Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-18287-9

1236: EXPLAINABLE ARTIFICIAL INTELLIGENCE SOLUTIONS FOR IN-THE-WILD HUMAN BEHAVIOR ANALYSIS

Exploring the scope of explainable artificial intelligence in link prediction problem-an experimental study

Mridula Dwivedi[1] · Babita Pandey[1] · Vipin Saxena[1]

Received: 7 August 2023 / Revised: 30 November 2023 / Accepted: 15 January 2024

Abstract

The realm of SN has witnessed remarkable developments, capturing the attention of researchers who seek to process and analyze user data in order to extract meaningful insights for future predictions and recommendations. Among the challenging problems in SN analysis is LP, which leverages available data and network knowledge, including node characteristics and connecting edges, to forecast potential associations in the near future. LP is used in data mining, commercial and e-commerce recommendation systems, and expert systems. This research presents a thorough LP taxonomy, including Similarity Metrics and Learning-based approaches, and their recent expansion in numerous network environments. This article also discusses XAI, a method that helps people understand and trust ML systems. LP taxonomy based on XAI is also proposed. The research also examines LIME, a popular XAI approach that illuminates ML and DL models. LIME provides model-independent local explanations for regression and classification tasks on structured and unstructured data. The study includes an extensive experimental evaluation of incorporating XAI with LP, which shows the XAI approach's ability to solve LP problems and interpret predictions. This research uses XAI to give users practical insights and a better knowledge of the LP problem.

Keywords Link prediction · Explainable artificial intelligence · Social networks · LIME · Machine learning · Similarity metrics

✉ Babita Pandey
shukla_babita@yahoo.co.in

Mridula Dwivedi
md.cs.bbau@gmail.com

Vipin Saxena
profvipinsaxena@gmail.com

[1] Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, UP, India

1 Introduction

The majority of human-related activities in the world can be presented in the form of network and graphs, where the links signify the association between the various entities. The information networks that we currently see are pervasive in real-world networks like the Worldwide Web, protein-protein interaction network, airlines-transport network, author-citation network, real-world SN, and so on.

LP in SN is particularly a challenging problem because of its highly dynamic nature. SNs tend to quickly expand and transform over time due to inclusion and exclusion of nodes and/or edges. In order to forecast lost edges in an existing network and brand-new or fading edges in upcoming networks, expert systems adopt LP approaches would surely generate data and retrieve information. LP algorithms can help identify fake or fraudulent links. However, it is important to note that some links that may appear unexpected or surprising could be mistakenly classified as false links. Removing these links without caution might generate a distorted knowledge of the system's architecture and behavior. An important question in the original environment may usually be mapped back to the network's LP in general, and vice versa.

In the beginning, the researchers have studied the network as connectedness (interaction) between node pairs, node pair connectivity as triangle closure and similarity of interaction between node pairs. Later on, the same was treated as closeness of network nodes (CN) and gave rise to "Link Prediction problem".

A fresh wave of applications for AI has been generated by recent developments in ML, which offer considerable benefits to a number of sectors. Recent successes in AI are mostly the result of current ML advancements that build models using the representations they have within themselves. They consist of SVM, DL, RL, RFs, and PGMs. Some models are challenging to understand despite having good performance. There can often be an accord between ML models performance, such as their expected accuracy, and their level of explainability. It is common for the most effective models, like decision trees, to be less explainable, while the most accurate ones, such as DL [1], may offer higher accuracy but lower interpretability.

The intention of an XAI system is to boost the understandability of its behavior by providing explanations. In order to develop potent and more interpretable AI systems, it is recommended that XAI systems be capable of describing its knowledge, skills, ongoing actions, future plans, and the most relevant information it considers. It is important to note that every explanation, whether comprehensive or incomplete, is contextual and relies on factors such as the task at hand, user expertise, and the expectations of an AI-based system. Therefore, interpretability and explainability are dependent on the specific domain and cannot be universally determined independently of it.

Table 1 contains the list of abbreviations and symbols used in the article along with their description.

