The Journal of Supercomputing (2024) 80:3902–3942
https://doi.org/10.1007/s11227-023-05591-8

A comprehensive survey of link prediction methods

Djihad Arrar[1] · **Nadjet Kamel**[1] · **Abdelaziz Lakhfif**[1]

Accepted: 17 August 2023 / Published online: 7 September 2023

Abstract
Link prediction aims to anticipate the probability of a future connection between two nodes in a given network based on their previous interactions and the network structure. Link prediction is a rapidly evolving field of research that has attracted interest from physicists and computer scientists. Over the years, numerous methods have been developed for link prediction, encompassing similarity-based indices, machine learning techniques, and more. While existing surveys have covered link prediction research until 2020, there has been a substantial surge in research activities in recent years, particularly between 2021 and 2023. This increased interest underscores the pressing need to comprehensively explore the latest advancements and approaches in link prediction. We analyse and present the most notable research from 2018 to 2023. Our goal is to offer a comprehensive overview of the recent developments in the field. Besides summarizing and presenting previous experimental results, our survey offers a comprehensive analysis highlighting the strengths and limitations of various link prediction methods.

Keywords Link prediction · Machine learning · Graph neural network

1 Introduction

Link prediction is the task that estimates the probability of a future link between two nodes within a network by considering their past interactions and the overall network topology. A network, also known as a graph, is a collection of nodes (entities) and edges (connections or interactions) between these nodes. It is a topic of significant interest in network analysis due to its diverse applications. The prediction of links

✉ Djihad Arrar
djihad.arrar@univ-setif.dz

Nadjet Kamel
nkamel@univ-setif.dz

Abdelaziz Lakhfif
abdelaziz.lakhfif@univ-setif.dz

[1] LRSD, Computer Science Department, University Ferhat Abbas Setif 1, Setif, Algeria

plays an important role in comprehending the network's structure and dynamics and making accurate predictions about its future behaviour.

Link prediction has various applications across different domains. First, in recommender systems [1–3], this task can be used to predict new items or products or services based on the user's actions and preferences, improving customer satisfaction and thereby increasing sales. In social networks [4, 5], link prediction helps users to find people they may know but have not yet connected with. Also, in online marketplaces like a marketplace on Facebook, link prediction can be used to recommend products to users based on their network connections. In the criminal networks [6, 7], by leveraging the analysis of relationships and interactions between individuals, law enforcement agencies can uncover criminal activities like drug trafficking or money laundering and determine the connections between those involved in such illegal activities. In the security domain [8], we use link prediction techniques to assess the trustworthiness of individual nodes in the network. For community detection [9, 10] problem, using link prediction could detect a community in a network. This can help analyse the network's structure and function. Identifying missing links through link prediction is a valuable tool in network analysis for anomaly detection [11]. It can help uncover unexpected or abnormal connections in the network that may not be immediately apparent and identify potential anomalies that may indicate suspicious or malicious activity. Finally, in the study of biological networks, specifically in predicting protein interactions [12, 13], link prediction algorithms can be used to predict new interactions between proteins based on existing interactions, allowing researchers to generate hypotheses about the function of previously uncharacterized proteins.

Much research has been published on link prediction techniques, covering topics such as the application of link prediction and link prediction in complex networks, summarized in Table 1. Wang et al. [14] provided a comprehensive review of the current state-of-the-art in complex networks. Martinez et al. [15] also surveyed the development of link prediction algorithms. In contrast, Kumar et al. [16] focused on the most advanced algorithms and applications of link prediction in social networks, including similarity, probabilistic, and algorithmic methods. However, the previous survey did not consider graph neural networks, which have become increasingly prominent recently.

In recent years, the field of link prediction has witnessed a significant increase in research activities. Particularly, between 2021 and 2023, there is an abundance

Table 1 Characteristics of the reviewed literature reviews

Article	Year	Objectives and topics
Linyuan et al. [17]	2011	Summarize popular link prediction algorithms for complex networks
Wang et al. [14]	2015	Link prediction for social networks in a systematic manner, techniques, and challenges
Martinez et al. [15]	2016	Link prediction in complex networks
Kumar et al. [16]	2020	The most advanced algorithms and applications of link prediction in social networks

of publications, including over 400 articles on platforms like ScienceDirect. This heightened interest emphasizes the need for a comprehensive exploration of the latest advancements and approaches in link prediction.

Our review aims to conduct a state-of-the-art survey on link prediction research since 2018, focusing on the studies not covered in previous surveys and those utilizing graph neural networks (GNNs). Our survey aims to offer a fresh perspective on current approaches in link prediction, providing insights into fundamental problems and strategies while assisting researchers in understanding this field's developments and future directions.

This paper presents a comprehensive survey of the latest link prediction methods. It distinguishes itself from other surveys in two aspects. Firstly, we provide simplified explanations of each method using schematic representations. Secondly, we delve into current techniques, particularly in machine learning and deep learning, specifically focusing on graph neural networks. We analyse various articles, extract relevant databases, and identify employed measures. Furthermore, we highlight the strengths and limitations of each method to aid researchers in selecting the appropriate method and dataset for their specific research question. We also discuss the trend and gaps in link prediction.

We have classified the methods into four categories with subcategories based on their underlying techniques:

- Similarity-based methods
  - Community detection-based approaches
  - Random walk-based approaches
- Dimensionality reduction-based methods
  - Embedding-based methods
  - Matrix factorization-based methods
- Machine learning technique-based methods
  - Supervised learning approaches
  - Unsupervised learning approaches
  - Deep learning techniques
  - Graph neural network-based methods
  - Reinforcement learning
- Other methods

In order to provide a comprehensive overview of recent advancements in link prediction techniques, we formulated several sub-questions to guide our research. These sub-questions encompassed the main categories of link prediction techniques, the types of networks concerned by ink prediction, the performance metrics used for evaluation, and the future research directions and potential areas for improvements in the field.

To address these sub-questions, we conducted an extensive search using prominent online databases such as Scopus, SpringerLink, ACM Digital Library,

IEEE Xplore, and ScienceDirect. Our search focused on paper titles, and we utilized various search keywords related to link prediction, machine learning, deep learning, social networks, complex networks, dynamic networks, and temporal networks.

Based on our research and analysis of relevant articles, Table 2 lists the top Universities in countries interested in link prediction research.

The rest of the paper is structured as follows: In Sect. 2, we introduce the problem of link prediction, the datasets, types of networks, and the evaluation measures used for the link prediction algorithms. In Sect. 3, we delve into the link prediction process and examine the various approaches, exploring their strengths and limitations in Sect. 5, we discuss the trends and gaps observed in link prediction. Finally, we present the conclusion in Sect. 6.

2 Link prediction

In this section, we define the link prediction problem and give an overview of the datasets, the evaluation measures, and the type of networks used in link prediction algorithms.

2.1 Definition

Link prediction is a task in network analysis that aims at predicting the likelihood of future connections between nodes in a network. It uses the existing network structure and sometimes node attributes to estimate the presence or absence of potential links. This task is often applied to graphs G(V, E), where V represents the set of nodes as entities and E represents the set of edges as connections or links. The network's existing relationships help predict potential connections between unconnected nodes.

Consider the example in Fig. 1. If node A is linked to both nodes B and D, nodes B and D are not linked, and node A is a common link between B and D, then link prediction would suggest a potential connection between D and B, as well as C and D.

Table 2 Top Universities in countries of interest for link prediction research

Country	University
USA	Stanford University
China	University of Electronic Science and Technology of China
India	The Department of Science and Technology (DST)
USA	Carnegie Mellon University
France	CNRS Centre National de la Recherche Scientifique
Germany	Max Planck Institute for Informatics

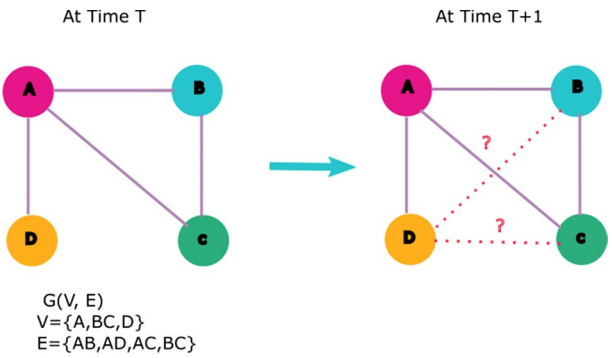

Fig. 1 Example of link prediction in network graph

2.2 Networks used in link prediction

There are several types of networks that are used in link prediction. The specific methods and techniques used for link prediction rely on the data type (network) and relationships being evaluated. The most prevalent networks employed in link prediction approaches are enumerated in this section and represented in Fig. 2.

In Table 3, we represent a compilation of different types of networks and the corresponding articles that have used these networks for link prediction.

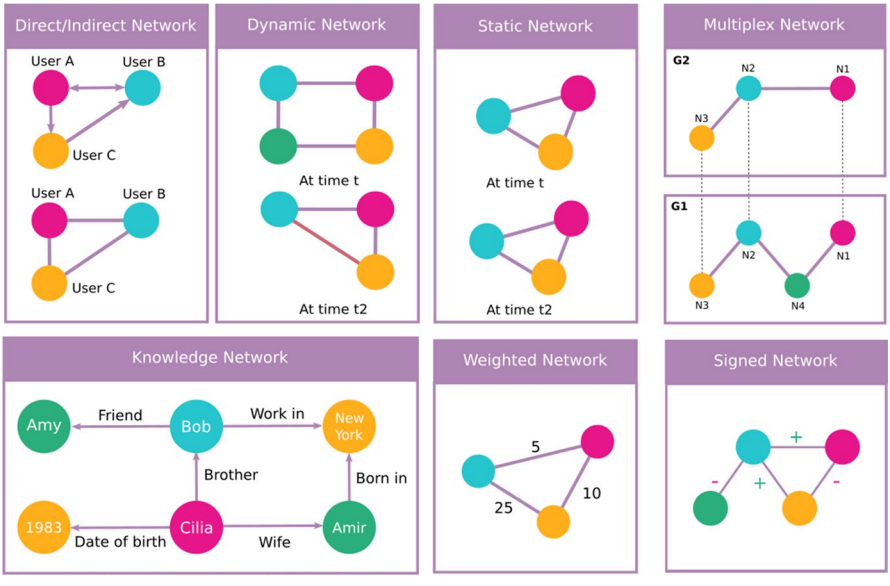

Fig. 2 Representation of different types of networks

Table 3 Type of the network with references

Type of networks	References
Ego network	[29–31]
Multiplex network	[24, 32–37]
Directed network	[38–43]
Sign network	[44, 45]
Dynamic network	[46, 47, 30, 40, 48–55]
Knowledge network	[27, 56, 57]
Static network	[39, 54, 58–61, 12, 62–68, 30, 58, 69–80, 7, 81–83]
Weighted network	[20, 84]
Heterogeneous network	[79, 85]
Undirected network	[64, 65, 86] [66, 70, 72, 73] [63, 64, 71]

2.2.1 Static network

A static network is a type of network that does not change over time, with nodes and edges remaining constant. The paper [18] presents a method for link prediction in static networks based on the popularity of nodes.

2.2.2 Dynamic network

A dynamic network is a kind of network that changes over time, with new nodes and edges developing or going away as the network progresses. A dynamic network is used in link prediction to represent relationships between nodes that change over time, such as interactions between people in a social network. The authors employ a dynamic network to identify missing links in [19–21]. The method used in [19] is the time-varying topological feature learning, which addresses the temporal dependencies among nodes and connections in dynamic networks.

2.2.3 Direct and indirect networks

A social network can be considered a direct or indirect network. A direct network is the asymmetric nature of links like Twitter. User A can follow user B or B following A. The indirect network is the symmetric nature of links like Facebook. The authors in [22] present an effective method for predicting directional links in directed networks.

2.2.4 Weighted network

A weighted network is a type of network where each link between nodes has a weight that can represent various aspects of the relationship, such as its strength

or importance. To predict links between nodes, link prediction algorithms often use weighted networks to capture their relationships. Recently, the paper [20] introduced a novel temporal link prediction model called GCN-GAN, specifically designed for predicting weighted links in dynamic networks. GCN-GAN is a nonlinear model that is well-suited for handling the complexities of weighted dynamic networks.

2.2.5 Multiplex network

A multiplex network is a network with many layers, where each layer represents a different type of relationship or interaction between nodes. Nodes are connected by multiple types of edges across different layers, capturing various dimensions of their associations. Multiplex networks provide a multidimensional representation of the network by considering diverse connections such as friendships, business partnerships, and familial ties. They offer a richer and more comprehensive understanding of complex social interactions in social networks. Najari et al. [23] presented an LP framework based on proximity-based characteristics and interlayer similarity (LPIS). Nasiri et al. [24] suggested a novel technique for predicting links in multiplex networks using topologically biased random walks (MLRW).

2.2.6 Knowledge network

A knowledge network is a multirelational graph composed of entities and relations regarded as nodes and different types of edges, respectively [25]. In [26], the authors performed a comparative analysis of knowledge graph embedding methods for link prediction. Tao et al. [27] tackled the problem of temporal link prediction in a dynamic knowledge graph (KG).

