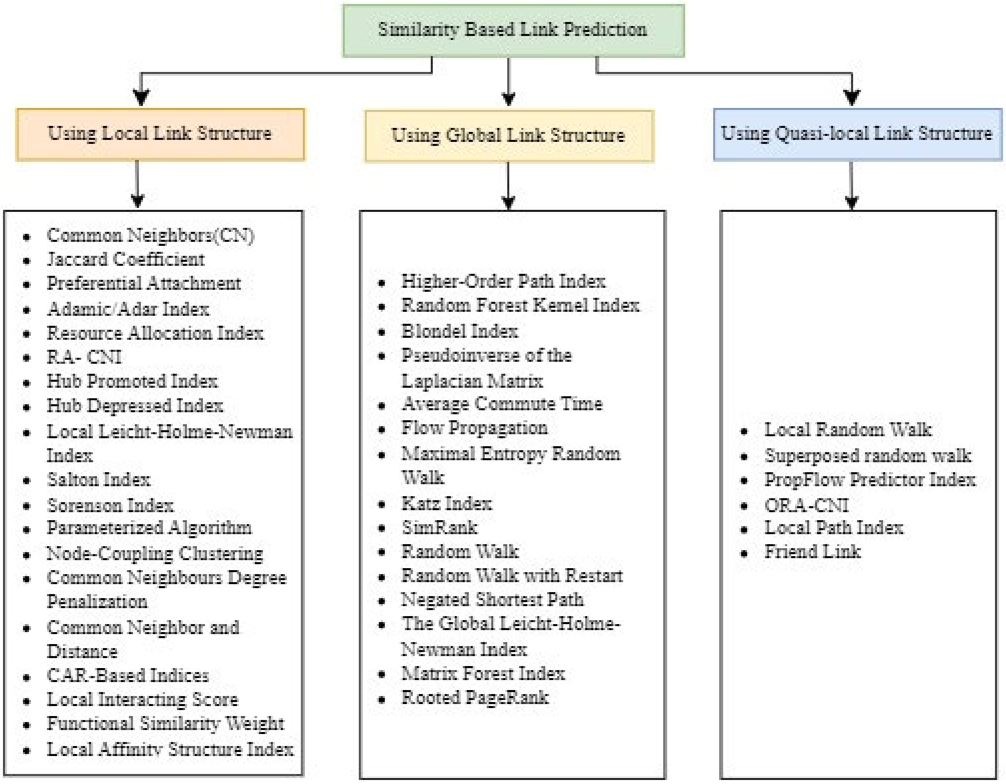


FIGURE 2. Various link prediction methods based on link structure.

3) PREFERENTIAL ATTACHMENT (PA)

Score SXY in the PA Index is calculated using the degrees of nodes x and y [43].Preferential Attachment Index is a comparison score that is thought to be self-sufficient of every node's neighborhood [37]. Community networks grow as new nodes join and join with active nodes that have higher degree evaluations than nodes with lower degree and it is define as [3] and [6],

$$S_{XY}=K_X.K_Y \tag{3}$$

4) ADAMIC/ADAR INDEX (AA)

Adamic/Adar Index [44] is a similarity compute, originally discovered by Lada Adamic and Eytan Adar; it is used to compute similarity among the nodes according to their feature of sharing. It is calculated as [45],

$$AA\,(a,b) = \sum_{(i\in\lceil(a)\cap(b)\rceil)} \frac{1}{\log\lceil(i)\rceil} \tag{4}$$

5) RESOURCE ALLOCATION INDEX

The dynamics of resource allocation in complex networks promote resource allocation [39], [46]. Let's understand, for instance, consider X and Y are two unconnected nodes. The common neighbor (CN) of node X serves as the transmitter as it transmits a small amount of resources to node Y [38].

The RAI is calculated as:

$$S_{XY}^{RA} = \sum\nolimits_{Z\in\lceil(X)\cap\lceil(Y)} \frac{1}{K_z} \tag{5}$$

6) RESOURCE ALLOCATION BASED ON COMMON NEIGHBOUR INTERACTIONS (RA- CNI)

The method of allocating resources supports the use of shared neighbor communications by having each node send a resource component to its neighbors [45]. On the other hand, this method also considers resource return in the opposite direction. The RA-CNI index is defined as,

$$S\,(x,y) = \sum_{z\in\lceil x\cap\lceil y} \frac{1}{|\lceil z|} + \sum_{e_{i,j}\in E,|\lceil i|<|\lceil j|,i\in\lceil x,j\in\lceil y} \left(\frac{1}{|\lceil i|} - \frac{1}{|\lceil j|}\right) \tag{6}$$

7) HUB PROMOTED INDEX (HPI)

The ratio of common neighbors to the network's lowest degree of nodes for x and y is referred to as the hub promoted index [47]. In order to measure the topological partially coverage of substrate pairings in metabolic networks and the similarity of HPI,

$$S_{xy} = \frac{|\lceil(x)\cap\lceil(y)\,|}{min\left\{k_x, k_y\right\}} \tag{7}$$

Due to the possibility of high scores being assigned to links near the hub based on the aforementioned metrics, the denominator is fixed at a modest degree only [38].

8) HUB DEPRESSED INDEX (HDI)

Using the hub promoted index as a base, the hub depressed index [47] is calculated. But have a reverse goal. This index is the ratio of common neighbors to the maximum degree of nodes of x and y in the network and the index similarity is defined as [45],

$$S_{xy} = \frac{|\lceil(x) \cap \lceil(y)\,|}{max\left\{k_x, k_y\right\}} \tag{8}$$

9) LOCAL LEICHT-HOLME-NEWMAN INDEX (LLHN)

It is the ratio of common neighbors to the degree of the x and y nodes as a product [48].This index is a much responsive measure of structural correspondence than Jaccard coefficient or Salton index. This index is defined as [45],

$$S_{xy} = \frac{|\lceil(x) \cap \lceil(y)\,|}{|\lceil(x)\lceil(y)\,|} \tag{9}$$

10) SALTON INDEX (SI)

Salton index [49]as well identify as the Salton Cosine index is employed to establish the cosine angle between rows of an adjacency matrix that takes nodes x and y [50]. The following formula is used to determine the Salton index [37]:

$$S_{xy} = \frac{|\lceil x \cap \lceil y|}{\sqrt{|\lceil x\,||\lceil y|}} \tag{10}$$

11) SORENSON INDEX

This index is proposed by Soreson [51] to compute the similarity between two species. Similar to the Jaccard coefficient, the Sorenson index, it is determined as the double common neighbours divided by the total of the degrees of nodes x and y, it is computed as,

$$S_{xy} = \frac{2\,|\lceil(x) \cap \lceil(y)\,|}{\|\lceil(x)\| * \|\lceil(y)\,\|} \tag{11}$$

12) PARAMETERIZED ALGORITHM

Parameterized Algorithm index [10], hold the numeral of common neighbor& the near-ness of two nodes is together occupied into relative to estimation the similarity among a pair of node. The score for similarity among x and y is computed as below equation, here between nodes x and y, d_{xy} is the shortest distance and α is a user-defined parameter [6].

$$S_{xy} = \alpha.\left(|\lceil(x) \cap \lceil(y)\,|\right) + (1-\alpha)\cdot\frac{N}{d_{xy}} \tag{12}$$

13) NODE-COUPLING CLUSTERING

To achieve the influence of every single common neighbor and likeness of each two nodes, this index uses clustering coefficient [6], [52]. Node-Coupling clustering is calculated as,

$$NCC_{ij} = \sum\nolimits_{v_n \epsilon \lceil i \cap \lceil j} \frac{\sum_{v_z \in CN_n^{(2)}}\left(\frac{1}{d_z} + C_z\right)}{\sum_{v_w \in \lceil n}\left(\frac{1}{d_w} + C_w\right)} \tag{13}$$

14) COMMON NEIGHBOURS DEGREE PENALIZATION

In this method, penalization of common neighbors is reflected. For this, the number of common neighbors for every pair of the two nodes' common neighbors is occupied into relation [6], [53].

$$CN_z^{(2)} = \{\lceil z \cap \lceil i \cap \lceil j\,\} \cup \left\{v_i, v_j\right\} \tag{14}$$

$$CNDP_{ij} = \sum_{v_z \in \lceil i \cap \lceil j} \left|CN_z^{(2)}\right|\left(d_z^{-\beta C}\right) \tag{15}$$

15) COMMON NEIGHBOR AND DISTANCE (CND)

This depends upon the two basic components that are common neighbor and distance. For calculating similarity score is shown below. Keep in mind that Γ(x) refers to a node's neighbours, CNxy is the quantity of an ordinary node between nodes x and y, and (dxy) is the separation between nodes x and y [10], [11].

$$S_{xy} = \begin{cases} \frac{CN_{xy}+1}{2} & \lceil(x) \cap \lceil(y) \neq \emptyset \\ \frac{1}{d_{xy}} & otherwise \end{cases} \tag{16}$$

16) CAR-BASED INDICES (CAR)

CAR-based indices are predicated on the idea that two nodes are more probable to be linked if their shared neighbors are individuals who are a part of a very close-knit group, called a local community (LC) [54]. By using this claim, we can give extra relevance to the neighboring nodes that are related to one another. A CAR-based description of common neighbors is described as [45].

$$s\,(x,y) = \sum\nolimits_{z \in \lceil x \cap \lceil y} 1 + \frac{|\lceil x \cap \lceil y \cap \lceil z|}{2} \tag{17}$$

As per the above calculation, one can compute a CAR-based deviation of the resource allocation as,

$$s\,(x,y) = \sum\nolimits_{z \in \lceil x \cap \lceil y} \frac{|\lceil x \cap \lceil y \cap \lceil z|}{\lceil z} \tag{18}$$

17) LOCAL INTERACTING SCORE (LIT)

Functional similarity weight in iterative form is used to calculate the local interacting score [55]. Originally, weights are assigned as $S^{xy}(0) = 1$ for linked pair of node and $S^{xy}(0) = 0$ for pair's respite. Hence, weights are iterated as,

$$s^{x,y}(t) = \frac{\sum_{u \in \lceil x \cap \lceil y} s^{z,x}(t-1) + \sum_{v \in \lceil x \cap \lceil y} s^{z,y}(t-1)}{\sum_{u \in \lceil x} s^{z,x}(t-1) + \sum_{v \in \lceil y} s^{z,y}(t-1) + \lambda(x) + \lambda(y)} \tag{19}$$

The computation of λ(x) is,

$$\lambda(\mathrm{x}) = \max(0, \frac{\sum_{u\in v}\sum_{v\in\lceil u} s^{u,x}(t)}{|V|} - \sum_{z\in\lceil x\cap\lceil y} s^{z,x}(t-1)) \tag{20}$$

In the functional similarity weight, λ(x) serves as a λ penalizing factor [45].