2.2.7 Signed network

A signed network is a type of network where edges have positive or negative weights. In a social network analysis, the positive weight of an edge indicates friendship between two nodes, and the negative weight of an edge indicates enmity between two nodes. A link prediction method is proposed in the work of Yuan et al. [28]. It leverages the structural information of signed social networks. The method involves comparing user similarity and utilizing the structural information of the signed social network to predict links.

2.2.8 Ego network

An ego network is a cluster of network nodes directly linked to a central node called the ego. The ego network encompasses the ego and its neighbouring nodes and their connections. The authors in [29] proposed an ELP algorithm that uses an ego network in a social network with relevant features based on the similarity approach.

2.2.9 Heterogeneous networks

Heterogeneous networks, also known as heterogeneous information networks or multi-modal networks, are characterized by multiple types of nodes and edges. These networks incorporate diverse entities, such as users, items, events, or concepts, where the edges represent different types of relationships or interactions, including friendship connections, co-occurrence, similarity, citation, and other relevant associations.

2.3 Popular datasets used in link prediction

Numerous datasets have been developed to support research in Link prediction for many areas, such as social networks, biology, and citation networks. In the following, we overview some of the most widely used datasets and summarize each dataset's key characteristics in Table 4.

- *The Zachary's Karate Club network* [87] is a social network of friendships among 34 karate club members at a US university.
- *Dolphin social network* [88] is a network of dolphins living in Doubtful Sound in New Zealand.
- *The Football network* [89] is a network of American football games played between Division IA colleges during the regular season in the fall of 2000.
- *The Jazz network* [90] is a collaboration network of 115 jazz musicians where a link between two musicians denotes music played by both in a band.
- *The USAir network* [91] is an airline network of US airports and their connectivity.
- *The Twitter dataset* [92] is a non-bipartite graph. It was made available as part of the 2020 RecSys Challenge, as described by Belli et al. (2020).
- *Ego-Facebook* [93] is a large dataset. This dataset comprises "circles" or "friends lists" extracted from Facebook.
- *Facebook NIPS network* [94] is a social network dataset made publicly available in 2012 by Julian McAuley and Jure Leskovc. This undirected and unweighted network represents user-to-user friendships on Facebook. Each user is represented by a node and an associated link between two nodes representing each friendship.
- *Email-Eu-core network* [95] is an email communication network generated by members of a large European institution. A node represents each user in the network, and an edge is established between two nodes if at least one email is generated by one of the two users.
- *Yeast* [96] is a well-known biological dataset that consists of a protein–protein interaction network in the budding yeast Saccharomyces cerevisiae.
- *The Power* [97] is a dataset representing the power transmission network covering the western region of the USA.

Table 4 Description of popular datasets used in link prediction

Dataset	Nodes	Edges	Dir/Undir	Type of networks	Paper	Link
Karate	34	78	Undirected	Static Network	[65, 66, 70] [63, 64, 71]	https://networkrepository.com/soc-karate.php
Dolphin	62	159	Undirected	Static Network	[59, 63, 65–68, 70]	https://networkrepository.com/soc-dolphins.php
Jazz	198	2742	Undirected	Static	[71, 86, 101]	http://konect.cc/networks/arenas-jazz/
Football	242	4100	Directed	Directed Network	[30, 63, 64, 67, 68, 72]	https://datahub.io/collections/football
USAir	332	2100	Undirected	Static Network	[64, 65, 72, 73, 86]	https://networkrepository.com/USAir97.php
Email-Eu-core	1005	25,571	Directed	Dynamic Network	[40, 43, 52, 58, 64–66, 71]	https://snap.stanford.edu/data/email-Eu-core.html
CollegeMsg	1899	T59835, S20296	Undirected	Dynamic Network	[52, 102]	https://snap.stanford.edu/data/CollegeMsg.html
Yeast	2375	11,693	Undirected	Static Network	[63, 64, 66, 70, 86]	https://humgenomics.biomedcentral.com/articles/10.1186/1479-7364-3-3-291
Facebook NIPS	2888	2981	Undirected	Ego Network	[58]	http://konect.cc/networks/ego-facebook/
The Ego-Facebook	4039	88,234	Undirected	Ego Network	[30, 31]	https://academictorrents.com/details/
Power	4941	6594	Directed	Static Network	[30, 63, 67, 68, 72, 73, 77, 103, 104]	Power
Twitter	1043	4860	Directed	Static Network	[33, 42, 47]	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5515441/
Twitter-Foursquare	10,989	16,104, 23,348, 11,459	Directed	Multiplex network	[33]	Twitter
Vickers	29	240, 126, 152	Undirected	Multiplex	[35, 36]	Figshare Link

- *CollegeMsg* [98] the dataset contains records of messages exchanged on an online social network at the University of California, Irvine. It includes information about the source user (sender), destination user (receiver), and timestamp of each message. The dataset covers a time period from April 2004 to October 2004. It provides insights into the communication patterns and interactions within the online social network during this duration.
- *Twitter-Foursquare* [99] is a dataset that combines data from both the Twitter and Foursquare platforms. It is collected between May and September 2012. It specifically focuses on three major cities in the USA. The dataset includes tweets and check-ins obtained from users who had checked in during that time and shared their check-ins on Twitter.
- *Vickers* [100] Dataset was collected by Vickers and it consists of data obtained from 29 seventh-grade students in a school located in Victoria, Australia. The students were asked to nominate their classmates based on several relations, focusing on three specific layers: The relationship layer, the Best friends layer, and the Working preference layer.

2.4 Evaluation metrics

In this section, we present the measures commonly used to evaluate the results of link prediction algorithms as shown in Fig. 3. The choice of evaluation measures depends on several factors, such as the method used and the types of datasets. Some measures may be more suitable for certain methods or datasets than others, and the choice of measures can also depend on the research objective. For instance, if the aim is to identify as many true positive links as possible, recall might be more significant than precision. Conversely, precision may be more crucial if the focus is on minimizing false positives.

- *Area Under the Receiver (AUC)* is the area under the receiver operating characteristic curve (ROC) [105], which is a measure of the model's ability

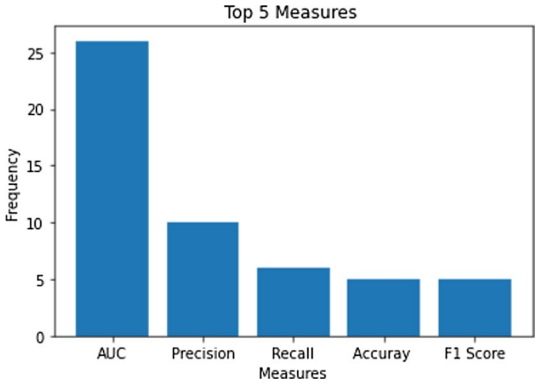

Fig. 3 Most commonly used measures among the papers included in our survey

to distinguish between positive and negative links. The formula to calculate the AUC score is shown in Equation (1), where n' is the number of times the missing links have scores greater than the non-existing links, and n'' is the number of times their scores are equal. Assuming that the scores are generated from an identical and independent distribution, the AUC score is around 0.5. A higher AUC value indicates a more efficient algorithm compared to random choice [106].

$$\text{AUC} = \frac{n' + 0.5n''}{n} \tag{1}$$

- *Accuracy* is a metric that indicates the percentage of correctly classified links by a link prediction model. It can be computed using the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where *TP (True Positive)* is the number of correctly predicted positive links. *TN (True Negative)* is the number of correctly predicted negative links.
FP (False Positive) is the number of incorrectly predicted positive links.
FN (False Negative) is the number of incorrectly predicted negative links.

- *Precision* is a metric used to evaluate the performance of a link prediction model. It measures the proportion of correctly classified positive links (i.e. links that actually exist) among all classified positive links.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

- *Recall* is the ratio of correctly classified positive links to the total number of positive links.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

- *F1 Score* is a measure of the balance between precision and recall.

$$\mathbf{F1} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5}$$

3 Methods for link prediction

We classify link prediction methods based on recent literature (2020–2023). We reviewed about 80 papers covering similarity heuristics, machine learning, dimension reduction, and other approaches.

3.1 Similarity-based methods

Similarity is a measure used to evaluate the level of similarity or connection between two nodes in a network. It is frequently used to predict the likelihood of a link forming between two nodes that are not currently connected in the network. The most similarity measures between nodes, used in link prediction studies, are local, quasi-local, and global indices. Local indices, such as Common Neighbours [107], Salton Index [108], Jaccard Index [109], Sorensen Index [110], Hub Promoted Index [111], Preferential Attachment Index [112], Adamic–Adar Index [113], and Resource Allocation Index [114], calculate scores based on the neighbourhood information of nodes with a path distance less than two. Global indices, like Katz [115], Random Walk [116], SimRank [117], Leicht-Holme-Newman Index (LHNI) [118], and Matrix Forest Index (MFI) [119], use information from the entire network to calculate scores with a path distance greater than two. Quasi-local indices, such as Local Random Walk [120] and Local Path Index [121], combine the advantages of local and global indices and calculate scores for nodes with a path distance of no more than two. These quasi-local approaches tend to have higher prediction accuracy compared to local methods.

In conclusion, these old metrics are adaptations of the common neighbour method that normalize or consider the significance of the neighbouring nodes to reduce biases resulting from uneven distribution of node degrees. These old measures are still being used in recent studies, often in combination with other methods. They are computed using the equations given in Table 5, where $\Gamma\, x$ and $\Gamma\, y$ represent the sets of neighbours of nodes x and y, respectively.

First, in [59], the authors introduce a new link prediction algorithm, CNDP (Common Neighbours Degree Penalization). The CNDP algorithm combines the network's average clustering coefficient and topological characteristics, such as the number of shared neighbours, to calculate the node similarity score. By including

Table 5 Most popular similarity-based metrics

Measure	Equation	References
Common Neighbours	$sCN_{xy} = \lvert\Gamma(x) \cap \Gamma(y)\rvert$	[107]
Jaccard Index (sJI)	$sJI(x,y) = \frac{\lvert\Gamma X \cap \Gamma Y\rvert}{\lvert\Gamma X \cup \Gamma Y\rvert}$	[109]
Salton Index (sSalton)	$sSalton(x,y) = \frac{\lvert\Gamma X \cap \Gamma Y\rvert}{\sqrt{\lvert\Gamma X\rvert \cdot \lvert\Gamma Y\rvert}}$	[122]
Hub Promoted Index (sHPI)	$sHPI(x,y) = \frac{\lvert\Gamma X \cap \Gamma Y\rvert}{\min(\lvert\Gamma X\rvert, \lvert\Gamma Y\rvert)}$	[111]
Hub Depressed Index (sHDI)	$sHDI(x,y) = \frac{\lvert\Gamma x \cap \Gamma y\rvert}{\max(\lvert\Gamma x\rvert, \lvert\Gamma y\rvert)}$	[111]
Leicht-Holme-Newman Index (sLLHN)	$sLLHN(x,y) = \frac{\lvert\Gamma x \cap \Gamma y\rvert}{\lvert\Gamma x\rvert \cdot \lvert\Gamma y\rvert}$	[118]
Adamic–Adar Index (sAA)	$sAA(x,y) = \sum_{w \in \Gamma x \cap \Gamma y} \frac{1}{\log(\Gamma w)}$	[113]
Resource Allocation Index (sRA)	$sRA(x,y) = \sum_{w \in \Gamma x \cap \Gamma y} \frac{1}{\Gamma w}$	[114]
Preferential Attachment (sPA)	$sPA(x,y) = \Gamma x \cdot \Gamma y$	[112]
SimRank (sSimRank)	$sSimRank(x,y) = \frac{C}{\Gamma x \cdot \Gamma y}$	[117]
Random Walk Index (sRandomWalk)	$sRandomWalk(x,y) = \frac{P(x,y)}{P(x) \cdot P(y)}$	[123]

common neighbours in the calculation, this method improves upon the previously proposed adaptive degree penalization (ADP) method. The results demonstrate that CNDP delivers substantial improvements in accuracy and computational efficiency compared to other similar techniques. Additionally, CNDP considers both local and global features of the network, providing a more comprehensive and nuanced examination of the network structure. In the work of Ahmad et al. [65], a novel algorithm called the Common Neighbour and Centrality-based Parameterized Algorithm (CCPA) was introduced for complex network link prediction. This algorithm is based on two key node characteristics: the number of common neighbours and the centrality of the nodes, which is calculated using closeness and betweenness centrality measures. The "common neighbour" feature reflects the shared nodes between two nodes in the network. The effectiveness of CCPA was evaluated through experiments on eight real-world datasets. The authors in [66] presented a novel similarity measure based on the local path information of nodes. The similarity score between two nodes is computed using the local information of nodes within a specified distance, which considers the direct connection between the nodes and all other shorter paths between them. This approach merges the strengths of two well-known similarity indices, the Katz Index and Adamic–Adar. The authors in [67] proposed a parameter-free new similarity metric based on degree and path depth. The metric extends their previous work (Jibouni et al. [68]) that focused on local features such as node degree and global features, including path structure. The metric considers different length pathways to reduce the impact of high-degree nodes by using path lengths 2 and 3 and considering the degree of the source and destination nodes. The authors applied machine learning techniques such as K-nearest neighbours, logistic regression, artificial neural network, decision tree, random forest, and support vector machine (SVM) for binary classification. In [69], the authors present a novel multidimensional network model incorporating multiple public opinion factors dimensions. The model is designed to capture the complexity of link formation in "We the Media" networks, where various factors, including structural information, occupational environments, interests, and social psychology, influence social user relationships. To evaluate the effectiveness of their model, the authors propose a link prediction algorithm specifically tailored for multidimensional network links. The algorithm is compared against baseline methods such as the Common-Neighbourhood-Driven model, the Jaccard index, and the SimRank method. The empirical analysis was conducted on Weibo.com public opinion data. In [36], the author explores link prediction in multiplex networks and proposes a multiple-attribute decision-making (MADM) approach to address the problem. By treating potential links in the target layer as alternatives and diverse layers in the network as attributes, the goal is to use information from all layers effectively. The proposed approach ranks alternatives based on ideality scores and assigns weights to different layers using a defined layer similarity measure based on cosine similarity. Extensive experiments demonstrate that the proposed method outperforms competing accuracy and running time approaches. In [53], the author proposed a model for growing networks and introduced novel time-sliced metrics to estimate the likelihood of missing links. These metrics, built upon established link prediction indices, exhibit superior performance compared to existing approaches, especially when the decay

factors are small. Additionally, the paper proposes function expressions for determining the optimal slice number and decay factor in real-world networks, enhancing the efficiency of link prediction. By leveraging these formulas, the method accurately and efficiently predicts missing links by estimating the ageing speed of growing networks. The implementation steps and advantages of the time-sliced metrics are exemplified using the preferential attachment metric, further highlighting their effectiveness in networks with ageing sites.