18) FUNCTIONAL SIMILARITY WEIGHT (FSW)

Given that in a directed network, the likelihood of a linking node 'x' to 'y' is independent of the likelihood that the 'y' would interact with 'x', the Functional Similarity Weight is a close analogue to the Sorensen index [56]. Nevertheless, this maintain score can also be used with an undirected network as [45],

$$s(x, y) = \left(\frac{2\,|\lceil x \cap \lceil y|}{|\lceil x - \lceil y| + 2\,|\lceil x \cap \lceil y| + \lambda}\right)^2 \tag{21}$$

When one of the nodes has a lower degree than the other, the similarity between the two nodes is penalized by integrating this parameter.

19) LOCAL AFFINITY STRUCTURE INDEX (LAS)

The closeness of two nodes to one other's neighbours is displayed by the LAS Index. According to the theory, a link is more likely if two nodes have a stronger attraction for each other and their common neighbor's [57], [58], that is:

$$S^{LAS}_{(v_x,v_y)} = \frac{|\lceil(v_x)\cap\lceil(v_y)|}{|\lceil(v_x)|} + \frac{|\lceil(v_x)\cap\lceil(v_y)|}{|\lceil(v_y)|} \tag{22}$$

B. SIMILARITY-BASED LINK PREDICTION METHODS ON GLOBAL LINK STRUCTURE

In order to determine the score of each connection, global similarity-based techniques employ the data from the entire topological network. These techniques do not just estimate distance between two nodes that are comparable. Diverse from local-based approach, global-based methods utilize the topological data of the complete network to arrange the node pairs [59]. However, because of their computational complexity, they can be impracticable for huge networks, and especially in distributed environments, its parallelization might be rather challenging where not every computational representative would necessarily be familiar with the network's whole structure.

1) HIGHER-ORDER PATH INDEX

Higher-Order Path index is depending on shared neighbors, and it proposes an iterative procedure by taking into consideration the significance of pathways among two nodes. The chance of a connection between the two nodes is determined by adding up the importance of the pathways between them. For this reason, the recommended path's length between nodes x and y is represented with the expression below [9].

$$s_{xy} = \sum_{v_n\epsilon\lceil x\cap\lceil y} \frac{1}{d_z} \tag{23}$$

With the following expression, we can determine the significance of the space last part to last part l > 2 path from x to y nodes depending on the significance of its component edges.

$$s_{ij} = \sum_{k=3}^{l-2} f1.f2.\alpha^{l-2}, \tag{24}$$

where f1 and f2 represent the importance of the component edge and the preceding iteration's path, respectively, and where α denotes an adjustable feature.

2) RANDOM FOREST KERNEL INDEX (RFK)

A linked, undirected sub-graph of G having all the vertices, some or all of the borders, and no cycle is defined in this Random Forest Kernel Index as a spanning tree of that graph. According to the matrix-tree theorem and Shamis, each cofactor of an access to G's Laplacian demonstration is equal to G's total number of spanning trees.

A cofactor is the main factor in the matrix that results from deleting a specific element's row and column. The collection of trees with displaced roots spreading over them is known as a root forest. The cofactor of (I+L) may be proven to be equal to the number of spanning rooted forests where x and y are found in the same x-rooted spanning tree [60].To determine how simple it will be to go from point x to point y, use the inverse of this value. Therefore, a similarity metric is described as,

$$S = (I + L)^{-1}$$

According to this similarity matrix, S (x, y) = Sxy describes, how similar two nodes are.

3) BLONDEL INDEX (BI)

The initial purpose of the Blondel index was to determine whether two vertices in two different graphs are comparable. However, ACM may be modified to operate in an accurate graph. Iteratively, it is constructed as S(t) [45]([61]

$$S(t) = \frac{AS(t-1)A^T + A^T S(t-1)A}{\left\|AS(t-1)A^T + A^T S(t-1)A\right\|_F}, \tag{25}$$

where S (0) = I and ‖M‖ F is the Frobenius matrix norm. The measure is computed iteratively in this index, same like in random-walk-based techniques. The matrix's Frobenius norm is calculated as,

$$\|M_{m\times n}\|\,F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}(M_{i,j})^2} \tag{26}$$

TABLE 2. Comparison of various methods considering local link structure.

Article	Key point	Method	Dataset	Advantages/ Applications	Drawback
[1]	In dynamic network, link forecast is anticipated to foresee the incident probabilities.	Common neighbor's (CN)	DBLP	Consider only the common neighbors in the one hop because CN and the enhanced methods depend on CN.	Common neighbors are not considered in the mutual relationship.
[42]	The Jaccard index was created for the presence-absence of data, giving equal weight to each species (attribute's) importance.	Jaccard coefficient	Interaction Web Database	The coefficient is used to compute interaction likeness in pollination networks as proposed by Olesen & al. (2007).	By creating a simplex diagram based on the Jaccard coefficient adjusted to include taxonomic functions, you can combine the two strategies.
[43]	The large scale features of complex networks are characterized by a high degree of self-organization.	Preferential Attachment	WWW or the citation patterns in science.	The node having higher degree will increase its connectivity or attracts other nodes at a higher rate.	All facets of the networks cannot be calculated using the proposed approach.
[44]	Determine whether a person is connected to another, and if so, rank all users according to how similar they are to that person.	Adamic/Ada r Index	MIT, Stanford	By examining the co-occurrence of terms connected with each cluster on homepages, it is possible to draw correlations between groups of individuals.	Makes a whole evaluation challenging, as it measures precision-recall rely on a complete data set.
[39]	If only consider the nearest neighbor then various links are allotted the similar prediction score.	Resource Allocation	PPI, NS, Grid, PB, INT, USAir	Resource allocation is similar to Adamic Adar. Works better for the networks having high average degrees.	A network with greater degree heterogeneity typically exhibits disassortativity because the assortativity coefficient is very sensitive to the degree sequence.
[48]	The widely accepted notion that the two nodes are similar if they have same close neighbors in the network.	Leicht-Holme-Newman Index (GLHN)	AddHealth, Roget's Thesaurus,	Based on network architecture, this measure can extract important data about vertex similarity.	This method will not applicable if nodes have inclination to link to dissimilar nodes.
[10]	The user-defined value of the parameter α ∈ [0, 1] influences the significance of the centrality and shared neighbour	Parameterize d Algorithm	Karate, USAir, Dolphins, Polbook, Word, Neural, Circuit , E-mail	They look for a statistical property that might determine the ideal value of α in an effort to increase the applicability of the suggested method in actual world.	AUC values presented differ slightly from those reported in the literature.
[52]	Approaches combine 1) The degree of coupling between common neighbor nodes and a network's clustering information and 2) The various node functions for link prediction.	Node coupling clustering	S D(2,100), SD(6,300), Jazz, PBl, Alines, Facebook, Karaclub, PtoP	This index is appropriate for massive networks.	The tested dataset's size expands together with its running duration, but the absence discernible variation in the durations of the numerous prediction approaches for various ratios of recognized edges..
[53]	Network structural property Clustering Coefficient" is used to propose new measure.	Common Neighbours Degree Penalization	BUP , CEG UAL,INF SMG, NSC and LDG EML ,HMT ADV ,PGP	Obtain high performance, superior to similarity-based approaches, with incredibly low computational cost, hence reduce run-time in benchmark datasets.	The absence of consideration for the form of common neighbors is this method's main flaw. Incorporated on a small dataset.
[54]	When the network's community structure is intact, the new CAR-based family of link predictors can be used to produce predictions that are more accurate.	CAR Based Indices	Mouse-single Neuron connectomes, Macaque Cortial connectomes, C.elegans single Neuron connectomes, American Football Games, Dolphin association, Zachary Karate club.	In many complex networks, link prediction performance has increased dramatically.	A small number of distinctive non-hub nodes centered on local modules are required for the local processing of the overall function.
[11]	Combines distance measure with Common Neighbors.	Common neighbors and distance (CND)	Power, Circuit Email, Karate, Dolphins, Polbook, Word, Neural	Boost prediction accuracy and identify missing linkages between nodes that lack neighbors in common.	In some of the dataset it calculates the low accuracy compared to different method.

TABLE 2. ***(Continued.)*** **Comparison of various methods considering local link structure.**

[55]	Protein pair topological data from both the local and global scales is used in the method.	Local Interacting Score	DIP (yeast interaction)	Detected a significant performance improvement.	Iterative approach used to calculate local interacting scores.
[56]	Test to a number of well-known existing techniques to get an idea of how functional likeness weighted averaging performs.	Functional Similarity Weight	GRID interaction dataset and MIPS	Considering depth in neighbors in prediction function, also makes it more likely to include inaccurate interaction data	False positives cannot be sufficiently reduced to make them effective for function prediction
[57]	Considering the deficiency of common neighbor algorithm.	Local Affinity Structure Index	C.elegans, Polblogs, Football, Polbook, Karate, Hep, Metabolic, USAir, Netscience.	Greater attraction between two nodes and their neighbors increases the likelihood that a link will emerge between them. .	Not consider the structural relationship between two nodes. Instead, only take into account the quantitative relationship of their common neighbor's.

4) PSEUDO INVERSE OF THE LAPLACIAN MATRIX (PLM)

By utilizing the Laplacian matrix $L = D-A$ instead of the adjacency matrix A, a substitute graph example is provided, where D is the diagonal matrix with most degree points [62]($D_{i,j} = 0$, and $D_{i,i} \sum_j A_{i,j}$). The Laplacian matrix's Moore-Penrose pseudo-inverse, denoted by the sign L^+, can be used to calculate the closeness process [63]. As a result of the fact that the literature identifies the pseudoinverse of the Laplacian matrix as "cosine similarity time",for the reason it is determined to be an internal product cosine similarity,& it is computed as [58] and [64],

$$S^{PLM}_{(v_x,v_y)} = \frac{L^+_{(v_x,v_y)}}{\sqrt{L^+_{(v_x,v_x)} L^+_{(v_y,v_y)}}} \tag{27}$$

5) AVERAGE COMMUTE TIME (ACT)

The distance travelled by random walker starting from node x to y and return from node y to x is utilized to compute the average commute time [45], [63]. If m(x, y) represents the quantity of steps required to move at node y from node x. Between the two nodes, the average commuting time value n(x, y) can be represented as [65],

$$n(x, y) = m(x, y) + m(y, x) \tag{28}$$

Additionally, the average commute time can be determined using the pseudo inverse of the Laplacian matrix L.

$$n(x, y) = |E| \left(L^+_{x,x} + L^+_{y,y} - 2L^+_{x,y}\right) \tag{29}$$

6) FLOW PROPAGATION (FP)

While it is true that the vector of probability is what the random walk with restart technique correspond to, their iterative descriptions relate to a transmission procedure. The adjacency matrix uses substitute normalizations. As an example, Vanunu and Sharan designed to be legitimate Random Walk with Restart, which is calculated as follows, by using the normalize Laplacian matrix in place of the normalize adjacency matrix [66],

$$M = D^l A D^r \tag{30}$$

where D^l and $AD^r$are slanting (diagonal)matrix whose essentials are correspondingly definite as, $D^l_{i,i} = \frac{1}{\sqrt{\sum jA_{i,j}}}$ and $D^T_{i,i} = \frac{1}{\sqrt{\sum jA_{j,i}}}$.The computational complexity of this technique, matrix be calculate using every adjacency matrix's development access via a scalar rate.

7) MAXIMAL ENTROPY RANDOM WALK (MERW)

In the maximum entropy random walk, a node in an ordered network is the likelihood of being associated with the innermost (central) nodes. MERW processed integrate the centrality of nodes in instruct to representation the performance. In this procedure, the approach seeks to increase the walk's entropy time μ [67], which is denoted as,

$$\mu = \lim_{l\to\infty} \frac{-\sum_{path^l_{x,y} \in paths^l} p(path^l_{x,y}) Inp(path^l_{x,y})}{l} \tag{31}$$

where $(path^l_{x,y}) = M_{x,h}M_{h,i}\ldots M_{i,j}M_{j,y}$. In instruct to make the most of maximum the entropy, every constituent of the evolution matrix is calculate as,

$$M_{i,j} = \frac{A_{i,j}}{\lambda} \frac{\psi_j}{\psi_i}, \tag{32}$$

where λ is the major eigen value of the adjacency matrix and ψ is the regularize eigenvector with deference to λ satisfying $\sum_{x\in V} \psi_x^2 = 1$.

8) KATZ INDEX (KI)

In this, the Katz measure computes the set of every path linking pair of nodes and gradually procedure the path based on its path length [68]. To put it another way, the wide path is considered a light weight, whereas the tight path is called a heavy weight. This metric is described as [59],

$$\begin{aligned} s^{Katz}_{(x,y)} &= \sum_{i=1}^{\infty} \beta^i \left|paths^{\langle i\rangle}_{xy}\right| = \sum_{i=1}^{\infty} \beta^i \left(A^i\right)_{x,y} \\ &= \beta A_{x,y} + \beta^2 \left(A^2\right)_{x,y} + \beta^3 \left(A^3\right)_{x,y} + \ldots \end{aligned} \tag{33}$$

where $\left|paths^{\langle i\rangle}_{xy}\right|$ is a compilation of every one route of length i between nodes x and y. The damping factor ß (free parameter), protected the path weights (ß > 0).The Katz measure will be relatively close to Indicators of Common Neighbors (CN) measure if ß is very small since extended-length pathways contribute relatively little to concluding similarity.

9) SIMRANK (SR)

According to the SimRank description in this index, which is self-consistent [69], if two nodes are connected to other related nodes, then they are comparable. Then directed or diversified networks can use this strategy.

$$s(x, y) = \beta \frac{\sum_{i\epsilon\tau_x} \sum_{j\epsilon\tau_y} s(i, j)}{|\tau_x| \left|\tau_y\right|}, \tag{34}$$

Hence, s(z, z) = 1 and 0 < ß < 1 is the moulder factor. Because the SimRank uses a random walk method, S(x, y) is employed to explain how two elements are put together starting from the corresponding nodes x and y.

10) RANDOM WALK (RW)

Given a network and a starting node, we randomly choose some of the node's neighbor's to move around in the random walk index. We then carry out the procedure once more for each node that is received, therefore explicitly making the method a random walk on the network [38], [70].The likelihood of acquiring all vertices can be iteratively approximated by if we refer to p^x as the likelihood vector to a few nodes that embark on a random walk from node x. If px is the likelihood vector for a few nodes that start a random walk from node x, the possibility of receiving every vertex can be iteratively estimated by [59].

$$\overrightarrow{p^x}(t) = M^T \overrightarrow{p^x}(t-1) \tag{35}$$

11) RANDOM WALK WITH RESTART (RWR)

Random Walk Restart uses Page Rank's algorithmic facts to this technique [8], along with its hypothesis is that the random walk element proceeds to the primary point with a definite possibility each move. This representation is recognized as a Random Walk Restart [71]. It is as follows,

$$\overrightarrow{q^x} = aP^T \overrightarrow{q^x} + (1-a)\overrightarrow{e^x}, \tag{36}$$

where P is the conversion possibility matrix, while Pxy = 1/kx, if x and y are associated, and Pxy = 0. Or else, the clarification is clear-cut as [72],

$$\overrightarrow{q^x} = (1-a)\left(1 - aP^T\right)^{-1} \overrightarrow{e^x} \tag{37}$$

Then, the RWR index is in consequence define as [59],

$$s_{XY}^{RWR} = q_{xy} + q_{yx},$$

12) NEGATED SHORTEST PATH (NSP)

To compute the NSP [73], one must first calculate the shortest route connecting two related nodes, which is a crucial graph likeness computation. The shortest routes can be precisely determined using the dijkstra method. The following formula can be used to determine the similarity and shortest path between two nodes, x and y.

$$s(x, y) = -\left|shortest\ path_{x,y}\right| \tag{38}$$

Given that, for each node in the network, the shortest pathways should be calculated.

13) THE GLOBAL LEICHT-HOLME-NEWMAN INDEX (GLHN)

The Kartz Index additionally takes into account a node's high similarity if there are many pathways linking these linked nodes, the knowledge underlying GLHN is very similar to that of Kartz Index [48]. GLHN is computed as:

$$s^{\text{GLHN}} = \beta_1 (I - \beta_2 A)^{-1} \tag{39}$$

where $\beta 1$ and $\beta 2$ are unrestricted, a slightly lower value for $\beta 2$ takes into account increased relevance for the shorter paths [58].

14) MATRIX FOREST INDEX (MF)

Matrix Forest Index is described as [60],

$$S^{MF} = (I - L)^{-1} \tag{40}$$

where the ratio of the number of spanning deep rooted forests, namely nodes x and y, to the identical tree rooted at x to each spanning rooted forest of the network may be used to compare x and y [38]. A MFI variant that is depending on parameters is,

$$S^{MF} = (I + \alpha L)^{-1}, \alpha > 0.$$

The comparison between nodes on a combined recommendation job has been computed using this Matrix Forest index [74].

15) ROOTED PAGERANK

In this Rooted PageRank is an additional alternative of PageRank centrality, this is used to position the look for results. The position is determined on the random walk of node in the graph. Furthermore, feature γ correspond to the visit of initial node to its neighbors [73], [75].

Consider, D consist of diagonal values of adjacency matrix Am,

$$D_{i,j} = \sum_j Am_{i,j} \tag{41}$$

Therefore, Rooted PageRank is approximate as below in expression,

$$RPR = (1-\gamma)\left(I - \gamma D^{-1} Am^{-1}\right) \tag{42}$$

C. SIMILARITY-BASED LINK PREDICTION METHODS ON QUASI-LOCAL LINK STRUCTURE

The approaches used in the local similarity approach have a minimal downtime complexity. On the other hand, global similarity approach, methods are the opposite. As a result, more and more associates started to understand how to trade

TABLE 3. Comparison of various methods considering global link structure.