In the article [29], the authors proposed an ELP algorithm that used an ego network in a social network with relevant features based on the similarity approach. The features considered are Egocommon Neighbours, Ego Resource Allocation, Ego Page Rank, Ego Node2vec, and Ego Clustering Coefficient, focusing on Egocommon Neighbours (ELP). They found that this algorithm is more accurate than state-of-the-art methods for the same network. Similarly, the authors in [86] focus on the initial information contribution of nodes as a critical factor in prediction accuracy. By incorporating topological information and an adjustable parameter, the algorithm quantifies the significance of the node's initial information. By analysing bidirectional information transmission between nodes, the algorithm measures the structural similarity based on the total information amount received by nodes. Experimental evaluations on real-world networks demonstrate the superiority of the proposed algorithm compared to existing benchmark indices. Next, in [52], the authors presented an enhanced feature set that addresses the link prediction problem in dynamic networks by incorporating individual snapshots and the overall network structure. Using a reward and penalty structure, the novel cost-based feature for link prediction (CFLP) estimates edge behaviour throughout the entire network. Four categories of similarity indices are used to measure edge activity in individual snapshots. Feature selection is performed on fourteen different snapshot-based features to identify the optimal combination. The combined feature set is evaluated with machine learning models and outperforms state-of-the-art approaches. Experimental results on real-world datasets demonstrate the superior performance of the proposed method. A novel link prediction approach is introduced in [30]. It addresses spurious links in static networks and predicts removing existing links in dynamic networks. The approach is based on the concept of attraction force between nodes and node-level assignment. For static networks, connection probabilities of existing links are calculated to detect spurious links. In dynamic networks, a virtual network is constructed based on network changes, and connection probabilities are computed to predict link removal. The author in [35] presented a challenge of link prediction in a multiplex network. The proposed method combines edge and node relevance to accurately predict links between unconnected nodes. It used an aggregation model to summarize the information from different layers into a weighted static network, considering the density of the layers. Node relevance is determined based on the node's importance in the overall graph structure, while edge relevance considers local information. The method outperforms existing approaches for link prediction in both weighted static networks and multiplex networks. Finally, the authors in [38] proposed a topological nearest neighbour similarity method for directed networks. The method improves upon the Sorensen index and its variants to consider the directionality of network edges. The proposed

method effectively combines local and global information of network nodes by deriving the topological nearest neighbours similarity index based on the Global Local Hybrid Neighbours (GLHN) similarity index. Empirical validation using real directed network datasets demonstrates the superior performance of the proposed method in terms of lower error, higher accuracy, and stronger robustness compared to benchmark indices. Although the proposed method shows promising results, future work can focus on addressing computational complexity challenges, exploring more efficient algorithms such as deep learning, and further cleaning global node information to improve link prediction accuracy. Yuliansyah et al. [64] address the cold-start problem in link prediction when new users with limited information join a network. They propose a Degree of Gravity for Link Prediction (DGLP) approach, inspired by Newton's law of gravity, which considers the gravity of node pairs and common neighbours in a single-layer network. Evaluations using multiple datasets and benchmark methods demonstrate that DGLP significantly improves prediction performance, achieving higher AUC values and successfully predicting the cold-start problem in 99.94 % of node pairs. However, the performance of DGLP depends on the presence of cold-start problems in the network. The authors suggest future research directions, including incorporating non-topological information and exploring complementary combinations of node attributes and network structure information for predicting future relationships.

3.1.1 Similarity approaches using random walk

The connections between nodes can be modelled through the random walk [120], where transition probabilities are used to determine the path a random walker would take from one node to another. Various link prediction metrics use random walks to calculate the similarity between nodes.

Nasiri et al. [24] proposed a new approach for link prediction in multiplex networks using topologically biased random walks (MLRW). They explored these methods to determine the likelihood of a new connection in a target layer by considering interlayer and intralayer relationships in the network. Bias functions were used to calculate cross-layer weights between different layers, and these methods were tested using real-world datasets from various social networks. Next, the author of [70] introduced the Mutual Influence Random Walk (MIRW) algorithm for link prediction. In MIRW, the next node is selected based on its impact on the current node and the effectiveness of the path. The technique was evaluated on 11 real-world networks and compared with existing local, quasi-local, and global indices.

3.1.2 Similarity approaches using community detection

Community detection involves identifying groups of nodes in a network that are more closely connected to each other than to the rest of the network. For example, detecting groups of people with common characteristics such as university, age, country, or interests. This technique helps predict missing links in the network, as the link is likely between two nodes if they belong to the same community.

Using community detection as the initial step in link prediction methods can lead to more effective link representations than similarity-based methods. Wang et al. [63] implement a community information (CI)-based local similarity index for link prediction in large-scale networks using Spark GraphX and a parallelization approach. They develop a family of nine CI-based local similarity indices, including Common Neighbours (CN), Salton (Sal), Jaccard (Jac), Sørensen (Sor), Hub Promoted Index (HPI), Hub Depressed Index (HDI), Leicht-Holme-Newman Index (LHN), Adamic–Adar (AA), and Resource Allocation (RA). The authors combine the similarity algorithm with a local index and use the parallel BGLL algorithm to discover community networks in larger-scale networks. The model is evaluated on both small and large datasets. On the other hand, in [71], the authors proposed to solve the link prediction problem with a framework based on community detection and a similarity-based approach.

They proposed community relationship strength (CRS) as a measure to calculate the proximity of communities and included CRS with basic similarity indices. They used two traditional algorithms to detect the community: FastQ and Louvain. These two algorithms are based on modularity optimization. They used three traditional local similarity indexes: Common Neighbours (CN), Adamic–Adar (AA), and Resource Allocation Index (RA). They evaluated CRS-AA CRS-CN CRS-RA on twelve networks and compared its results to those of other link prediction methods. Singh et al. [72] present a new algorithm, Community Link Prediction via Information Diffusion (CLP-ID), for predicting missing links in networks. The algorithm is based on community detection and employs CD methods to determine the community structure of the networks using information diffusion. They use the IC model to integrate the information diffusion model. Then, they calculate the overall index of each existing link and the likelihood score of each non-existing link. Next, the authors proposed in [33] a new approach for link prediction in multiplex networks called Community-Guided Link Prediction based on External Similarity (CLPES). The authors use a more advanced MOEA/D-TS algorithm for community detection, which determines the network layer ordering and generates IP using the clustering coefficient metric. The proposed CLPES computes external similarity across network layers using a new external similarity metric (ExSim) and predicts internal links by considering various intralayer features, such as the Jaccard coefficient, preferential attachment, and Adamic–Adar while taking into account the community structure of the network. The likelihood of connection development between nodes is determined through the integration of these processes.

In many real-world networks, community identification has been found to be a useful strategy for link prediction, and its combination with other approaches can enhance performance.

Several studies have demonstrated that similarity-based link prediction approaches can achieve high accuracy. These approaches have the ability to capture the global structure of the network, which enhances its accuracy. In addition, these similarity approaches are both efficient and low-cost, making them practical and cost-effective solutions for link prediction. However, it is important to note that these approaches have limitations. The high time cost of the algorithms can make them impractical for large or complex networks, while their sensitivity to noise and

sparse networks can result in inaccurate predictions. In social network analysis, the similarity measure based on the number of common neighbours is often used as a first-order heuristic function to predict potential friendship relations. It has been shown to achieve satisfactory performance. However, this heuristic may not be effective when it is applied to protein–protein interaction networks. In these types of networks, two proteins sharing many common neighbours may still have a low probability of interacting, making this heuristic a non-reliable predictor in this context. Also, many similarity indices, such as Common Neighbour (CN) and Resource Allocation (RA) indices, are commonly used in link prediction for static networks. However, these methods have a limited ability to handle high levels of nonlinearity. Despite these limitations, the similarity-based approaches remain highly desirable options for many applications in network analysis due to their high level of accuracy and efficiency.

3.2 Machine learning

Machine learning has been widely used for various tasks, such as text and image classification [124, 125]. It has also been applied to the problem of link prediction in networks. A common approach in machine learning-based link prediction is using feature engineering to extract relevant features from the network data. These features are then used as input for a learning algorithm to predict the probability of a link forming between two nodes based on these features. We classify the approaches based on machine learning for link prediction into four categories: supervised, unsupervised, reinforcement learning, and deep learning.

3.2.1 Supervised learning

Supervised learning is a machine learning technique for binary classification problems like link prediction. In this approach, a graph (E) is transformed into data points (x, y), where each data point represents a pair of nodes. The data are labelled with a positive class label (+1) if a relationship exists between the nodes and negative class labels (− 1) if no edge (link) exists between them. The labels of x and y are defined as follows:

$$s(x,y) = \begin{cases} (x,y) \in V \\ +1 \qquad \text{if } (x,y) \in E \\ -1 \qquad \text{if } (x,y) \notin E \end{cases}$$

Figure 4 depicts the typical workflow in many studies that use machine learning for link prediction. First, we collect or select a correctly labelled database and partition it into training and testing datasets. Next, we extract the features from the network to describe each node pair. These features may include measures of similarity or other domain-specific features. To have a trained model, a suitable machine learning algorithm is selected and trained on the labelled training set to learn a function that maps the features of a node pair to a predicted probability of having a link between them.

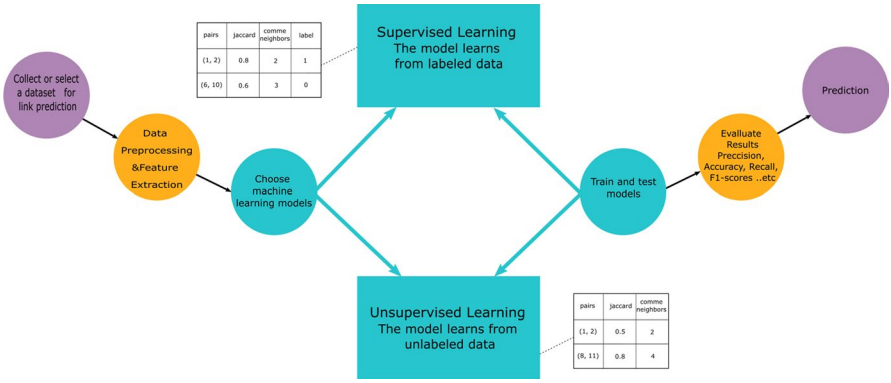

Fig. 4 Supervised and Unsupervised Learning Workflow Process

Finally, the test dataset is used to evaluate the trained model. Precision, recall, and AUC metrics evaluate the model's performance.

Support vector machine (SVM) approaches

Shan et al. [32] also applied SVM, random forest, and Adaboost algorithms to a multiplex network. This research employed hybrid approaches that combined different similarity measures, both local and global, to create a feature vector for node pairs. The feature vector included common neighbourhood, node degree, and clustering coefficient, as well as two new features, "neighbour friendship (FoN)" and "friendship in auxiliary layers (FAL)," which considered the structural information from all layers in the multiplex network. The feature vector was then fed into a classifier to predict the relationships between node pairs. The effectiveness of this method was evaluated by conducting experiments on six multiplex networks and comparing the results to NSILR [37] methods. This approach showed the benefits of integrating structural information from all layers and avoiding the issue of parameter setting. On the other hand, in [31], a binary classification approach was applied to the ego network in a social network. This approach extracts features such as age and location from user profiles based on the homophily theory. This approach uses similarity-based methods with various supervised learning algorithms, including adaptive boosting (AB), extremely randomized trees, gradient boosting (GB), K-nearest neighbours (KNN), linear discriminant analysis (LD), logistic regression (LR), Naive Bayes classifier (NB), neural network (NN), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XG). The proposed method reports the highest accuracy using the extreme gradient boosting method. The advantage of this method is that it uses essential similarity information extracted from user profiles, which cannot be deduced from the graph structure alone. However, it is worth noting that this information may not always be available or correct, which could impact the accuracy of the predictions. Kumari et al. [58] also employed machine learning algorithms such as support vector machine (SVM), K-nearest neighbours (KNN), decision tree (DT), random forest (RF), and artificial neural network (ANN) to predict links in network data by extracting features from the network's structural information. The study found that relying solely on a

similarity measure was insufficient for capturing link information in all networks. Therefore, network structure and individual features were considered for link prediction. The authors evaluated the models on real-world networks. They found that SVM performed better on the Dolphin, Facebook, and Hamsterster friendship datasets, while random forest had better accuracy on the email communication network. This approach's benefits include using machine learning methods to extract characteristics from network structure information. However, it is essential to note that this approach only considers the static and unweighted network for link prediction and does not consider content-based features.

Naive Bayes approach

In [44], the authors proposed the two motif Naive Bayes (TMNB) model in the sign network. The TMNB model extends the single motif Naive Bayes model (SMNB) by combining information from two different types of motifs and uses the maximal information coefficient (MIC) matrix to discover the relationship between the motifs. The proposed solutions were evaluated and shown to be superior to existing methods. The (SMNB) model considers the contribution of each neighbouring node or edge in sign prediction.

3.2.2 XGBoost classifier

This paper [75] presented the self-configured framework (SCF) integrated with spark for enhancing link prediction in large-scale social networks. The SCF autonomously configures the best settings based on the dataset size, workload, and cluster specifications. It utilizes the XGBoost classifier to predict the optimal number of executors per node. The framework demonstrates a 40 % reduction in prediction time and balanced resource consumption, efficiently using limited clusters. The SCF offers an advantage over manual configurations, improving performance and prediction quality without requiring extensive hardware setup.

One advantage of the supervised learning approach for link prediction in network analysis is its ability to leverage a wide range of features and representations for nodes and edges in the graph, leading to more robust and accurate models. Another advantage is the availability of several metrics, including precision, recall, F1-score, and ROC-AUC, for the performance evaluation of supervised learning models.

However, there are some drawbacks to using supervised learning for link prediction. One major issue is that building models through supervised learning require high computational complexity. Another problem is that social networks increase and the classification models can become outdated, requiring frequent updates. Additionally, obtaining a large labelled dataset is a challenging and time-consuming task. Furthermore, the quality and accuracy of the link prediction results depend significantly on the quality of features extracted from the network. However, extracting high-quality features from the network can be challenging and may require domain expertise and manual feature engineering.

In conclusion, despite its limitations, supervised learning remains a popular and the most used technology for link prediction, offering precise predictions and firm performance in various contexts.

3.2.3 Unsupervised learning

Unlike supervised learning techniques, unsupervised learning techniques use unlabelled data. Muniz et al. [54] proposed the unsupervised link prediction Contextual-Temporal-Topological (CTT) criterion based on a weighted concept. This weighted concept focuses on three weighting criteria: Temporal-Topological, Contextual-Topological, and Contextual-Temporal-Topological. The CTT criteria merge the three weighting criteria, differentiating it from other approaches based on these weighting criteria without combining them. Ghorbanzadeh et al. [41] proposed a method to solve the problem of two nodes having no common neighbours meaning that they are not always predictable to link in the future. They consider neighbourhood direction and the hub and authority of neighbours. The performance of this method was evaluated using both supervised and unsupervised prediction models and was compared to several widely used baseline methods such as Node2Vec, DeepWalk, LINE, and M-NMF.

K-means

Mavromatis and Karypis [62] applied popular unsupervised learning algorithms, K-means, using a graph representation learning method called Graph InfoClust (GIC). The K-means algorithms were used to compute clusters based on simultaneous mutual information maximization. The performance of GIC was compared against other unsupervised methods such as deep graph infomax, variational graph auto-encoders, adversarially regularized graph autoencoder, DeepWalk, deep neural network for graph representation, and spectral clustering.

3.2.4 Deep learning

Deep learning has garnered significant attention recently due to its effectiveness in solving various problems, including link prediction. The deep learning (DL) approach is an automated learning technique that uses neural networks to extract the best features from the structure and content information for link prediction. Unlike other supervised learning methods, DL overcomes limitations in feature extraction, as it can perform this task automatically.

Zhang and Chen [74] introduced the Weisfeiler-Lehman neural machine (WLNM), a deep learning approach for link prediction. The WLNM method utilizes a fully connected neural network to identify the local enclosing subgraphs around links that are highly correlated with link presence. By extracting these subgraphs as the training data, this technique has achieved impressive results in link prediction, outperforming other state-of-the-art methods. Afterwards, Zhang and Chen [73] proposed a novel link prediction method called SEAL. This later method addresses limitations of the previously proposed Weisfeiler-Lehman neural machine (WLNM) method [74]. SEAL improves upon WLNM using a graph neural network (GNN) instead of a fully connected neural network to learn graph structure features from local enclosing subgraphs. This enables SEAL to learn from latent and explicit node attributes in addition to subgraph topologies. To demonstrate the ability to unify various high-order heuristics, SEAL presented a α-decaying theory. The method was tested on eight datasets and compared with

different heuristics, latent feature techniques, and network embedding algorithms. Chen et al. [46] introduced an encoder–LSTM–decoder (E-LSTM-D) model for link prediction in dynamic networks. The key innovation of the method is integrating structural and temporal features into the same framework. The authors employ an encoder–decoder to automatically describe the network and use LSTM to capture the temporal evolution of the network by stacking and representing a sequence of graphs. The model's effectiveness was evaluated using a newly developed error rate metric. The results demonstrate that the E-LSTM-D model can effectively tackle the challenges posed by high dimensionality, nonlinearity, and sparsity in dynamic network link prediction, thanks to its encoder–decoder construction. On the other hand, Rossi et al. [47] proposed using temporal graph networks (TGNs) with deep learning techniques for dynamic link prediction. To balance parallel processing efficiency with the capability to learn from the sequentiality of the input, the authors suggest a novel training method: TGN. TGNs are able to memorize the long-term dependencies between nodes in the graph. The authors comprehensively studied various framework components and evaluated the speed and accuracy trade-offs. The proposed models' performance was compared to static and dynamic baseline models. In their study, Sankar et al. [49] presented a new neural network architecture called the dynamic self-attention network (DySAT). This approach captures a node's temporal dynamics and structural neighbourhoods in a combined self-attentional representation. The authors aimed to learn low-dimensional vectors that describe the structural properties of a node and its surroundings. To evaluate the performance of DySAT, the authors tested it on four real-world datasets. They found that it outperformed existing state-of-the-art static and dynamic graph embedding baselines. Keikha et al. [76] proposed the DeepLink framework for link prediction in social networks. The framework utilizes deep learning to extract target features from both the content and structure information of nodes. Firstly, a Word2Vec framework is used to learn the structural feature vector of nodes, and then a Doc2Vec algorithm is applied to learn the feature vector of content information for each node. Finally, the weight vectors of structure and content information are aggregated into a single vector. The effectiveness of DeepLink was evaluated on the Telegram and irBlogs networks and compared with other link prediction methods.

In [103], the authors address the limitation of information loss in layers of graph pooling in graph neural networks. They propose a solution by learning the features of the target link directly instead of extracting features from the whole enclosing subgraph. To facilitate this, the original graph is transformed into a graph line, which enables efficient feature learning. The proposed model is evaluated against the baseline techniques on 14 datasets. Zulaika et al. [84] proposed a link weight prediction Weisfeiler-Lehman (LWP-WL) method. Inspired by the Weisfeiler-Lehman neural machine, LWP-WL extracts an enclosing subgraph around the target link and applies a graph labelling algorithm to create an ordered subgraph adjacency matrix. Then, this matrix is fed into a neural network, which includes a convolutional neural network (CNN) layer with specialized filters designed for the input graph representation. Extensive evaluations demonstrate that LWP-WL

outperforms state-of-the-art methods in various weighted graphs. Additionally, an ablation study is conducted to showcase the performance improvement achieved by incorporating different features into the approach.

Deep learning in link prediction presents several key advantages. Firstly, deep learning models excel in capturing intricate patterns within network data by automatically extracting meaningful features, eliminating the need for manual feature engineering and reducing analysis efforts. They demonstrate remarkable scalability, capable of handling large-scale networks with millions or billions of nodes and edges, enabling the analysis of massive datasets. Moreover, deep learning models effectively capture nonlinear relationships between nodes, enabling accurate predictions within complex network dynamics. Additionally, using techniques like long short-term memory (LSTM) enables the modelling of the temporal evolution of networks by stacking and representing sequences of graphs. This empowers the accurate prediction of links by considering the evolving relationships among nodes over time.

However, it is important to consider certain drawbacks. Deep learning models typically necessitate a substantial amount of labelled data for effective training, which can be challenging to acquire. Training deep neural networks can be computationally intensive, requiring significant computational resources. The interpretability of deep learning models is limited, posing challenges in understanding the rationale behind their predictions. Overfitting is a potential concern, particularly when working with limited or imbalanced training data. The effectiveness of deep learning models heavily relies on diverse and representative training data, as biased or limited data can result in poor generalization.

In summary, deep learning in link prediction offers advantages in capturing complex patterns and scalability. However, it is crucial to address challenges such as data requirements, computational complexity, interpretability, overfitting, and the need for diverse training data for successful application in link prediction tasks.

3.2.4.1 Graph neural network (GNN) A graph neural network (GNN) is a machine learning model that applies an optimized transformation to all graph attributes, including nodes, edges, and global-context information. This transformation is designed to preserve graph symmetries, such as permutation invariance, which are important for accurately representing and analysing graph data.

In the field of deep learning, graph neural network (GNN) are a type of method specifically designed to operate on graph structured data. GNNs aim to effectively combine the feature information and graph structure to learn better node representations through feature propagation and aggregation. GNNs aim to learn a low-dimensional vector representation for each node in the graph. "The goal of GNNs is to iteratively update the node representations by aggregating the representations of node neighbours and their own representation in the previous iteration" from the book Graph Neural Networks: Foundations, Frontiers, and Applications [126].

Graph neural networks (GNNs) can be seen as a method for learning node embeddings by iteratively combining information from a node's local neighbourhood. In each iteration, the first information is learned about the direct neighbours, followed

by the neighbours of the neighbours, and so on. There are various types of GNNs, each with its own unique update and aggregation functions. The three most common tasks GNNs perform are downstream graph analysis and prediction at the node, edge, and graph levels. A simple representation of the GNN process is depicted in Fig. 5.

Several techniques for neighbour embeddings in Graph Neural Networks (GNNs) have been proposed. Kipf and Welling present the first notable work in this area in [127], where they proposed aggregating neighbour information as a normalized sum of states, incorporating the update operation into this aggregation through the addition of a self-loop for specific nodes. Another popular technique is using a multi-layer perceptron (MLP) for aggregation [128]. This involves using a feed-forward network to perform the aggregation operation, where the weights can be optimized to achieve the best aggregation of neighbouring states. Gated graph neural networks [129] use an attention mechanism in the aggregation process, considering the importance of the neighbouring node's features. This results in an updated embedding that contains more information about the neighbour's features. These networks use a recurrent unit to update the state iteratively over time.