Article	Key point	Method	Dataset	Advantages/ Applications	Drawback
[9]	SHOPI attempts to stop information leakage through shared neighbors by punishing them. Longer pathways are punished by using higher-order paths as discriminating features.	Higher-Order Path Index	Social Networks (Facebook, Twitter, Dolphins, Political blogs), Citation Networks (Cora , Smagri), Collaboration Networks (Jazz , Ca-GrQc, Ca-HepTh, Netscience), Biological Networks (Celegansneural, Protein–protein interaction)	SHOPI is powerful than Katz and Local Path index (LPI).	Prediction accuracy affected, computational complexity increases considerably
[61]	First ten results are studied on four search terms that were chosen for their variety.	Blondel Index	Four query words- Webster dictionary	ArcRank provides better outcomes than for other words but is still not good than the two other methods.	Measures that take into account adjacency matrices that don't just have integer entries but also have arbitrary real ones.
[63]	No parameter adjustment is required for Laplacian pseudo-inverse in the direct approach.	Pseudo inverse of the Laplacian matrix and Average Commute Time	MovieLens Database	Used in relational databases that contains machine learning and pattern recognition tasks. Additionally used to determine word or document similarity	For huge Databases, it does not scale effectively.
[66]	Parameters: c -controlling the LR transformation, α – controlling the relative position of prior info, # of propagation repetitions employed.	Flow Propagation	PPI network, Human Protein Reference Database (HPRD)	Method attained the finest performance in ranking the accurate gene at top in 50.9% of the cases.	Parallel results were found in the tuning method for diseases with more than one known gene.
[67]	Decide that the longest well is where the entire stationary probability will asymptotically localize.	Maximal Entropy Random Walk	Square lattice with periodic boundary condition.	MERW maximizes global entropy of random trajectories	MERW needs the information of the entire system
[68]	The idea of "attenuation" in a chain link is required to construct appropriate weights.	Katz Index	a six-person group with a decision matrix	Benefit in calculations to take 1/a equal to an integer.	When 1/a is smaller than the biggest characteristic root of C, the method divides.
[69]	Proposed an algorithm that determines only the similarity of structural context.	SimRank	A corpus of scientific research papers from Research Index, Transcripts of 1030 under- graduate students at Stanford University.	Fixed-point algorithm for calculating SimRank scores, as well as methods to reduce its time and space requirements.	It is restricted experiments to those objects for which co-citation had at least 50. Candidates to c (p)>=50.
[70]	Node centrality in a graph is disregarded by RW.	Random Walk	Hep-Th, Hep-ph, NetScience, Gr-Qc, Yeast Power, USAir C.elegans, ER, SW, BA	Random walk can produce the greatest amount of entropy.	As a result, nodes with 0 in-degree or out-degree may have centrality that is equal to zero.
[8]	Powered by power iteration and divided into two stages: normalization and iteration.	Random Walk with Restart	HepPh, HepTh, Wikipedia and, Polblogs	RWR enables it to return preferences for a query node based on relevance scores by allowing each node to have a configurable restart probability.	To determine the relevance of a query node to other nodes, RWR cannot take into account a query node's preferences.
[48]	Each term will differ from unity by a fraction depending on how many or few pathways are present at the associated vertex.	Global Leicht-Holme-Newman Index (GLHN)	AddHealth, Roget's Thesaurus,	Using an iterative technique, the resemblance of two vertices is expressed according to how similar their neighbors are.	An area where more established similarity metrics, like cosine similarity, perform poorly.
[60]	The network of a tree is a forest, and a rooted tree is one having a single identifiable vertex known as a root.	Matrix Forest Index	Kirchhoff matrix	Representation for the adjugate of the characteristic matrix.	These theorems have dual counterparts concerning converging forests.
[73]	Many pairings that work together are separated by distances larger than two.	Rooted PageRank	hep-th, astro-ph, hep-ph, gr-qc to, cond-mat,	As the rank drops, performance rises until it reaches rank one.	There are numerous opportunities to enhance performance on this work and discover methods of utilizing the training data's information more effectively.

off design process accuracy and time complexity for more accurate link prediction techniques. They demonstrate the computational effectiveness of local approaches and the topological visibility of global methods. The most often used quasi-local techniques are founded on path counting and models of random walks. So, quasi-local similarity become known and remarkable a balance between global and local similarity based methods [59]. Here are some illustrations of quasi-local techniques.

1) LOCAL RANDOM WALK (LRW)

The random walk with limited moves [76], which was useful for large and light networks, provides the foundation for the LRW proposal. As determined by the t-process random walk, this similarity can be stated as follows, where the q is an initial composition function.

$$s_{xy}^{LRW}(t) = q_x \pi_{xy}(t) + q_y \pi_{xy}(t) \tag{43}$$

2) SUPERPOSED RANDOM WALK (SRW)

The earlier result depends on LRW and the t-step is examined in SRW [76]. The purpose is to connect as many nodes as possible that are close to the objective node [59]. It's described as follows:

$$s_{xy}^{SRW}(t) = \sum\nolimits_{\tau=1}^{t} s_{xy}^{SRW}(\tau) = \sum\nolimits_{\tau=1}^{t} [q_x \pi_{xy}(\tau) + q_y \pi_{xy}(\tau)] \tag{44}$$

3) PROPFLOW PREDICTOR (PFP) INDEX

In order to anticipate if a link will occur, PropFlow [77] uses PageRank as support and analyses the information shared by nodes. The probability of outlook links is determined by the dimension of the PropFlow rate, and establishing relationships is made simpler by increasing value. PropFlow is at the present commonly used in light, weighted, unweighted, directed, undirected or dense networks.

$$s_{xy}^{PFP} = s_{(a,x)}^{PFP} \frac{w_{xy}}{\sum_{k \in \tau(x)} w_{xy}} \tag{45}$$

From above equation w_{xy} corresponds to the link weight between the pair of nodes, and x is starting node. Then $s_{(a,x)}^{PFP} = 1$, or else, $s_{(a,x)}^{PFP}$ is the collection of routes that are the shortest between x and y.

4) THIRD-ORDER RESOURCE ALLOCATION BASED ON COMMON NEIGHBOR INTERACTIONS (ORA-CNI)

The algorithm for this index increases the distribution of resources based on interactions between common neighbor's to also take into account a distance of three pathways. For nodes at a distance of three, it redefines resource allocation [45]. It is determined as,

$$s(x,y) = \sum_{z \in \lceil x \cap \lceil y} \frac{1}{|\lceil z|} + \sum_{e_{i,j} \in E | \lceil i | < | \lceil j |, i \in \lceil x, j \in \lceil y} \left(\frac{1}{|\lceil i|} - \frac{1}{|\lceil j|} \right) + \beta \sum_{[x,p,q,y] \in paths_{x,y}^{3}} \frac{1}{|\lceil p| \, |\lceil q|}, \tag{46}$$

In this case, β acts as a dampening factor to limit the impact of the three-hop resource allocation expression. The resource distribution in this index is dependent on the intricacy of interactions between shared neighbours.

5) LOCAL PATH INDEX (LP)

This index was developed form Katz Index. The major dissimilarity among Local Path and Katz Index is that it takes into account the local path lengths of 2 and 3. It is explained as,

$$S^{LP} = A^2 + \varepsilon A^3 \tag{47}$$

There is an open parameter ε. When $\varepsilon = 0$, LP = CN. In this instance, A2 and A3 stand for the number of nodes that are adjacent with two and three path lengths, respectively [59].

6) FRIEND LINK (FL)

A new dimension called "Friend Link" [78] measures the path numbers of potential nodes and is comparable to Local Path. This method utilizes a number of path length consequence techniques in addition to normalization. The accuracy of the prediction has improved once more. It's described as,

$$S_{xy}^{FL} = \sum\nolimits_{i=1}^{l} \frac{1}{i-1} \cdot \frac{\left| paths_{x,y}^{i} \right|}{\prod_{j}^{i} = 2^{(n-j)}}. \tag{48}$$

Using the formula above, n, demonstrate the diversity of the network's nodes. The route length from nodes x and y is represented by 'i' and the collection of route length 'I' from node x to y is paths $paths_{x,y}^{i}$.

V. LINK PREDICTION ALGORITHMS

Calculating the similarity or likelihood of possible links between pairs of nodes in a network, ranking these pairs, and then forecasting the links with the highest scores are the steps involved in using scores and indices for link prediction. There are two ways that these link prediction techniques can be used:

A. DIRECT CALCULATION AND HYPOTHESIS TESTING

To get the scores or indices for each pair of nodes in the network, we can apply one or more of the prediction techniques covered in section IV. Common neighbors, the Jaccard coefficient, and preferential attachment are a few examples of these strategies. For every pair of nodes, each technique assigns a numerical score or index that represents the probability that a link exists between them. We can create a connection prediction hypothesis based on these scores. For instance, we could speculate that node pairs are more likely to have a link if their scores are higher than a given threshold. Next, we apply this hypothesis to forecast whether links will exist in the network or not.

TABLE 4. Comparison of various methods considering quasi-local link structure.

Article	Key point	Method	Dataset	Advantages/ Applications	Drawback
[76]	Predict each component's links separately, and then investigate each link. It is unpredictable how two components will connect.	Local Random Walk (LRW), Superposed Random Walk (SRW)	USAir, NetScienc, Power, Yeast, C.elegans	SRW and LRW move more faster than RWR and ACT. And gives a superior prediction to a certain extent with a lower computational complexity.	For link prediction, lots of dendrograms are acquired. Every single step takes a specific time to do any random selections.
[77]	Similar to Rooted PageRank, PropFlow is a ranking system. To maintain balance, every training set is undersampled.	PropFlow Predictor Index	Phone and Cond-mat	PropFlow performs better than all other unsupervised approaches combined when some fundamental properties are used, such as node and degree.	In Cond-mat dataset PropFlow suffers mediocre performance. In Condmat, PropFlow is ineffective and there is no good analogy.
[3]	The LP index outperforms the Katz index in terms of practicality. Compared to the CN & KI, prediction accuracy is substantially greater and efficiency is quite high.	Local Path Index(LP)	Grid, INT, USAir, NS, PB PPI	LP index uses a lot less memory and CPU resources. Consequently, the LP index's advantage for large networks is exceptional.	Along with runtime complexity, memory space is still another barrier to the development of algorithms for enormous-size networks.
[78]	FriendLink algorithm consist two functions. CombinePaths() & Compute Similarity().	FriendLink (FL)	Epinions , Facebook, Hi5 web sites.	Collecting both local and global graph properties in an efficient manner.	FriendLink cannot perform better, if the strength of randomness is high in dataset

We can compute precision and recall to evaluate the efficacy of our predictions and the validity of our hypothesis. The precision of our forecasts is determined by calculating the percentage of actual links out of our anticipated links. The percentage of real links that are accurately predicted is measured by recall, which shows how comprehensive our predictions are. We can assess the effectiveness of our link prediction techniques and make any required modifications to increase their accuracy by looking at these metrics.