In a new study presented by Cai and Ji [77], a novel node aggregation method was proposed to transform the enclosing subgraph into different scales while preserving information. The multi-scale approach used the subgraphs at different scales to extract graph structure features, and it was evaluated on 14 datasets from various fields. In the field of heterogeneous graph analysis, a new model named metapath aggregated graph neural network (MAGNN) was proposed in [79]. This model addresses three limitations of previous models, including the omission of node content features, the discarding of intermediate nodes along the metapath, and the consideration of only one metapath. MAGNN combines both structural and semantic information from neighbour nodes and metapath context to generate final node embeddings through the use of an attention mechanism for intra-metapath aggregation. The effectiveness of MAGNN was evaluated on three real-world heterogeneous graph datasets for tasks such as node clustering, link prediction, and node classification. In [80], the authors presented a new approach to graph analysis by representing the problem as a graph convolutional network. They proposed a novel multi-level graph convolutional network (MGCN) to

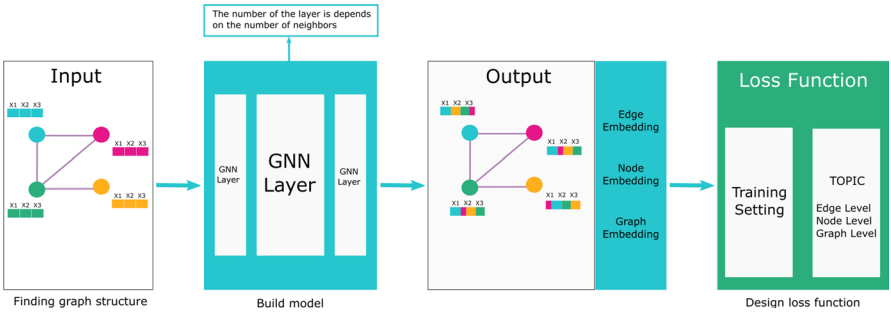

Fig. 5 Graph neural network workflow process

uncover the embeddings of each network. The authors introduced a new paradigm integrating multi-level graph convolutions into the local network structure and the hypergraph structure. They presented several solutions, such as network splitting and space reconciliation, to manage the dispersed training process and make their proposed framework scalable for handling large-scale social networks. The paper [130] introduced CPAGCN (Community Preserving Adaptive Graph Convolutional Networks), a novel method for link prediction in attributed networks. CPAGCN effectively combines attribute information and link information by utilizing the AGCN (Adaptive Graph Convolutional Networks) algorithm for network embedding. It incorporates a community detection model to preserve the community structure within the node representations. The link prediction module employs multi-layer perception (MLP) to learn prediction scores for potential links directly. Experimental results on real-world attributed networks demonstrate that CPAGCN outperforms state-of-the-art methods in link prediction. In [48], the authors address the challenge of link prediction in complex dynamic networks by proposing a novel approach called STEM-GCN. STEM-GCN is a gated graph convolutional network incorporating spatiotemporal semi-variogram analysis to capture spatial and temporal correlations. It introduces a correlation smoothing strategy to enhance prediction accuracy and reduce noise. STEM-GCN effectively captures network dynamics between consecutive time steps using stacked memory cell structures. Experimental results demonstrate that STEM-GCN outperforms existing methods, showcasing its potential to uncover evolving mechanisms of real-world dynamic networks. In [57], the authors proposed ComplexGCN, a novel extension of graph convolutional networks (GCNs) in complex space for knowledge graph embedding. ComplexGCN utilizes complex-valued embeddings and paratuck2 decomposition-based scoring function to enhance representation quality and predict missing links in knowledge graphs. The model outperforms existing methods on standard link prediction datasets. It introduces complex graph convolutional layers and residual connections to preserve initial embedding information. The paper referenced by [131] introduces DATGN, a novel model for link prediction in dynamic graphs. DATGN addresses existing methods' limitations by considering the nodes' local time-space environment. It utilizes an activity-based sampling algorithm and a global attention network to aggregate global information. Additionally, a local attention network aggregates information from sampled sequences. Experimental results show that DATGN outperforms state-of-the-art models' accuracy and efficiency, particularly in the inductive link prediction task. The local spatial-temporal network layer captures evolutionary patterns, improving link prediction accuracy. The study presented in [56] introduces LCILP, a novel strategy for inductive link prediction in knowledge graphs. LCILP used a Personalized PageRank (PPR)-based local clustering technique to sample tightly related subgraphs around target links, improving the capture of meaningful local context. The approach employs a GNN-based model for reasoning over the extracted subgraphs. Experimental results demonstrate the superior performance of LCILP compared to state-of-the-art models on three benchmark datasets. The study also explores the relationship between graph properties, such as average clustering coefficient and average node degree, and the effectiveness of link prediction. Huang and Lei [50] proposed a novel approach

called temporal group-aware graph diffusion network (TGGDN) for dynamic link prediction. The TGGDN incorporates a group affinity matrix to capture mutual interactions between neighbours and integrates it into the graph diffusion process to simultaneously capture group and long-distance interactions. Additionally, a transformer block with self-attention is employed to model the temporal information and enhance interpretability. Experimental results on real-world datasets of varying sizes demonstrate that TGGDN outperforms state-of-the-art methods, achieving significant improvements in terms of accuracy (ACC) and area under the curve (AUC). The proposed method shows promise in dynamic link prediction tasks, and future work aims to address scalability for large-scale dynamic networks. Zhang et al. [81] proposed a novel approach called IEA-GNN (Anchor-aware graph neural network fused with information entropy) to address the limitations of existing methods for capturing global location information and distinguishing located nodes in graphs symmetrically. IEA-GNN calculates the information entropy of nodes and constructs candidate sets of anchor points to capture the relative distance information between nodes. It uses a nonlinear distance-weighted aggregation learning strategy based on these anchor points to enhance node feature information and improve discrimination between homogeneous neighbourhood nodes. The model avoids aggregating anchor points and highlights positional differences by selecting anchor points based on information entropy. Experimental results on multiple datasets demonstrate that IEA-GNN outperforms baseline models in node classification and link prediction tasks. However, the model's performance may be affected when nodes are over-aggregated, an aspect that could be addressed in future research to enhance its generalization ability. Zhao et al. [85] introduce an end-to-end link prediction method for heterogeneous networks, leveraging metapath projection and semantic graph aggregation. This approach learns node pair embeddings from different metapaths, projecting the network into multiple semantic graphs and employing a graph neural network. A semantic aggregation module combines node pair embeddings using an attention mechanism. Experimental results demonstrate the method's superior accuracy compared to competing approaches.

Graph neural networks (GNNs) offer significant advantages in link prediction due to their ability to capture and leverage the underlying structural information of a graph. GNNs excel at extracting complex features from the graph structure, enabling accurate predictions. For example, Cai et al. [77] introduced a multi-scale approach using subgraphs at different scales to extract graph structure features, resulting in improved link prediction performance.

Another advantage of GNNs in link prediction is their ability to handle sparse graphs effectively. Traditional methods often struggle with sparse graphs due to the reliance on explicit features for each node and edge. However, GNNs can utilize the graph structure itself as an implicit feature, making them particularly effective in sparse graph scenarios. This advantage has been demonstrated in various studies, such as the work by Huang and Lei [50], where the temporal group-aware graph diffusion network (TGGDN) successfully tackled dynamic link prediction tasks on real-world datasets.

Despite these advantages, GNNs also have limitations in link prediction. Scalability to very large graphs is one such limitation, as the computational and

memory requirements of GNNs can become prohibitive. This challenge has been addressed in studies like [80], where the authors proposed a multi-level graph convolutional network (MGCN) to manage the dispersed training process and enhance scalability for large-scale social networks.

Another limitation is the risk of overfitting on small and noisy datasets, which can negatively impact link prediction performance. To mitigate this, techniques that incorporate temporal information have been introduced. For instance, STEM-GCN, proposed by the authors in [48], integrates spatiotemporal semi-variogram analysis to capture spatial and temporal correlations, resulting in improved prediction accuracy and reduced noise in dynamic networks.

In summary, GNNs offer advantages in capturing structural information and handling sparse graphs in link prediction. However, scalability to large graphs and the risk of overfitting should be carefully considered. Incorporating temporal information can further enhance link prediction accuracy in dynamic networks.

3.2.5 Reinforcement learning (RL)

RL is a method used in machine learning where programmes, called agents, interact with both familiar and unfamiliar environments while continually adapt and learn depending on the grants received as points (positive or negative) for their job performance. These points can be positive or negative, and they are, respectively, branded as rewards or penalties [132]. In link prediction, an agent can be taught to explore a network, collect information about nodes and edges, and then use that data to predict the links that are likely to exist. The agent receives rewards when it anticipates links correctly, while it receives penalties when it makes errors. The problem with the use of machine learning techniques such as support vector machine and Naive Bayes is the dependency on the availability of large datasets for training purposes. For this reason, Lim et al. [7] proposed a deep reinforcement learning (DRL) based criminal network link prediction model that was compared to a gradient boosting machine (GBM) in machine learning, using a relatively smaller dataset. The DRL model transformed a graphical dataset into a feature matrix for each pair of nodes and used Jaccard, Adamic, and Adar as the layer input, with the SNA metrics of the hub index and preferential attachment index functioning as weights for the hidden layer. The results showed that the DRL model achieved better AUC (0.85, 0.82, and 0.76) scores than the GBM models of JUANES, MAMBO, and JAKE. Later, the same authors [55] applied a reinforcement learning algorithm to a temporal dataset for a criminal network and found that the time-based link prediction model (TDRL) has higher prediction accuracy than the previous DRL model. In another study [133], two methods were compared for link prediction in a small dataset produced through self-simulation. One method used a model without a metadataset (CNA-DRL), while the other used a metadataset (MCNA-DRL) to improve performances. The results showed that the MCNA-DRL model achieved an AUC score of 79 %, while the CNA-DRL model achieved a score of 70%. Tao et al. [27] also tackled the problem of temporal link prediction in a dynamic knowledge graph (KG) using reinforcement learning models and presented a novel policy network that could learn predictable evolutionary patterns. An updating system was

also suggested to allow the agent to adapt to KG changes without fresh training. Next, the authors in [134] proposed a link prediction approach called RLPath, combining representation learning and reinforcement learning-based attentive relation paths. The approach obtains relation routes through reinforcement learning using a Markov decision process (MDP) model with two agents: a relation agent for choosing relations and an entity agent for choosing entities, which are represented as the policy network. The comparison results with the most advanced link prediction techniques demonstrate the competitive performance of RLPath.

3.3 Dimensionality reduction

In the context of link prediction, reducing the dimensions of graphs is an important task. High-dimensional data can be challenging to process and analyse, especially large-scale datasets. By reducing the dimensions, the data can be simplified, making it easier to work with. This can lead to faster and more efficient processing and improved model accuracy. This section summarizes the three main approaches for reducing the dimensions of graphs. The first approach, known as embedding methods, involves mapping the graph's information to a lower-dimensional space while preserving its structural information and components. The second approach is matrix factorization, where the graph is represented as a matrix and then factorized to obtain the node embedding vectors. The third approach, graph neural networks, also focuses on reducing the dimensions. It was already presented in the machine learning Sect. 3.2, allowing for a comprehensive explanation of all machine learning techniques.

3.3.1 Embedding-based methods

The network embedding approach converts network nodes into low-dimensional vectors, retaining their neighbourhood structures and allowing for learning informative features from the graph. To be more precise, the primary objective of node embedding techniques is to transform the original high-dimensional node representations into lower dimensions such that nodes that exhibit similar features in the original network are mapped into close proximity in the embedded space. In [42], a framework for joint link prediction and network alignment is presented with the aim of improving the recall for both tasks. A cross-graph embedding technique based on structural and topological neighbours was developed to effectively embed nodes from separate graphs. This approach is based on random walks, and a new formula is proposed to support network alignment. The results of the experiments show that the proposed approach outperforms other state-of-the-art methods in terms of link prediction accuracy and network alignment quality. One advantage of this work is its applicability to network alignment in scenarios where the degree scatter plot is narrow or attribute information is unavailable. The study presented in [135] proposes a novel link prediction method for social networks. It addresses the complexity of network features. The method leverages network embedding to represent the network structure as low-dimensional vectors, capturing spatial

relationships and user relevance. Additionally, word embedding models are employed to convert user text into vectors, considering the diversity and complexity of text semantics. To account for the dynamic nature of user behaviours, a time decay function is applied to quantify the impact of user text on link establishment. To simplify the complexity, the top-k relevant users are selected for each user. Moreover, an attention mechanism is introduced in a convolutional neural network to enhance the expression of user interests in text information. By integrating and analysing structural and text features, the proposed method achieves the objective of accurately predicting links in social networks. In two articles, the DeepWalk method is modified for the purpose of link prediction. The first article Nasiri et al. [12] focuses on protein–protein interactions, while the second [78] focuses on social networks. Nasiri presented a novel approach for link prediction in protein–protein interactions using embedding-based methods. The author proposes using the DeepWalk algorithm, a graph embedding method that utilizes Random Walking as a similarity measure, to address the nonlinearity issue. The DeepWalk algorithm is modified with a feature selection-based approach to generate a graph embedding. The results are compared with those obtained using other embedding methods such as node2vec, DeepWalk, Line, and GraphSAGE. The authors in [45] proposed a model for signed graphs, where positive and negative links provide insights into user associations. The authors tackle the challenges of imbalanced class distribution and hidden community structures often overlooked in existing methods. They propose an ensemble framework called ϵNeLp, consisting of network embedding and classifier prediction phases, to leverage hidden network communities for predicting negative links. The framework incorporates techniques such as community generation, optimization, probabilistic network embedding, and classifier prediction. Extensive experimental evaluation demonstrates the promising performance of ϵNeLp in terms of pertinency, robustness, and scalability. Barracchia et al. [102] proposed the LP-ROBIN method, which used incremental embedding based on a random walk to capture network dynamism and predict new links. LP-ROBIN can handle the addition of new nodes over time without prior knowledge. When new data arrives, LP-ROBIN updates the model by learning the embeddings of new nodes and links while updating the latent representations of existing ones. Experimental results demonstrate that LP-ROBIN achieves superior performance in terms of AUC and F1-score, outperforming baselines, static node embedding methods, and state-of-the-art dynamic node embedding methods. The paper referenced by [61] introduces NNWLP, a method based on network embedding that utilizes natural nearest neighbours to improve link prediction accuracy. Unlike existing methods that select neighbour nodes with equal probability, NNWLP leverages the network features to find the nearest neighbours. The clustering coefficients are employed to assess the contribution of nearest neighbour nodes and direct neighbour nodes, generating node sequences and forming node vectors. These node vectors are then converted into edge vectors and used for link prediction. Experimental results demonstrate that NNWLP effectively utilizes neighbour information and significantly enhances link prediction accuracy. In the paper referenced by [101], the authors introduce LRNP (low-rank network projection), a novel link prediction algorithm. LRNP is designed by leveraging optimal interactive coefficients derived from solving the objective

function, and it uses the adjacency matrix of a fully connected network as the base matrix to capture local structures in observed networks. Experimental findings demonstrate that LRNP surpasses existing state-of-the-art methods regarding link prediction accuracy. In [136], the authors introduce Rotate4D, a novel model for knowledge graph embedding that performs 4D rotations in quaternion space using a special orthogonal group. By embedding entities in quaternion space and applying rotations, the model improves link prediction performance on standard datasets compared to existing models. The Rotate4D model utilizes group theory and quaternion scaling to represent rotations and handle hierarchical relations efficiently. Experimental results demonstrate significant improvements in various evaluation metrics, such as MRR and Hits@k. The paper suggests further exploring group-like structures, and their combination with neural networks can enhance link prediction in quaternion and octonion spaces. With a line graph, Zhang et al. [104] proposed the line graph contrastive learning (LGCL) method for link prediction tasks. LGCL addresses information loss and limited generalization in similarity-based approaches by leveraging h-hop subgraph sampling and line graph transformation. The link prediction task is transformed into a node classification task using graph convolution, enhancing edge embedding learning. A cross-scale contrastive learning framework is introduced to maximize the mutual information between line graphs and subgraphs, integrating structural and feature information. Experimental results demonstrate that LGCL outperforms existing methods, offering better generalization and robustness.

The random walk-based embedding approach for missing link prediction presents several advantages, including its simplicity and computational efficiency, which makes it a suitable method for large-scale networks. The approach has been found to perform well in homophilic networks where nodes are linked to similar ones and can capture nonlinear relationships. However, a challenge for this approach is its accuracy in predicting links for nodes with high centrality. These nodes tend to have complex connectivity patterns and numerous connections to other nodes.

3.3.2 Factorization-based methods

Factorization-based methods are a commonly used set of techniques for link prediction. These methods use matrix factorization, a technique of dividing a matrix into smaller matrices, to reflect the latent representations of nodes in a graph. With these representations, predictions of missing links can be made efficiently. Chen et al. [43] proposed two approaches for link prediction, NMF-AP [43] and MS-RNMF [137]. NMF-AP combines the information from both local and global network structures by using PageRank to determine the impact score of nodes, which reflects the whole network structure, and the asymmetric link clustering coefficient approach to obtain local information. The performance of NMF-AP was evaluated on ten networks and compared with other methods. NMF models with a single-layer map the original network and its corresponding low-dimensional latent space. However, the technique is limited in uncovering hidden multi-layer information in complex networks such as biological systems, social networks, and transportation networks, which contain hierarchical information with

implicit lower-level features. On the other hand, MS-RNMF uses the heat kernel method to measure local similarity and the k-medoids algorithm to capture global information about the network structure. The nonnegative matrix factorization model in MS-RNMF is regularized using manifold regularization and sparse learning techniques, which reduces random noise and spurious links. This model was tested on seven networks using various measurements and parameters, and the results showed better performance than other methods, especially for weak sparse networks and susceptibility to random noise. Additionally, Chen et al. [82] introduced the fusing structure and sparsity constrained via deep nonnegative matrix factorization (FSSDNMF) approach to address this issue. To extract the topological details of each hidden layer, the common neighbour approach was employed to compute similarity scores and translate them into a multi-layer low-dimensional latent space. The authors eliminated random noise by jointly using the l2,1-norm restricted factor matrix at each hidden layer. In this article [60], the authors address the challenge of cold-start link prediction, where the network structure contains isolated nodes. They propose a multi-nonnegative matrix factorization model that integrates three types of information: community membership, attribute similarity, and first-order structure characteristics. The proposed model successfully predicts missing edges in the disconnected network structure by leveraging these multiple perspectives. This article [40] also presents a novel approach called GNMFCA (graph regularized nonnegative matrix factorization algorithm) for temporal link prediction in directed temporal networks. The proposed algorithm incorporates both local and global information of temporal networks by utilizing PageRank centrality and asymmetric link clustering coefficient. Graph regularization is employed to capture local information in each network slice, while PageRank centrality measures the importance of nodes, capturing global information. By jointly optimizing these factors in a nonnegative matrix factorization model, the GNMFCA model simultaneously preserves local and global information. The paper also introduces effective multiplicative updating rules for solving the model and provides a convergence analysis of the iterative algorithm. Experimental results on artificial and real-world temporal networks demonstrate that the GNMFCA outperforms existing temporal link prediction algorithms. Yang et al. [34] introduced a novel anchor link prediction method called multiple consistency (MC), which leverages interlayer and intralayer structures for improved performance. The MC method iteratively uses interlayer structure information and employs matrix factorization-based network representation learning to capture the global structural features of nodes. It further trains a radial basis neural network as a mapping function to align embedding vectors from different spaces. The method predicts anchor links between node pairs by considering interlayer and intralayer structures. Experimental results demonstrate the superiority of the proposed approach over existing methods. Agibetov [39] introduced an enhanced approach to learning neural graph embeddings by incorporating information from unlikely node pairs, addressing the limitation of traditional methods that truncate such information. Through experiments on various networks, the proposed approach demonstrates significant improvements in link prediction performance compared to baseline methods. The research sheds light on the relationship between skip-gram powered neural graph embeddings and matrix

factorization, revealing that the accuracy of graph embeddings in link prediction depends on the transformations applied to the original graph co-occurrence matrix. Notably, smoothening low-frequency pair entries instead of truncating them leads to better performance. The findings contribute to a deeper understanding of graph embedding algorithms and offer insights for designing future approaches based on matrix factorization.

Factorization matrix methods can enhance network structure analysis by removing random noise and identifying multiple link types, improving prediction accuracy. Integrating global and local structure information also helps reduce the impact of random noise. However, the limitation of using matrix factorization for link prediction is that it heavily depends on accurately representing the observed network through a low-rank matrix, which may not be feasible for networks with complex structures. Furthermore, the technique requires significant computational resources and may result in overfitting if it is not regulated properly, particularly in large-scale networks. These challenges make matrix factorization difficult for link prediction in real-world dynamic networks.

3.4 Other methods

In this section, we summarize the articles using different methods and approaches. In [138], the authors proposed a quantum algorithm QLP designed for path-based link prediction in diverse networks. QLP encodes prediction scores for both even and odd-length paths using a controlled continuous-time quantum walk. Through classical simulations, it demonstrates comparable performance to established path-based predictors. The proposal highlights the potential of QLP to achieve a quantum speedup in link prediction, distinguishing it from conventional methods by utilizing quantum computing techniques for calculations and predictions. Kumar et al. [139] proposed a novel strategy for link prediction using quantum kernel-enhanced machine learning models that incorporate local and global information for feature generation. The aim is to develop a quantum-assisted feature-based approach that combines projected quantum kernel (PQK) with machine learning models to improve prediction performance. By leveraging high-dimensional Hilbert spaces and a mathematical structure similar to quantum mechanics, the proposed approach enhances data for more accurate link prediction. Experimental results show that the quantum-enhanced models, such as PQK-neural networks and PQK-random forest classifier, outperform the corresponding classical machine learning. Comparative analysis of dynamic datasets demonstrates the superiority of the quantum-assisted methodology over individual link prediction approaches and state-of-the-art methods. The article [51] focused on temporal link prediction (TLP) and the need for high-accuracy white-box methods to explain network evolution mechanisms. Existing black-box models, such as network embeddings and graph neural networks, provide high prediction accuracy but lack interpretability. To overcome this lack, the authors propose the Develop-Maintain Activity Backbone (DMAB) model, which considers node dynamics at a microscopic level to predict future links. The DMAB model extracts and quantifies two dynamic properties of nodes: activity and loyalty. Comparative experiments with

state-of-the-art black-box methods demonstrate the excellent prediction performance and ability of DMAB to capture network evolution mechanisms. The study highlights the effectiveness of considering node dynamics in understanding temporal networks' dynamic link generation process. It emphasizes the need for further exploration of temporal network evolution mechanisms. Singh et al. [140] introduced a fuzzy-based link prediction algorithm, FLP-ID, designed to address the challenges of accuracy and efficiency in growing and multiplex social networks. FLP-ID considers critical factors such as different interaction channels, information diffusion, and group norms to form new connections. The algorithm generates a multiplex network by combining various relationship types and identifies the community structure. It computes node and relative relevance based on fuzzy criteria and group norms. By calculating the likelihood score of each non-existing link, FLP-ID predicts missing links with improved accuracy compared to crisp algorithms. Zheng et al. [141] proposed an explainable friend link recommendation method that leverages fusion embedding of heterogeneous context information. It integrates user content interests and external knowledge semantics to develop a fusion user embedding method. Using collaborative neighbourhood attention mechanisms, the method calculates direct and indirect similarity relationships between user pairs. It also incorporates a hybrid personalized and neighbour attention model for friend link prediction. The proposed method predicts users' friends and explains the link prediction results. In their study, [142] proposed a novel approach to improve link prediction accuracy by combining different link prediction methods. They comprehensively analysed the hybrid method and introduced the Precision-Noise Ratio (PNR) metric to evaluate the accuracy of uncertain link predictions. They developed a scalable and parameter-free algorithm based on posterior Bayesian estimation to combine different methods. The results showed that the PNR-based combination outperformed traditional combination methods regarding prediction accuracy. Additionally, the proposed approach offered a general and efficient framework for integrating various existing link prediction methods without increasing computational complexity. The effectiveness and efficiency of the approach were validated through extensive experiments on real datasets. This research provides valuable insights and opens opportunities for enhancing link prediction systems using a combination of existing methods. The article [83] introduced a novel causal model called causal lifting for link prediction tasks with path dependencies, where the outcome of link interventions depends on existing links. Existing causal models are unsuitable for such scenarios due to challenges in identifying causal effects or requiring many control variables. Causal lifting addresses this by allowing the identification of causal link prediction queries using limited intervention data. The article also investigates using structural pairwise embeddings, which offer lower bias and better represent the causal structure than traditional node embedding methods like GNNs and matrix factorization.

4 Discussion

This section comprehensively compares link prediction methods, encompassing similarity-based approaches, machine learning techniques, deep learning models (graph neural networks), and dimensionality reduction methods.

Similarity-based methods, which solely rely on the network's structure, offer simplicity and interpretability. They excel in scenarios where nodes with similar network neighbourhoods are likely to form links. For instance, users with common friends may be likelier to become friends on social networks. The input to similarity-based methods includes the network's adjacency matrix or edge list, which represents the connections between nodes in the network. The output of these methods is a similarity score or ranking that indicates the likelihood of a link between pairs of nodes. However, these methods may not fully capture complex global network dynamics, limiting their applicability in some cases.

Machine learning techniques provide greater flexibility by incorporating network structure, node attributes, and contextual information. The inputs to machine learning methods include the network's adjacency matrix or edge list and any available node attributes or features. These attributes can be nodes' characteristics, like age or location, used to enrich the information for link prediction. By leveraging this additional information, machine learning methods can capture more intricate relationships, making them well-suited for diverse networks. The output of these methods is a predictive model that can be used to infer potential links between nodes in the network. However, these methods often require a considerable amount of labelled data for training the predictive model, which may not always be available. Additionally, the increased complexity introduced by node attributes and deep learning models can make the resulting models less interpretable than simpler methods.

Dimensionality reduction methods aim to reduce the dimensionality of the input data while preserving essential network properties. The input of these methods is typically the network structure represented as an adjacency matrix or a graph. Some dimensionality reduction techniques, such as matrix factorization, embedding-based methods, and graph neural networks can incorporate node attributes as additional input. The output of dimensionality reduction methods is a lower-dimensional network representation, often referred to as embeddings. These embeddings capture the essential information of the network in a reduced space. However, it is important to note that the interpretability of the embeddings may be limited. As the original node features are abstracted into lower-dimensional representations, it becomes challenging to directly interpret the meaning of individual dimensions in the embedding space. Therefore, while dimensionality reduction methods offer computational efficiency and robustness for large networks, the interpretability of the resulting embeddings should be considered in the analysis and interpretation of the link prediction results.

After selecting suitable input and output configurations for link prediction methods, it becomes important to consider the availability of computational resources, particularly when using GNNs and deep learning models. These advanced techniques often require substantial computational power and may even require specialised hardware for efficient training and inference.

Each method has advantages and limitations, and the choice of method should be carefully considered based on the network's characteristics and the application domain. Similarity-based methods may be appropriate for smaller networks with clear community structures due to their simplicity and interpretability. On the

other hand, for larger and more complex networks with available node attributes, machine learning or GNN-based approaches could be preferred to capture intricate relationships and patterns.

5 Trends and gaps

In addition to providing a comprehensive overview of existing link prediction methods, it is important to discuss the field trends and the gaps that require further investigation. This section examines the trends observed in link prediction research regarding Attributes, Type of network, and Algorithms.

5.1 In terms of attributes

One significant trend in link prediction is the increasing emphasis on integrating node and edge attributes. While traditional link prediction methods primarily rely on network topology, researchers have recognized the importance of incorporating additional information to improve prediction accuracy. Researchers aim to capture the diverse factors influencing link formations in real-world networks by considering attributes such as node features and textual content. For example, Xiao et al. [135] proposed a method incorporating structural and text features for link prediction. They leverage the textual content of nodes in a social network to capture the semantic similarity between nodes and consider the network's structural properties. By combining these features, their approach achieves more accurate link predictions compared to methods that only rely on network topology. Furthermore, Giubilei and Brutti [31] employed supervised algorithms and extracted attributes such as age and location from user profiles. Integrating attribute information in their approach significantly improved link prediction accuracy by effectively capturing the similarity between nodes based on their attribute similarities.