B. MACHINE LEARNING APPROACH

The collection of techniques offered in section IV can be used to determine the prediction scores for each pair of nodes. These scores can function as a complete feature set by quantifying the likelihood of linkages between nodes. After obtaining this feature set, we split it up into two groups: a testing set and a training set. Different machine learning models, including logistic regression, decision trees, and neural networks, are trained using the training set. The models discover patterns and relationships in the data during training that show whether or not there are links between nodes. The models are tested on the testing set after training. To do this, a comparison between the expected and actual linkages in the test set must be made. As described in section VI, the effectiveness of the link prediction approach is assessed using standard metrics like precision, recall, F1-score, and AUC-ROC. These metrics aid in determining the model's prediction accuracy and link prediction performance, offering a thorough assessment of the efficacy of the link prediction technique.

These two methods allow us to predict linkages in a network by utilizing different similarity-based link prediction scores and indices. While the machine learning approach allows us to create and evaluate more sophisticated predictive models, the direct computation and hypothesis testing approach allows us to construct and evaluate simpler predictions. When combined, these techniques offer a strong framework for network analysis's link prediction and performance assessment.

VI. EVALUATION METRICS

To assess the effectiveness of the classification models for a certain set of test data, the confusion matrix is utilized. Figure 3 depicts the binary classification confusion matrix, which simply has the positive and negative conditions.

The following cases are included in the table below.

- **True Positive (TP):** YES, as predicted by the model, and YES, as measured by the actual value.
- **True Negative (TN):** The model predicted NO, and the actual value was also NO.
- **False Positive (FP):** YES was predicted by the model, but the actual value was NO.
- **False Negative (FN):** When the model predicted NO but the actual value was YES.

We'll now look at methods for determining how well our machine learning algorithms produce their models. The metrics for evaluating the learning model are as follows:

TABLE 5. Evaluation metrics, description and article.

Evaluation Metrics	Description	Formula	Article
Classification Accuracy	It establishes how frequently the model forecasts the outcome accurately.	$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$	[78], [86]
Precision	Positive Predictive Value is another name for precision. The accuracy of a forecast is measured as a percentage of all successful predictions.	$\text{Precision} = \frac{TP}{TP + FP}$	[8], [9], [11], [14], [15], [38], [53], [55], [56], [59], [66], [70], [73], [76], [78], [79], [86], [87]
Recall	It is defined as the positive classes out of all positive classes that our model correctly forecasted.	$\text{Recall} = \frac{TP}{TP + FN}$	[1], [14], [55], [56], [63], [66], [73], [74], [78]
F1-Score	If one model has low precision and the other has high recall, it is challenging to compare the two. F-score is a useful tool for this. When recall and precision are identical, the F-score is at its highest.	$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	[14], [86]
AUC	AUC is the probability that a random positive case comes after a random negative example. AUC varies from 0 to 1. A model whose forecasts are 100% inaccurate has an AUC of 0.0; one whose predictions are 100% right has an AUC of 1.0.	$\text{AUC} = \frac{\text{Recall} + \text{Sensibility}}{2}$	[2], [3], [6], [8], [9], [10], [11], [13], [15], [38], [39], [40], [52], [53], [54], [57], [59], [70], [76], [77], [78], [79], [80], [81], [86]
Mean Average Precision (MAP)	It measures the average precision across all relevant links in the predicted ranking	$\text{MAP} = \frac{1}{N}\sum_{i=1}^{N}(AP_i)$	[8], [9], [14], [78]

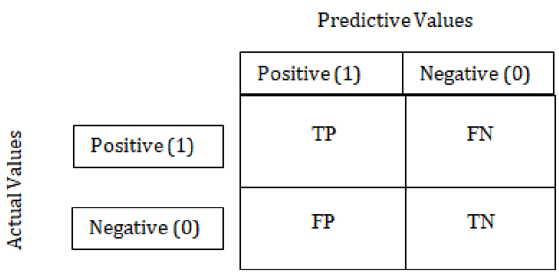


FIGURE 3. Confusion matrix.

VII. APPLICATIONS IN REAL-WORLD SCENEROS

Table 6 gives the scenario based practical use of similarity based link prediction methods.

VIII. CHALLENGES AND FUTURE PLANS

The advantages and disadvantages of link prediction applications in social networks have been the subject of numerous studies. The connection prediction tasks, however, grow increasingly difficult as social networks develop and become more complex. A number of current link predictions in social networks challenges have not yet been thoroughly examined. Predicting links in social networks faces a number of challenges and potential future paths are presented in this section:

- **Simple and diverse network:**
  In Simple network, the nodes and links have same features whereas in diverse network the nodes and links may have some variation in features, meaning they include a wide variety of node and edge types. Because of this, it is challenging to use a universal link prediction method across different network architectures. Most of the methods proposed by the various authors for link prediction are based on the simple network where all nodes and edges share the common features but the many networks are of diverse types of nodes and edges that create challenges. For addressing such problems, a lot of new methods have been introduced [82], [83].
- **Scalability:**
  As the size of social networks grows, the computing cost of running link prediction algorithms may rise. The problem of scalability of network is the particular challenges in Link prediction which is to be solved. The scalability of link prediction algorithms is always doubtful until and unless the link prediction algorithm implemented on large size of dataset. But based on the literature available on link prediction, it is observed that most of the researchers implemented their link prediction algorithms on small data which is not enough to make sure the scalability. The developed link prediction method should be implemented on the enough size of dataset i.e. on large scale networks [84]to ensure the scalability so that the same method can able to predict the links in big network.
- **Multi model network:**
  Social networks are multi-modal, meaning that they include not only text and links but also things like photos

TABLE 6. Methods, scenario and practical use.

Methods	Scenario	Practical Use
CN, JC, AA Index, RAI, RA-CNI, HPI, HDI	Recommendation on Social Media Platforms and in E-Commerce sites.	Raising user satisfaction by improving recommendation accuracy by making suggestions for products that are consistent with a user's more general interests or preferences as determined by their network interaction patterns.
	Collaboration Networks in Academia.	Establishing clusters that make use of a variety of expertise to improve multidisciplinary research projects or funding allocation techniques, which will encourage innovation and knowledge integration.
	Biological Networks.	Finding important proteins or genes involved in complex biological processes might help understand complex diseases and facilitate the development of therapeutic targets or biomarkers.
PA	Professional Networking Platforms.	LinkedIn suggests possible relationships using an algorithm that is driven by preference attachment.
	Citation Networks.	Author impact scores are determined by ResearchGate through the analysis of citation trends.
	Social Media Influence.	On the Explore page, Instagram highlights posts from users who are very active.
	Collaboration Networks in Movies.	Preferential Attachment-Based Casting Suggestions and Network Analysis Talent Scouting
SOI	Ecology and Biodiversity.	Researchers can help efficient conservation and management efforts by gaining important insights about the composition of species and the effects of environmental changes.
	Community Ecology	Researchers are able to evaluate how comparable species assemblages are across several sampling sites or time intervals.
	Genetics and Molecular Biology	The genomes of several harmful bacterial strains are compared by researchers. They examine genetic content variations, the existence of virulence factors, and antibiotic resistance genes.
	Information Retrieval and Text Analysis.	The Sorensen Index is used by educational institutions to identify plagiarism in student essays and research papers.
CNDP	Traffic and Transportation Networks.	Lowering fuel costs, cutting commute times, and enhancing air quality by using effective traffic management techniques.
HOPI	Information Network.	By recommending relevant documents or concepts that are connected through higher-order thematic similarities, knowledge discovery and information retrieval systems are improved and comprehensive research and decision-making processes are made easier.
FP	Information Spread in Social Networks	Identifying key informants and the best times to spread important information during emergencies in order to improve crisis communication plans.
	Epidemic Spread in Public Health.	Providing information for public health initiatives like immunization campaigns, isolation protocols, and travel bans to reduce the spread of disease.
	Resource Allocation in Supply Chains.	Lowering transportation costs, cutting down on stockouts, and enhancing the dependability of deliveries in intricate supply chain networks.
	Opinion Formation in Online Forums.	Improving online engagement tactics through the identification of important influencers and comprehension of the variables affecting the convergence or divergence of opinions.
	Financial Contagion in Banking Networks.	Bolstering regulatory actions and stress testing as financial stability measures to stop the spread of negative impacts during recessions.
MERW	Influence Maximization in Social Networks.	Focusing on influential users to increase reach and engagement through marketing campaigns or message distribution.
	Disease Spread Modeling in Epidemiology.	Providing information on immunization programs, isolation protocols, and the distribution of healthcare resources to effectively control disease outbreaks.
	Financial Risk Management.	Improving regulatory actions and stress testing programs for financial stability in order to reduce the potential for spillover during market crises.
SR	Collaborative Filtering in Movie or Music Recommendation.	By making content recommendations that suit each user's unique interests and preferences, you can increase user happiness and engagement.
	Fraud Detection in Financial Transactions.	Monitoring transactions that have traits with known fraudulent activity in order to improve fraud detection algorithms.
	Network Security and Intrusion Detection.	Improving cyber security protocols by identifying anomalous network activity and averting possible breaches or assaults.

TABLE 7. Prediction methods & its complexity [45], [52], [59], [64], [75], [80], [84].