The availability of datasets with comprehensive attribute information is often limited in the field of link prediction. Many existing datasets lack certain attributes or have incomplete information, making it challenging to apply attribute-based methods effectively. This scarcity of datasets hampers the development and evaluation of accurate link prediction models. Moreover, even when datasets are available, they often require preprocessing to address issues such as missing values, outliers, and inconsistencies. Data preprocessing is crucial to ensure the quality and reliability of attribute data. It involves techniques like data cleaning and normalization to enhance the accuracy and effectiveness of using the attributes in link prediction algorithms.

5.2 In terms of networks

Another important trend is the recognition of the dynamic nature of networks. Real-world networks are constantly evolving, with new nodes, edges, and changes in network structure over time. Therefore, link prediction approaches need to

account for temporal dynamics and capture the changing nature of connections. Researchers [50, 102] are actively investigating dynamic link prediction techniques that can model and predict link formations over time. However, challenges remain in developing accurate and scalable methods to handle large-scale dynamic networks. Future research should focus on designing efficient algorithms and techniques to capture temporal patterns and predict links in evolving networks. Another important trend is using a knowledge network. This involves incorporating additional information or domain knowledge about the nodes, edges, or the network structure. External knowledge can provide valuable insights and enhance prediction accuracy by capturing contextual information, domain-specific relationships, or expert knowledge. Using external knowledge demonstrates the importance of considering beyond the network structure alone and leveraging relevant information to improve link prediction performance.

5.3 In terms of algorithms

In terms of algorithms, one notable trend in link prediction is the increasing use of dimensionality reduction techniques, particularly in conjunction with graph neural networks (GNNs). These techniques aim to overcome the challenge of high-dimensional feature spaces in network datasets by reducing the dimensionality of the input data while preserving relevant information. By doing so, these algorithms can improve the efficiency and effectiveness of link prediction models.

However, gaps and challenges still need to be addressed in this area. One such gap is the need for algorithms with lower complexity and reduced computational resource requirements. As network datasets continue to grow in size and complexity, it becomes increasingly important to develop algorithms that can handle the computational demands efficiently. This can involve exploring more efficient optimization strategies, model architectures, and algorithms that scale well to large networks.

6 Conclusion

In conclusion, our comprehensive literature review highlights the significance of link prediction in various domains and provides an up-to-date overview of the advancements in the field. We propose a classification framework that categorizes existing methods into machine learning, similarity-based methods, and dimensionality reduction methods, with further subdivisions within each category. We review representative algorithms within each submethod, discussing their respective advantages and disadvantages. In addition to categorizing and reviewing existing link prediction methods, we discuss the current trends and identify the gaps in the field.

Overall, this survey contributes to advancing link prediction by providing researchers with a comprehensive analysis of the latest research trends and methodologies. It guides researchers towards developing more accurate and

context-aware models and offers a rich resource of articles, datasets, and measures for further exploration. With this information, researchers can make informed decisions about applying link prediction methods in their specific domains. The survey serves as a foundation for future investigations and paves the way for advancements in the field of link prediction and its applications.

Author contributions This work is mainly carried out by the DA under the guidance and correction of the NK. The third author assisted.

Data availability Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Su Z, Zheng X, Ai J, Shen Y, Zhang X (2020) Link prediction in recommender systems based on vector similarity. Physica A 560:125154
2. Vahidi Farashah M, Etebarian A, Azmi R, Ebrahimzadeh Dastjerdi R (2021) A hybrid recommender system based-on link prediction for movie baskets analysis. J Big Data 8:1–24
3. Su Z, Zheng X, Ai J, Shang L, Shen Y (2019) Link prediction in recommender systems with confidence measures. Chaos Inter J Nonlinear Sci 29(8):083133
4. Abdolhosseini-Qomi AM, Yazdani N, Asadpour M (2020) Overlapping communities and the prediction of missing links in multiplex networks. Physica A 554:124650
5. Daud NN, Ab Hamid SH, Saadoon M, Sahran F, Anuar NB (2020) Applications of link prediction in social networks: a review. J Netw Comput Appl 166:102716
6. Berlusconi G, Calderoni F, Parolini N, Verani M, Piccardi C (2016) Link prediction in criminal networks: a tool for criminal intelligence analysis. PLoS ONE 11(4):0154244
7. Lim M, Abdullah A, Jhanjhi N, Supramaniam M (2019) Hidden link prediction in criminal networks using the deep reinforcement learning technique. Computers 8(1):8
8. Alnumay W, Ghosh U, Chatterjee P (2019) A trust-based predictive model for mobile ad hoc network in internet of things. Sensors 19(6):1467
9. De Bacco C, Power EA, Larremore DB, Moore C (2017) Community detection, link prediction, and layer interdependence in multilayer networks. Phys Rev E 95(4):042317
10. Esslimani I, Brun A, Boyer A (2011) Densifying a behavioral recommender system by social networks link prediction methods. Soc Netw Anal Min 1(3):159–172
11. Huang Z, Zeng DD (2006) A link prediction approach to anomalous email detection. In 2006 IEEE International Conference on Systems, Man and Cybernetics, vol 2, pp 1131–1136. IEEE
12. Nasiri E, Berahmand K, Rostami M, Dabiri M (2021) A novel link prediction algorithm for protein-protein interaction networks by attributed graph embedding. Comput Biol Med 137:104772
13. Cannistraci CV, Alanis-Lobato G, Ravasi T (2013) From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. Sci Rep 3(1):1–14
14. Wang P, Xu B, Wu Y, Zhou X (2015) Link prediction in social networks: the state-of-the-art. Sci China Inf Sci 58(1):1–38
15. Martínez V, Berzal F, Cubero J-C (2016) A survey of link prediction in complex networks. ACM Comput Surv 49(4):1–33
16. Kumar A, Singh SS, Singh K, Biswas B (2020) Link prediction techniques, applications, and performance: a survey. Physica A 553:124289
17. Lü L, Zhou T (2011) Link prediction in complex networks: a survey. Physica A 390(6):1150–1170
18. Wang T, He X-S, Zhou M-Y, Fu Z-Q (2017) Link prediction in evolving networks based on popularity of nodes. Sci Rep 7(1):7147

19. Zhang Z, Wen J, Sun L, Deng Q, Su S, Yao P (2017) Efficient incremental dynamic link prediction algorithms in social network. Knowl-Based Syst 132:226–235
20. Lei K, Qin M, Bai B, Zhang G, Yang M (2019) Gcn-gan: a non-linear temporal link prediction model for weighted dynamic networks. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications, pp 388–396. IEEE
21. Singh AK, Lakshmanan K (2021) Pilhnb: popularity, interests, location used hidden naive bayesian-based model for link prediction in dynamic social networks. Neurocomputing 461:562–576
22. Bütün E, Kaya M, Alhajj R (2018) Extension of neighbor-based link prediction methods for directed, weighted and temporal social networks. Inf Sci 463:152–165
23. Najari S, Salehi M, Ranjbar V, Jalili M (2019) Link prediction in multiplex networks based on interlayer similarity. Physica A 536:120978
24. Nasiri E, Berahmand K, Li Y (2021) A new link prediction in multiplex networks using topologically biased random walks. Chaos Solitons Fractals 151:111230
25. Ji S, Pan S, Cambria E, Marttinen P, Philip SY (2021) A survey on knowledge graphs: representation, acquisition, and applications. IEEE Trans Neural Netw Learn Syst 33(2):494–514
26. Rossi A, Barbosa D, Firmani D, Matinata A, Merialdo P (2021) Knowledge graph embedding for link prediction: a comparative analysis. ACM Trans Knowl Discov Data 15(2):1–49
27. Tao Y, Li Y, Wu Z (2021) Temporal link prediction via reinforcement learning. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 3470–3474. IEEE
28. Yuan W, He K, Guan D, Zhou L, Li C (2019) Graph kernel based link prediction for signed social networks. Inform Fusion 46:1–10
29. Mishra S, Singh SS, Kumar A, Biswas B (2022) Elp: link prediction in social networks based on ego network perspective. Physica A 605:128008
30. Chi K, Qu H, Yin G (2022) Link prediction for existing links in dynamic networks based on the attraction force. Chaos Solitons Fractals 159:112120
31. Giubilei R, Brutti P (2022) Supervised classification for link prediction in facebook ego networks with anonymized profile information. J Classif 5:1–24
32. Shan N, Li L, Zhang Y, Bai S, Chen X (2020) Supervised link prediction in multiplex networks. Knowl-Based Syst 203:106168
33. Karimi F, Lotfi S, Izadkhah H (2021) Community-guided link prediction in multiplex networks. J Informet 15(4):101178
34. Yang Y, Wang L, Liu D (2022) Anchor link prediction across social networks based on multiple consistency. Knowl-Based Syst 257:109939
35. Mishra S, Singh SS, Kumar A, Biswas B (2022) Mnerlp-mul: merged node and edge relevance based link prediction in multiplex networks. J Comput Sci 60:101606
36. Luo H, Li L, Zhang Y, Fang S, Chen X (2021) Link prediction in multiplex networks using a novel multiple-attribute decision-making approach. Knowl-Based Syst 219:106904
37. Yao Y, Zhang R, Yang F, Yuan Y, Sun Q, Qiu Y, Hu R (2017) Link prediction via layer relevance of multiplex networks. Int J Mod Phys C 28(08):1750101
38. Guo F, Zhou W, Wang Z, Ju C, Ji S, Lu Q (2023) A link prediction method based on topological nearest-neighbors similarity in directed networks. J Comput Sci 69:102002
39. Agibetov A (2023) Neural graph embeddings as explicit low-rank matrix factorization for link prediction. Pattern Recogn 133:108977
40. Lv L, Bardou D, Hu P, Liu Y, Yu G (2022) Graph regularized nonnegative matrix factorization for link prediction in directed temporal networks using pagerank centrality. Chaos Solitons Fractals 159:112107
41. Ghorbanzadeh H, Sheikhahmadi A, Jalili M, Sulaimany S (2021) A hybrid method of link prediction in directed graphs. Expert Syst Appl 165:113896
42. Du X, Yan J, Zhang R, Zha H (2020) Cross-network skip-gram embedding for joint network alignment and link prediction. IEEE Trans Knowl Data Eng 34(3):1080–1095
43. Chen G, Xu C, Wang J, Feng J, Feng J (2020) Nonnegative matrix factorization for link prediction in directed complex networks using pagerank and asymmetric link clustering information. Expert Syst Appl 148:113290
44. Liu S-Y, Xiao J, Xu X-K (2020) Sign prediction by motif naive bayes model in social networks. Inf Sci 541:316–331
45. Abbasi F, Muzammal M, Qureshi KN, Javed IT, Margaria T, Crespi N (2022) Exploiting optimised communities in directed weighted graphs for link prediction. Online Soc Netw Media 31:100222