LP Methods	Local link structural similarity	Global link structural similarity	Quasi-Local link structural similarity	Time-Complexity
Common neighbors	✓			$O(vk^2)$
Jaccard index	✓			$O(vk^2)$
Preferential Attachment	✓			$O(vk^2)$
Resource-Allocation Based on Common neighbors	✓			$O(vk^4)$
Resource Allocation Index	✓			$O(vk^3)$
Adamic Adar Index	✓			$O(vk^6)$
Local LeichtHolme Newman Index	✓			$O(vk^3)$
Hub Depressed Index	✓			$O(vk^3)$
Hub promoted Index	✓			$O(vk^3)$
Salton Index	✓			$O(vk^3)$
Sorenson Index	✓			$O(vk^3)$
Parameterized Algorithm	✓			–
Node Coupling Clustering	✓			$O(v^2)$
Common neighbors Degree Penalization	✓			$O(vk^2)$
Common neighbors and Distance	✓			–
CAR-Based Indices	✓			$O(vk^4)$
Local Interacting Score	✓			$O(lvk^3)$
Functional Similarity Weight	✓			$O(vk^3)$
Local Affinity Structure Based	✓			–
Higher Order Path Index		✓		–
Average Compute Time		✓		$O(v^3)$
Pseudoinverse of Laplacian Matrix		✓		$O(v^3)$
Blondel Index		✓		$O(v^2k)$
Random Forest Kernel Index		✓		$O(v^3)$
Flow Propagation		✓		$O(v^2k)$
Maximal Entropy Random Walk		✓		$O(v^2k)$
Katz Index		✓		$O(v^3)$
SimRank		✓		$O(v^2k^{2l+2})$
Random Walk		✓		$O(v^2k)$
Random Walk with Restart		✓		$O(v^2k)$
Negated Shortest Path		✓		$O(ev\ \log v)$
Global LeichtHolme Newman Index		✓		$O(v^2k)$
Matrix Forest		✓		–
Rooted Page Rank		✓		–
Local Random Walk			✓	$O(lv^2k)$
Superposed Random Walk			✓	$O(lv^2k)$
PropFlow Predictor			✓	$O(vlk^l)$
TORA-CNI			✓	$O(vk^6)$
Local Path Index			✓	$O(lv^2k)$
Friend Link			✓	$O(lv^2k)$

and videos. One potential path forward for link prediction models is to include more modalities like these [85].

- **Explainability:**
  In machine learning, explaining the logic behind the connection predictions may be difficult. In order to make progress in the field of link prediction in social networks, it will be crucial to develop models that can be easily explained. Developing explainable models is an important future direction for link prediction in social networks.
- **Including tangential information:**
  Link prediction in social networks may benefit from the addition of ancillary data, such as demographics, timestamps, and other features of the nodes.
- **Mixing several kinds of ties together:**
  In actual networks, friends, coworkers, and love partners are just a few of the many possible kinds of ties that may be represented by a link. One potential path forward for link prediction is the creation of models with the flexibility to anticipate a wide variety of different connection types. Developing models that can predict multiple types of links can be a future direction for link prediction
- **Including network structure:**
  The structure of a network may give useful information for link prediction, thus it's important to take it into account. One potential path forward is for link prediction models to take into account the underlying network topology.
- **Dynamic in time:**
  Time variable Network is a network which grows over a period of time. The difficulty in time varying network for link prediction is that we have to continuously check the growing network. This makes it hard for link prediction algorithms to represent the dynamic nature of networks over time [7].
- **Data Gaps:**
  In many cases, a significant amount of data is missing in social networks, which can make it difficult to accurately predict links. It might be challenging to develop reliable connection predictions in social networks because of the large amounts of missing data that often occur.

IX. SUMMARY AND DISCUSSION

The local-similarity based strategy makes use of approaches for predicting links based on structure, which is the first type of similarity approach. These approaches base their similarity score calculation on the node and path having a maximum length of 2 and no more. As a result, a few noteworthy and likely connections could be missing, in addition to information. For each node in the network, computing the comparison score will be time-consuming and challenging. In the second types of methods which are based on global structure, the link prediction calculate the similarity score depending on the nodes' paths, the graph's overall link structure, and the number of nodes with paths length greater than two. However, when dealing with huge networks, like online social networks, where several bytes of information must be examined to predict the link, generating similarity scores based on the global environment takes time and is difficult [78], [79] [81]. Link prediction techniques based on quasi-local similarity appear to be more accurate than those based on local similarity, which is the last type of similarity-based approach. Methods investigates quasi-local link structure wrap complete network while allowing for pathways between nodes that are longer than two. Also given the comparison of the link predictions methods as per the link structure they have used.

Example:

Complexity calculation for Common Neighbors

Common Neighbors: need to predict the number of common neighbors for each node pair. Let graph G=(V, E), where V is a set of vertices and E stands for the edge. For any pair u, v ∈ V, the common neighbors of nodes connected to both nodes.

Steps to Calculate Common Neighbors:

Step 1: Retrieve the neighbors of node u: N(u).

Step 2: Retrieve the neighbors of node v: N(v).

Step 3: Compute the intersection N(u)∩N(v), which gives the common neighbors.

Time Complexity Analysis:

1. Retrieving neighbors:
   - If the graph is represented as an adjacency list, then time complexity of extracting neighbors for a node can be defined by the degree of said node.
   - Let d(u) and d(v) be the degree of nodes u and v respectively. Retrieving N(u) takes O(d(u)) time, and retrieving N(v) takes O(d(v)) time.
2. Computing the intersection:
   - The time complexity is O(min(d(u),d(v))) to compute the intersection of two sets N(u) and N(v) when those are implemented by hash set
   - In case the sets are not hashed, and we have to compare each element of one set with another. The worst time complexity would be O(d(u)×d(v)).

Time Complexity:

Best/Average case: O(d(u)+d(v))

Worst case: O(d(u)×d(v))

X. CONCLUSION

In this paper we briefly summarized comparative study about the several methods for predicting the links in social network with similarity based approach. In the beginning, we first make clear that the problem of predicting the links in social networks. Discussions are made of several similarity-based methodology kinds that are usually used while solving link a prediction problem. In these approaches, we also check and discussed time complexity of several methods with its type. A lot of challenging issues are yet unfinished in the relatively new academic field of social network link prediction. To comprehend why certain strategies perform better or worse depends on the network to which they are applied to, more knowledge must be gained. A key research challenge is determining which network attributes result in enhanced effectiveness for each approach. The research on various link prediction approaches using a similarity-based approach is lacking, and this paper helps to address that gap.

The field of link prediction continues to evolve and advance, and there are many exciting future directions for research, including incorporating node attributes, handling dynamic networks, developing scalable methods, and integrating link prediction with other methods. The findings of research in link prediction suggest that this is a rich and active field with many exciting developments and opportunities for future research.

AUTHOR CONTRIBUTIONS

Conceptualization, Sachin U. Balvir, Mukesh M. Raghuwanshi, and Purushottam D. Shobhane; methodology, Sachin U. Balvir, Mukesh M. Raghuwanshi, and Purushottam D. Shobhane; validation, Mukesh M. Raghuwanshi, and Purushottam D. Shobhane; formal analysis, Sachin U. Balvir, and Purushottam D. Shobhane; investigation, Sachin U. Balvir, Mukesh M. Raghuwanshi, and Purushottam D. Shobhane; writing—original draft preparation, Sachin U. Balvir, and Mukesh M. Raghuwanshi; writing—review and editing, Sachin U. Balvir, Mukesh M. Raghuwanshi and Purushottam D. Shobhane; supervision, Mukesh M. Raghuwanshi; All authors have read and agreed to the published version of the manuscript.

FUNDING

This research received no external funding.

DATA AVAILABILITY STATEMENT

Not Applicable.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ABBREVIATIONS

The following abbreviations are used in this manuscript:

SN	Social Network
DSLP	Deep Learning Link Prediction
CCPA	Common Neighbor and Centrality based Parameterize Algorithm
GCCDC	Global Clustering Coefficient-dependent Degree Centrality
DMNS	Conditional Diffusion-based Multi-level Negative Sampling
NN	Neural network
XGB	XGBoost
LDA	Linear Discriminant Analysis
RFC	Random Forest based Classifier
CN	Common neighbors
JC	Jaccard Coefficient
PA	Preferential Attachment
AA	Adamic/Adar
RAI	Resource Allocation Index
RA- CNI	Resource Allocation Based on Common Neighbor Interactions
HPI	Hub Promoted Index
HDI	Hub Depressed Index
LLHN	Local Leicht-Holme-Newman Index
SI	Salton Index
SOI	Soreson Index
CNDP	Common neighbors Degree Penalization
CND	Common Neighbor and Distance
LC	Local community
LIT	Local Interacting Score
HOPI	High Order Path Index
FSW	Functional Similarity Weight
LAS	Local Affinity Structure Index
GLHN	Leicht-Holme-Newman Index
RFK	Random Forest Kernel Index
BI	Blondel Index
PLM	Pseudoinverse of the Laplacian Matrix
ACT	Average Commute Time
FP	Flow Propagation
MERW	Maximal Entropy Random Walk
KI	Katz Index
SR	SimRank
RW	Random Walk
RWR	Random Walk with Restart
NSP	Negated Shortest Path
MF	Matrix Forest Index
LPI	Local Path index
HPRD	Human Protein Reference Database
LRW	Local Random Walk
SRW	Superposed random walk
PFP	PropFlow Predictor
ORA-CNI	Third-Order Resource Allocation Based on Common Neighbor Interactions
FL	Friend Link
LRW	Local Random Walk
SRW	Superposed Random Walk
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
MAP	Mean Average Precision

REFERENCES

[1] A. Rawashdeh, "Performance based comparison between several link prediction methods on various social networking datasets (including two new methods)," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 12, pp. 1–8, 2020, doi: 10.14569/IJACSA.2020.0111201.

[2] L. Dong, Y. Li, H. Yin, H. Le, and M. Rui, "The algorithm of link prediction on social network," *Math. Problems Eng.*, vol. 2013, pp. 1–7, Nov. 2013, doi: 10.1155/2013/125123.

[3] L. Lü, C.-H. Jin, and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 4, pp. 1–24, Oct. 2009, doi: 10.1103/PHYSREVE.80.046122.

[4] L. Getoor and C. P. Diehl, "Link mining," *ACM SIGKDD Explor. Newslett.*, vol. 7, no. 2, pp. 3–12, Dec. 2005, doi: 10.1145/1117454.1117456.

[5] K. P. Udagepola and F. Chiroma, "Review social network analysis and mining: Link prediction," *Int. J. Inf. Electron. Eng.*, vol. 6, no. 4, pp. 2–6, 2016.

[6] A. Zareie and R. Sakellariou, "Similarity-based link prediction in social networks using latent relationships between the users," *Sci. Rep.*, vol. 10, no. 1, pp. 1–24, Nov. 2020.