46. Chen J, Zhang J, Xu X, Fu C, Zhang D, Zhang Q, Xuan Q (2019) E-lstm-d: a deep learning framework for dynamic network link prediction. IEEE Trans Syst Man Cybern Syst 51(6):3699–3712
47. Rossi E, Chamberlain B, Frasca F, Eynard D, Monti F, Bronstein M (xxxx) Temporal graph networks for deep learning on dynamic graphs
48. Yang L, Jiang X, Ji Y, Wang H, Abraham A, Liu H (2022) Gated graph convolutional network based on spatio-temporal semi-variogram for link prediction in dynamic complex network. Neurocomputing 505:289–303
49. Sankar A, Wu Y, Gou L, Zhang W, Yang H (2020) Dysat: deep neural representation learning on dynamic graphs via self-attention networks. In: Proceedings of the 13th International Conference on Web Search and Data Mining, pp 519–527
50. Huang D, Lei F (2023) Temporal group-aware graph diffusion networks for dynamic link prediction. Inform Process Manag 60(3):103292
51. Wu J, He L, Jia T, Tao L (2023) Temporal link prediction based on node dynamics. Chaos Solitons Fractals 170:113402
52. Kumar M, Mishra S, Pandey RD, Biswas B (2022) Cflp: a new cost based feature for link prediction in dynamic networks. J Comput Sci 62:101726
53. Zou L, Wang C, Zeng A, Fan Y, Di Z (2021) Link prediction in growing networks with aging. Soc Netw 65:1–7
54. Muniz CP, Goldschmidt R, Choren R (2018) Combining contextual, temporal and topological information for unsupervised link prediction in social networks. Knowl-Based Syst 156:129–137
55. Lim M, Abdullah A, Jhanjhi N, Khan MK, Supramaniam M (2019) Link prediction in time-evolving criminal network with deep reinforcement learning technique. IEEE Access 7:184797–184807
56. Mohamed HA, Pilutti D, James S, Del Bue A, Pelillo M, Vascon S (2023) Locality-aware subgraphs for inductive link prediction in knowledge graphs. Pattern Recogn Lett 167:90–97
57. Zeb A, Saif S, Chen J, Haq AU, Gong Z, Zhang D (2022) Complex graph convolutional network for link prediction in knowledge graphs. Expert Syst Appl 200:116796
58. Kumari A, Behera RK, Sahoo KS, Nayyar A, Kumar Luhach A, Prakash Sahoo S (2022) Supervised link prediction using structured-based feature extraction in social network. Concurr Comput Practice Exp 34(13):5839
59. Rafiee S, Salavati C, Abdollahpouri A (2020) Cndp: link prediction based on common neighbors degree penalization. Physica A 539:122950
60. Tang M, Wang W (2022) Cold-start link prediction integrating community information via multi-nonnegative matrix factorization. Chaos Solitons Fractals 162:112421
61. Zhou M, Han Q, Li M, Li K, Qian Z (2023) Nearest neighbor walk network embedding for link prediction in complex networks. Physica A 620:128757
62. Mavromatis C, Karypis G (2021) Graph infoclust: maximizing coarse-grain mutual information in graphs. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp 541–553. Springer
63. Wang J, Ma Y, Liu M, Shen W (2019) Link prediction based on community information and its parallelization. IEEE Access 7:62633–62645
64. Yuliansyah H, Othman ZA, Bakar AA (2023) A new link prediction method to alleviate the cold-start problem based on extending common neighbor and degree centrality. Physica A 616:128546
65. Ahmad I, Akhtar MU, Noor S, Shahnaz A (2020) Missing link prediction using common neighbor and centrality based parameterized algorithm. Sci Rep 10(1):1–9
66. Aziz F, Gul H, Muhammad I, Uddin I (2020) Link prediction using node information on local paths. Physica A 557:124980
67. Ayoub J, Lotfi D, El Marraki M, Hammouch A (2020) Accurate link prediction method based on path length between a pair of unlinked nodes and their degree. Soc Netw Anal Min 10(1):1–13
68. Jibouni A, Lotfi D, El Marraki M, Hammouch A (2018) A novel parameter free approach for link prediction. In 2018 6th International Conference on Wireless Networks and Mobile Communications (WINCOM), pp 1–6. IEEE
69. Wang G, Wang Y, Li J, Liu K (2021) A multidimensional network link prediction algorithm and its application for predicting social relationships. J Comput Sci 53:101358
70. Berahmand K, Nasiri E, Forouzandeh S, Li Y (2022) A preference random walk algorithm for link prediction through mutual influence nodes in complex networks. J King Saud Univ Comput Inf Sci 34(8):5375–5387
71. Li L, Fang S, Bai S, Xu S, Cheng J, Chen X (2019) Effective link prediction based on community relationship strength. IEEE Access 7:43233–43248

72. Singh SS, Mishra S, Kumar A, Biswas B (2020) Clp-id: community-based link prediction using information diffusion. Inf Sci 514:402–433
73. Zhang M, Chen Y (2018) Link prediction based on graph neural networks. Adv Neural Inf Process Syst 31:25
74. Zhang M, Chen Y (2017) Weisfeiler-lehman neural machine for link prediction. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 575–583
75. Daud NN, Hamid SHA, Seri C, Saadoon M, Anuar NB (2022) Scalable link prediction in twitter using self-configured framework. arXiv preprint arXiv:2208.09798
76. Keikha MM, Rahgozar M, Asadpour M (2021) Deeplink: a novel link prediction framework based on deep learning. J Inf Sci 47(5):642–657
77. Cai L, Ji S (2020) A multi-scale approach for graph link prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, vol 34, pp 3308–3315
78. Berahmand K, Nasiri E, Rostami M, Forouzandeh S (2021) A modified deepwalk method for link prediction in attributed social network. Computing 103:2227–2249
79. Fu X, Zhang J, Meng Z, King I (2020) Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In Proceedings of The Web Conference 2020, pp 2331–2341
80. Chen H, Yin H, Sun X, Chen T, Gabrys B, Musial K (2020) Multi-level graph convolutional networks for cross-platform anchor link prediction. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp 1503–1511
81. Zhang P, Chen J, Che C, Zhang L, Jin B, Zhu Y (2023) Iea-gnn: anchor-aware graph neural network fused with information entropy for node classification and link prediction. Inf Sci 634:665–676
82. Chen G, Wang H, Fang Y, Jiang L (2022) Link prediction by deep non-negative matrix factorization. Expert Syst Appl 188:115991
83. Cotta L, Bevilacqua B, Ahmed N, Ribeiro B (2023) Causal lifting and link prediction. arXiv preprint arXiv:2302.01198
84. Zulaika U, Sanchez-Corcuera R, Almeida A, Lopez-de-Ipina D (2022) Lwp-wl: link weight prediction based on cnns and the weisfeiler-lehman algorithm. Appl Soft Comput 120:108657
85. Zhao Y, Sun Y, Huang Y, Li L, Dong H (2023) Link prediction in heterogeneous networks based on metapath projection and aggregation. Expert Syst Appl 2:120325
86. Liu Y, Liu S, Yu F, Yang X (2022) Link prediction algorithm based on the initial information contribution of nodes. Inf Sci 608:1591–1616
87. Zachary WW (1977) An information flow model for conflict and fission in small groups. J Anthropol Res 33(4):452–473
88. Lusseau D, Schneider K, Boisseau OJ, Haase P, Slooten E, Dawson SM (2003) The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: can geographic isolation explain this unique trait? Behav Ecol Sociobiol 54:396–405
89. Kreft H, Jetz W (2007) Global patterns and determinants of vascular plant diversity. Proc Natl Acad Sci 104(14):5925–5930
90. Gleiser PM, Danon L (2003) Community structure in jazz. Adv Complex Syst 6(04):565–573
91. Batagelj V, Mrvar A (2006) Pajek datasets http://vlado.fmf.uni-lj.si/pub/networks/data/mix.USAir97.net
92. Anelli VW, Delić A, Sottocornola G, Smith J, Andrade N, Belli L, Bronstein M, Gupta A, Ira Ktena S, Lung-Yut-Fong A et al. (2020) Recsys 2020 challenge workshop: engagement prediction on twitter's home timeline. In Proceedings of the 14th ACM Conference on Recommender Systems, pp 623–627
93. Leskovec J, Mcauley J (2012) Learning to discover social circles in ego networks. Adv Neural Inf Process Syst 25:58
94. Anonymous: Facebook wall posts network dataset. http://konect.cc/networks/facebook-wosn-wall/ (2017)
95. Yin H, Benson AR, Leskovec J, Gleich DF (2017) Local higher-order graph clustering. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 555–564
96. Von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) Comparative assessment of large-scale data sets of protein-protein interactions. Nature 417(6887):399–403
97. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442

98. Panzarasa P, Opsahl T, Carley KM (2009) Patterns and dynamics of users' behavior and interaction: network analysis of an online community. J Am Soc Inform Sci Technol 60(5):911–932
99. Hristova D, Noulas A, Brown C, Musolesi M, Mascolo C (2016) A multilayer approach to multiplexity and link prediction in online geo-social networks. EPJ Data Sci 5:1–17
100. Vickers M, Chan S (1981) Representing classroom social structure. Victoria Institute of Secondary Education, Melbourne
101. Chai L, Tu L, Yu X, Wang X, Chen J (2023) Link prediction and its optimization based on low-rank representation of network structures. Expert Syst Appl 219:119680
102. Barracchia EP, Pio G, Bifet A, Gomes HM, Pfahringer B, Ceci M (2022) Lp-robin: link prediction in dynamic networks exploiting incremental node embedding. Inf Sci 606:702–721
103. Cai L, Li J, Wang J, Ji S (2021) Line graph neural networks for link prediction. IEEE Trans Pattern Anal Mach Intell 2:56
104. Zhang Z, Sun S, Ma G, Zhong C (2023) Line graph contrastive learning for link prediction. Pattern Recogn 140:109537
105. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology 143(1):29–36
106. Mumin D, Shi L-L, Liu L (2022) An efficient algorithm for link prediction based on local information: considering the effect of node degree. Concurr Comput Pract Exp 34(7):6289
107. Newman ME (2001) Clustering and preferential attachment in growing networks. Phys Rev E 64(2):025102
108. Salton G, Yang C-S (1973) On the specification of term values in automatic indexing. J Doc 2:58
109. Jaccard P (1901) Étude comparative de la distribution florale dans une portion des alpes et des jura. Bull Soc Vaudoise Sci Nat 37:547–579
110. Sorensen TA (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. Biol Skar 5:1–34
111. Liben-Nowell D, Kleinberg J (2003) The link prediction problem for social networks. In Proceedings of the Twelfth International Conference on Information and Knowledge Management, pp 556–559
112. Barabási A-L, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512
113. Adamic LA, Adar E (2003) Friends and neighbors on the web. Soc Netw 25(3):211–230
114. Zhou T, Kuscsik Z, Liu J-G, Medo M, Wakeling JR, Zhang Y-C (2010) Solving the apparent diversity-accuracy dilemma of recommender systems. Proc Natl Acad Sci 107(10):4511–4515
115. Katz L (1953) A new status index derived from sociometric analysis. Psychometrika 18(1):39–43
116. Tong H, Faloutsos C, Pan J-Y (2006) Fast random walk with restart and its applications. In Sixth International Conference on Data Mining (ICDM'06), pp 613–622. IEEE
117. Jeh G, Widom J (2002) Simrank: a measure of structural-context similarity. In Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 538–543
118. Leicht EA, Holme P, Newman ME (2006) Vertex similarity in networks. Phys Rev E 73(2):026120
119. CHEBOTAREV P (1997) The matrix-forest theorem and measuring relations in small social groups. Autom Remote Control 58(9):1505–1514
120. Liu W, Lü L (2010) Link prediction based on local random walk. Europhys Lett 89(5):58007
121. Zhou T, Lü L, Zhang Y-C (2009) Predicting missing links via local information. Eur Phys J B 71:623–630
122. Salton G (1983) Introduction to modern information retrieval. McGraw-Hill, London
123. Pons P, Latapy M (2005) Computing communities in large networks using random walks. In Computer and Information Sciences-ISCIS 2005: 20th International Symposium, Istanbul, Turkey, October 26–28, 2005. Proceedings 20, pp 284–293. Springer
124. Kowsari K, Jafari Meimandi K, Heidarysafa M, Mendu S, Barnes L, Brown D (2019) Text classification algorithms: a survey. Information 10(4):150
125. Sen PC, Hajra M, Ghosh M (2020) Supervised classification algorithms in machine learning: A survey and review. In Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018, pp 99–111. Springer

126. Wu L, Cui P, Pei J, Zhao L, Guo X (2022) Graph neural networks: foundation, frontiers and applications. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp 4840–4841
127. Kipf TN, Welling M (xxxx) Semi-supervised classification with graph convolutional networks
128. Zaheer M, Kottur S, Ravanbakhsh S, Poczos B, Salakhutdinov RR, Smola AJ (2017) Deep sets. Adv Neural Inf Process Syst 30:58
129. Li Y, Zemel R, Brockschmidt M, Tarlow D (2016) Gated graph sequence neural networks. In Proceedings of ICLR'16
130. He C, Cheng J, Fei X, Weng Y, Zheng Y, Tang Y (2023) Community preserving adaptive graph convolutional networks for link prediction in attributed networks. Knowl-Based Syst 5:110589
131. Mi Q, Wang X, Lin Y (2023) A double attention graph network for link prediction on temporal graph. Appl Soft Comput 136:110059
132. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533
133. Lim M, Abdullah A, Jhanjhi N (2020) Data fusion-link prediction for evolutionary network with deep reinforcement learning. Int J Adv Comput Sci Appl 11(6):245
134. Chen L, Cui J, Tang X, Qian Y, Li Y, Zhang Y (2022) Rlpath: a knowledge graph link prediction method using reinforcement learning based attentive relation path searching and representation learning. Appl Intell 52(4):4715–4726
135. Xiao Y, Li R, Lu X, Liu Y (2021) Link prediction based on feature representation and fusion. Inf Sci 548:1–17
136. Le T, Tran H, Le B (2023) Knowledge graph embedding with the special orthogonal group in quaternion space for link prediction. Knowl-Based Syst 266:110400
137. Chen G, Xu C, Wang J, Feng J, Feng J (2020) Robust non-negative matrix factorization for link prediction in complex networks using manifold regularization and sparse learning. Physica A 539:122882
138. Moutinho JP, Melo A, Coutinho B, Kovács IA, Omar Y (2023) Quantum link prediction in complex networks. Phys Rev A 107(3):032605
139. Kumar M, Mishra S, Biswas B (2022) Pqklp: projected quantum kernel based link prediction in dynamic networks. Comput Commun 196:249–267
140. Singh SS, Srivastva D, Kumar A, Srivastava V (2022) Flp-id: Fuzzy-based link prediction in multiplex social networks using information diffusion perspective. Knowl-Based Syst 248:108821
141. Zheng J, Qin Z, Wang S, Li D (2022) Attention-based explainable friend link prediction with heterogeneous context information. Inf Sci 597:211–229
142. Xu R-Q, Zhou M-Y, Liao H (2022) Pnr: How to optimally combine different link prediction approaches? Inf Sci 584:342–359