[7] S. Das and S. K. Das, "A probabilistic link prediction model in time-varying social networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6, doi: 10.1109/ICC.2017.7996909.

[8] W. Jin, J. Jung, and U. Kang, "Supervised and extended restart in random walks for ranking and link prediction in networks," *PLoS One*, vol. 14, no. 3, Mar. 2019, Art. no. e0213857, doi: 10.1371/JOURNAL.PONE.0213857.

[9] A. Kumar, S. Mishra, S. S. Singh, K. Singh, and B. Biswas, "Link prediction in complex networks based on significance of higher-order path index (SHOPI)," *Phys. A, Stat. Mech. Appl.*, vol. 545, May 2020, Art. no. 123790, doi: 10.1016/J.PHYSA.2019.123790.

[10] 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, Jan. 2020, doi: 10.1038/S41598-019-57304-Y.

[11] J. Yang and X.-D. Zhang, "Predicting missing links in complex networks based on common neighbors and distance," *Sci. Rep.*, vol. 6, no. 1, pp. 1–10, Dec. 2016, doi: 10.1038/SREP38208.

[12] K. Li, S. Gu, and D. Yan, "A link prediction method based on neural networks," *Appl. Sci.*, vol. 11, no. 11, p. 5186, Jun. 2021, doi: 10.3390/APP11115186.

[13] A. Saxena, G. Fletcher, and M. Pechenizkiy, "NodeSim: Node similarity based network embedding for diverse link prediction," *EPJ Data Sci.*, vol. 11, no. 1, pp. 1–22, Dec. 2022, doi: 10.1140/EPJDS/S13688-022-00336-8.

[14] F. Sarhangnia, S. Mahjoobi, and S. Jamshidi, "A novel similarity measure of link prediction in bipartite social networks based on neighborhood structure," *Open Comput. Sci.*, vol. 12, no. 1, pp. 112–122, Mar. 2022, doi: 10.1515/COMP-2022-0233.

[15] H. Saeidinezhad, E. Parvinnia, and R. Boostani, "A new similarity-based link prediction algorithm based on combination of network topological features," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 12, no. 3, p. 2802, Jun. 2022, doi: 10.11591/IJECE.V12I3.PP2802-2811.

[16] C. Zhang, K.-K. Shang, and J. Qiao, "Adaptive similarity function with structural features of network embedding for missing link prediction," *Complexity*, vol. 2021, no. 1, pp. 1–15, Jan. 2021.

[17] K.-K. Shang, T.-C. Li, M. Small, D. Burton, and Y. Wang, "Link prediction for tree-like networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 29, no. 6, pp. 1–29, Jun. 2019.

[18] K.-K. Shang and M. Small, "Link prediction for long-circle-like networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 105, no. 2, Feb. 2022, Art. no. 024311.

[19] K.-K. Shang, M. Small, X.-K. Xu, and W.-S. Yan, "The role of direct links for link prediction in evolving networks," *EPL (Europhysics Lett.)*, vol. 117, no. 2, Jan. 2017, Art. no. 28002.

[20] K.-K. Shang, M. Small, and W.-S. Yan, "Link direction for link prediction," *Phys. A, Stat. Mech. Appl.*, vol. 469, pp. 767–776, Mar. 2017.

[21] R. Khushal and U. Fatima, "Fuzzy machine learning logic utilization on hormonal imbalance dataset," *Comput. Biol. Med.*, vol. 174, May 2024, Art. no. 108429, doi: 10.1016/J.COMPBIOMED.2024.108429.

[22] N. Choudhury, "Community-aware evolution similarity for link prediction in dynamic social networks," *Mathematics*, vol. 12, p. 285, Jun. 2024.

[23] C. Liu, Y. Han, H. Xu, S. Yang, K. Wang, and Y. Su, "A community detection and graph-neural-network-based link prediction approach for scientific literature," *Mathematics*, vol. 12, no. 3, p. 369, Jan. 2024, doi: 10.3390/MATH12030369.

[24] J. Yang and Y. Wu, "Link prediction based on depth structure in social networks," *Int. J. Mach. Learn. Cybern.*, vol. 1, pp. 1–27, May 2024, doi: 10.1007/S13042-024-02178-4.

[25] Y. V. Nandini, T. J. Lakshmi, M. K. Enduri, and H. Sharma, "Link prediction in complex networks using average centrality-based similarity score," *Entropy*, vol. 26, no. 6, p. 433, May 2024, doi: 10.3390/E26060433.

[26] C. Meng and H. Motevalli, "Link prediction in social networks using hyper-motif representation on hypergraph," *Multimedia Syst.*, vol. 30, no. 3, p. 123, Jun. 2024, doi: 10.1007/S00530-024-01324-W.

[27] M. Li, S. Zhou, D. Wang, and G. Chen, "Missing link prediction using path and community information," *Computing*, vol. 106, no. 2, pp. 521–555, Feb. 2024.

[28] M. Abedini and H. Shakibian, "Inter-layer similarity-based graph neural network for link prediction in social multiplex networks," in *Proc. 10th Int. Conf. Web Res. (ICWR)*, Apr. 2024, pp. 1–24, doi: 10.1109/ICWR61162.2024.10533376.

[29] X. Liu, X. Li, G. Fiumara, and P. De Meo, "Link prediction approach combined graph neural network with capsule network," *Expert Syst. Appl.*, vol. 212, pp. 0957–4174, Jun. 2022.

[30] Q. Tan, X. Zhang, N. Liu, D. Zha, L. Li, R. Chen, S. H. Choi, and X. Hu, "Bring your own view: Graph neural networks for link prediction with personalized subgraph selection," in *Proc. 16th ACM Int. Conf. Web Search*, 2023, pp. 1–24.

[31] Z. Zhao, A. Hu, N. Zhang, J. Xie, Z. Du, L. Wan, and R. Yan, "Mining node attributes for link prediction with a non-negative matrix factorization-based approach," *Knowl.-Based Syst.*, vol. 299, Sep. 2024, Art. no. 112045, doi: 10.1016/J.KNOSYS.2024.112045.

[32] U. Fatima, S. Hina, and M. Wasif, "A novel global clustering coefficient-dependent degree centrality (GCCDC) metric for large network analysis using real-world datasets," *J. Comput. Sci.*, vol. 70, Jun. 2023, Art. no. 102008, doi: 10.1016/J.JOCS.2023.102008.

[33] A. Saxena, G. Fletcher, and M. Pechenizkiy, "FairSNA: Algorithmic fairness in social network analysis," *ACM Comput. Surveys*, vol. 56, no. 8, pp. 1–45, Aug. 2024, doi: 10.1145/3653711.

[34] T.-K. Nguyen and Y. Fang, "Diffusion-based negative sampling on graphs for link prediction," in *Proc. ACM Web Conf.*, vol. 34, May 2024, pp. 948–958, doi: 10.1145/3589334.3645650.

[35] S. U. Balvir, M. M. Raghuwanshi and P. S. Borkar, "Node2 Vec and machine learning: A powerful duo for link prediction in social network," *J. Electr. Syst.*, vol. 20, no. 2s, pp. 639–649, Apr. 2024, doi: 10.52783/JES.1530.

[36] M. Kumar, S. Mishra, S. S. Singh, and B. Biswas, "Community-enhanced link prediction in dynamic networks," *ACM Trans. Web*, vol. 18, no. 2, pp. 1–32, May 2024, doi: 10.1145/3580513.

[37] P. Srilatha and R. Manjula, "Similarity index based link prediction algorithms in social networks: A survey," *J. Telecommun. Inf. Technol.*, vol. 2, no. 2016, pp. 87–94, Jun. 2016.

[38] 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, doi: 10.1016/J.PHYSA.2010.11.027.

[39] 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, doi: 10.1140/EPJB/E2009-00335-8.

[40] H. Yuliansyah, Z. A. Othman, and A. A. Bakar, "Taxonomy of link prediction for social network analysis: A review," *IEEE Access*, vol. 8, pp. 183470–183487, 2020, doi: 10.1109/ACCESS.2020.3029122.

[41] L. Yao, L. Wang, L. Pan, and K. Yao, "Link prediction based on common-neighbors for dynamic social network," *Proc. Comput. Sci.*, vol. 83, pp. 82–89, Jul. 2016, doi: 10.1016/J.PROCS.2016.04.102.

[42] J. Podani, "The wonder of the Jaccard coefficient: From Alpine floras to bipartite networks," *Flora Mediterr.*, vol. 31, pp. 105–123, Jul. 2021, doi: 10.7320/FLMEDIT31SI.105.

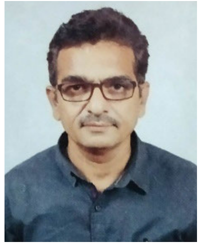
[43] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999, doi: 10.1126/SCIENCE.286.5439.509.

[44] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social Netw.*, vol. 25, no. 3, pp. 211–230, Jul. 2003.

[45] 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, Dec. 2017, doi: 10.1145/3012704.
[46] Q. Ou, Y.-D. Jin, T. Zhou, B.-H. Wang, and B.-Q. Yin, "Power-law strength-degree correlation from resource-allocation dynamics on weighted networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 75, no. 2, pp. 1–5, Feb. 2007, doi: 10.1103/PHYSREVE.75.021102.
[47] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, pp. 1551–1555, 2002.
[48] E. A. Leicht, P. Holme, and M. E. J. Newman, "Vertex similarity in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 73, no. 2, pp. 1–10, Feb. 2006, doi: 10.1103/PHYSREVE.73.026120.
[49] G. Salton and M. J. McGill, *Introduction to Modem Information Retrieval*. New York, NY, USA: McGraw-Hill, 1986.
[50] A. Rodriguez, B. Kim, M. Turkoz, J.-M. Lee, B.-Y. Coh, and M. K. Jeong, "New multi-stage similarity measure for calculation of pairwise patent similarity in a patent citation network," *Scientometrics*, vol. 103, no. 2, pp. 565–581, May 2015, doi: 10.1007/S11192-015-1531-8.
[51] A. Ceska, "Estimation of the mean floristic similarity between and within sets of vegetational relevés," *Folia Geobotanica Et Phytotaxonomica*, vol. 1, no. 3, pp. 93–100, Jun. 1966, doi: 10.1007/BF02852438.
[52] F. Li, J. He, G. Huang, Y. Zhang, Y. Shi, and R. Zhou, "Node-coupling clustering approaches for link prediction," *Knowl.-Based Syst.*, vol. 89, pp. 669–680, Nov. 2015, doi: 10.1016/J.KNOSYS.2015.09.014.
[53] S. Rafiee, C. Salavati, and A. Abdollahpouri, "CNDP: Link prediction based on common neighbors degree penalization," *Phys. A, Stat. Mech. Appl.*, vol. 539, Feb. 2020, Art. no. 122950, doi: 10.1016/J.PHYSA.2019.122950.
[54] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi, "From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks," *Sci. Rep.*, vol. 3, no. 1, pp. 1–13, Apr. 2013, doi: 10.1038/SREP01613.
[55] L. W. G. Liu and J. Li, "Assessing and predicting protein interactions using both local and global network topological metrics," *Genom Informat.*, vol. 149, pp. 138–149, Jul. 2008.
[56] H. N. Chua, W.-K. Sung, and L. Wong, "Exploiting indirect neighbours and topological weight to predict protein function from protein–protein interactions," *Bioinformatics*, vol. 22, no. 13, pp. 1623–1630, Jul. 2006, doi: 10.1093/BIOINFORMATICS/BTL145.
[57] Q. Sun, R. Hu, Z. Yang, Y. Yao, and F. Yang, "An improved link prediction algorithm based on degrees and similarities of nodes," in *Proc. IEEE/ACIS 16th Int. Conf. Comput. Inf. Sci. (ICIS)*, May 2017, pp. 13–18, doi: 10.1109/ICIS.2017.7959962.
[58] E. C. Mutlu, T. Oghaz, A. Rajabi, and I. Garibay, "Review on learning and extracting graph features for link prediction," *Mach. Learn. Knowl. Extraction*, vol. 2, no. 4, pp. 672–704, Dec. 2020, doi: 10.3390/MAKE2040036.
[59] H. Wang and Z. Le, "Seven-layer model in complex networks link prediction: A survey," *Sensors*, vol. 20, no. 22, p. 6560, Nov. 2020, doi: 10.3390/S20226560.
[60] P. Chebotarev and E. Shamis, "Matrix-forest theorems," 2006, *arXiv:0602575*.
[61] V. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, "A measure of similarity between graph vertices: Applications to synonym extraction and web searching," 2004, *arXiv preprint cs/0407061*.
[62] D. A. Spielman, "Spectral graph theory and its applications," in *Proc. 48th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2007, pp. 29–38, doi: 10.1109/FOCS.2007.4389477.
[63] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 3, pp. 355–369, Mar. 2007, doi: 10.1109/TKDE.2007.46.
[64] 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, doi: 10.1007/S11432-014-5237-Y.
[65] L. Lu and T. Zhou, "Link prediction in weighted networks: The role of weak ties," *EPL (Europhysics Lett.)*, vol. 89, no. 1, p. 18001, Jan. 2010, doi: 10.1209/0295-5075/89/18001.
[66] O. Vanunu and R. Sharan, "A propagation-based algorithm for inferring gene-disease associations," in *Proc. Ger. Conf. Bioinf.*, 2008, pp. 54–63.
[67] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, "Localization of the maximal entropy random walk," *Phys. Rev. Lett.*, vol. 102, no. 16, pp. 1–4, Apr. 2009, doi: 10.1103/PHYSREVLETT.102.160602.
[68] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.
[69] G. Jeh and J. Widom, "SimRank: A measure of structural-context similarity," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2002, pp. 538–543.
[70] R. H. Li, J. X. Yu, and J. Liu, "Link prediction: The power of maximal entropy random walk," in *Proc. 20th ACM Int. Conf. Inf.*, Oct. 2011, pp. 1147–1156.
[71] H. Tong, C. Faloutsos, and J.-Y. Pan, "Fast random walk with restart and its applications," in *Proc. 6th Int. Conf. Data Mining (ICDM)*, Dec. 2006, pp. 613–622, doi: 10.1109/ICDM.2006.70.
[72] N. Z. Gong, A. Talwalkar, L. Mackey, L. Huang, E. C. R. Shin, E. Stefanov, E. Shi, and D. Song, "Joint link prediction and attribute inference using a social-attribute network," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 2, pp. 1–20, Apr. 2014.
[73] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007, doi: 10.1002/ASI.V58:7.
[74] F. Fouss, L. Yen, A. Pirotte, and M. Saerens, "An experimental investigation of graph kernels on a collaborative recommendation task," in *Proc. 6th Int. Conf. Data Mining (ICDM)*, Dec. 2006, pp. 863–868, doi: 10.1109/ICDM.2006.18.
[75] A. Samad, M. Qadir, I. Nawaz, M. Islam, and M. Aleem, "A comprehensive survey of link prediction techniques for social network," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 7, no. 23, May 2020, Art. no. 163988, doi: 10.4108/EAI.13-7-2018.163988.
[76] W. Liu and L. Lü, "Link prediction based on local random walk," *EPL (Europhysics Lett.)*, vol. 89, no. 5, p. 58007, Mar. 2010, doi: 10.1209/0295-5075/89/58007.
[77] 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*, Jul. 2010, pp. 243–252, doi: 10.1145/1835804.1835837.
[78] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "Fast and accurate link prediction in social networking systems," *J. Syst. Softw.*, vol. 85, no. 9, pp. 2119–2132, Sep. 2012, doi: 10.1016/J.JSS.2012.04.019.
[79] A. Chartsias, "Link prediction in large scale social networks using Hadoop," Dept. Elect. Comput. Eng., Tech. Univ. Crete, Tech. Rep., 2010.
[80] H. Wu, C. Song, Y. Ge, and T. Ge, "Link prediction on complex networks: An experimental survey," *Data Sci. Eng.*, vol. 7, no. 3, pp. 253–278, Sep. 2022, doi: 10.1007/S41019-022-00188-2.
[81] M. Fire, L. Tenenboim, O. Lesser, R. Puzis, L. Rokach, and Y. Elovici, "Link prediction in social networks using computationally efficient topological features," in *Proc. IEEE 3rd Int. Conf. Privacy, Secur., Risk Trust IEEE 3rd Int. Conf. Social Comput.*, Oct. 2011, pp. 73–80, doi: 10.1109/PASSAT/SOCIALCOM.2011.20.
[82] S. Negi and S. Chaudhury, "Link prediction in heterogeneous social networks," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2016, pp. 609–617, doi: 10.1145/2983323.2983722.
[83] J.-C. Li, D.-L. Zhao, B.-F. Ge, K.-W. Yang, and Y.-W. Chen, "A link prediction method for heterogeneous networks based on BP neural network," *Phys. A, Stat. Mech. Appl.*, vol. 495, pp. 1–17, Apr. 2018, doi: 10.1016/J.PHYSA.2017.12.018.
[84] A. Zhiyuli, X. Liang, and Y. Chen, "HSEM: Highly scalable node embedding for link prediction in very large-scale social networks," *World Wide Web*, vol. 22, no. 6, pp. 2799–2824, Nov. 2019, doi: 10.1007/S11280-018-0649-Z.
[85] P. Symeonidis and C. Perentis, "Link prediction in multi-modal social networks," in *Proc. Eur. Conf.*, 2014, pp. 147–162.
[86] S. K. Sahu, A. Mokhade, and N. D. Bokde, "An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance: Recent progress and challenges," *Appl. Sci.*, vol. 13, no. 3, p. 1956, Feb. 2023, doi: 10.3390/APP13031956.
[87] S. U. Balvir, M. M. Raghuwanshi, and K. R. Singh, "A comprehensive survey on learning based methods for link prediction problem," in *Proc. 6th Int. Conf. Inf. Syst. Comput. Netw. (ISCON)*, Mar. 2023, pp. 1–7, doi: 10.1109/ISCON57294.2023.10112010.

SACHIN U. BALVIR received the M.Tech. degree in computer science and engineering form the Visvesvaraya National Institute of Technology (VNIT), Nagpur, India. He is currently pursuing the Ph.D. degree in computer science and technology from the Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur. He is also working as an Assistant Professor with the Department of Computer Science and Engineering, S. B. Jain Institute of Technology, Management and Research, Nagpur. He has been engaged in teaching for more than 15 years. He has more than ten research publications in various international and national journals and conferences. His research interests include web mining, social network analysis, and link prediction.

MUKESH M. RAGHUWANSHI received the M.Tech. degree in computer science from Indian Institute of Technology, Kharagpur, India, and the Ph.D. degree in computer science from the Visvesvaraya National Institute of Technology (VNIT), Nagpur, India. He is currently the Dean of Engineering with the S. B. Jain Institute of Technology, Management and Research, Nagpur. He is a highly ingenious mentor and a passionate technocrat with more than 35 years of experience in academics, administration, software project development and management, and hardware installations and maintenance. He has a proven track record in establishing new institutes and departments. He is an expertise in the design and delivery of engineering curriculum and courses. His areas of specialization are genetic algorithm, multi-objective optimization, algorithms, and data science.



PURUSHOTTAM D. SHOBHANE received the M.Sc. degree in mathematics from RTM Nagpur University, Nagpur, India, and the Ph.D. degree in mathematics from Gondwana University, Gadchiroli, India. He has been engaged in teaching for more than 30 years and in research for more than ten years. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Symbiosis Institute of Technology, Nagpur Campus, Nagpur, Symbiosis International (Deemed University), Pune, Maharashtra, India. He has authored four books on mathematics and has presented more than 15 papers in international/national journals/conferences. His research interests include general theory of relativity and cosmology and mathematical modeling.