1.1 Motivation and research gaps

LP and XAI both have separately been topics of interest among researchers. However, the usage of XAI in LP was not much observed. Figure 1 shows the number of Google searches on LP and XAI separately since past five years taken from Google Trends. Many emerging LP techniques fail to provide explainability of their results. This key motivation

Table 1 Table of abbreviations/symbols used in the article

Symbol/Abbreviation	Description
AA	Adamic Adar
AI	Artificial Intelligence
ANN	Artificial Neural Network
ARI	Adjusted Rand Index
AUC	Area Under Curve
AUPRC	Area Under Precision-Recall Curve
AUROC	Area Under Receiver Operating Characteristic Curve
CD	Community Detection
CN	Common Neighbor
CN-AH	Common Neighbors Authority Hub
CN-HA	Common Neighbors Hub Authority
CNN	Convolutional Neural Network
COND	Conductance Of Detected Communities
CORLP	Complex Number Representation Link Prediction
CTT	Contextual-Temporal-Topological
DL	Deep Learning
DR	Dimensionality Reduction
ERGM	Exponential Random Graph Model
GNN	Graphical Neural Network
Hits@k	Hit rate
JC	Jaccard's Coefficient
KG	Knowledge Graph
KI	Katz Index
KNN	K-Nearest Neighbor
LAS	Local Affinity Structure
LIME	Local Interpretable Model-agnostic Explanations
LP	Link Prediction
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
Symbol/Abbreviation	Description
MAP	Mean Average Precision
MCC	Matthew's Correlation Coefficient
MF	Matrix Factorization
MGW	Music Genre Weigh
ML	Machine Learning
MLRW	Multiplex Local Random Walk
MRR	Mean Reciprocal Rank
NLPM	Neighbor-based Link Prediction Measures
NMI	Normalized Mutual Information
PA	Preferential Attachment
PCC	Pearson's Correlation Coefficient
PGM	Probabilistic Graphical Models
PR	Precision-Recall
RA	Resource Allocation

Table 1 (continued)

Symbol/Abbreviation	Description
RF	Random Forest
RL	Reinforcement Learning
RMSE	Root Mean Squared Error
RGNMF-AN	Robust Graph Regularization Nonnegative Matrix Factorization for Attributed Networks
ROC	Receiver Operating Characteristic
RW	Random Walk
RWR	Random Walk with Restart
SBM	Stochastic Block Model
SCNHA	Sum of Common Neighbors with Hub and Authority
SN	Social Network
SULP	Shabaz-Urvashi Link Prediction
SVM	Support Vector Machine
TPR	True Positive Rate
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boost

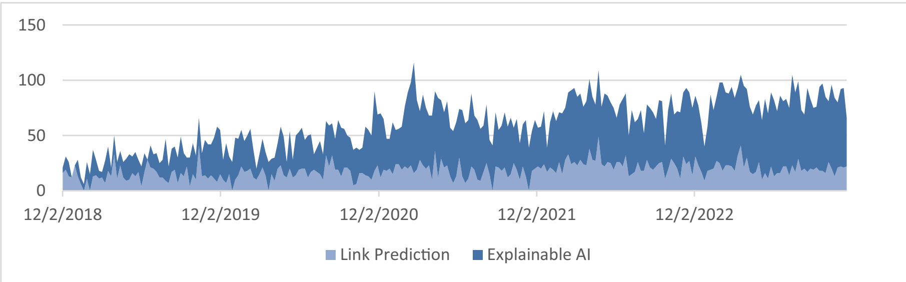


Fig. 1 Graph showing the worldwide Google searches on Link Prediction and Explainable AI in past 5 years taken from Google Trends

for conducting this research is to establish a novel approach to LP by combining XAI for clear and understandable decision-making in complicated networks.

Studies carried out in [2–4] only discussed LP, similarity measures, ML approaches and challenges whereas [5, 6] discussed taxonomy, summary and research directions in XAI. No existing literatures implemented XAI in LP. This article discusses this major issue faced in existing literatures and provides a method to implement XAI with similarity metrics.

1.2 Contributions

Our previous works [7–10] lack some key properties in LP that we have contributed in this article. The following key contributions make this work more thorough and in-dept than previous studies:

- A comprehensive exploration of phases of LP is conducted which also provides a basic idea on various evaluation metrics and their usage in LP.
- A generic taxonomy stating Similarity Based and Learning Based LP techniques is provided along with their limitations and utilities based year-wise.
- The evolution of LP methods proportional to the network types from 2013 to 2023 are picturized.
- The inclusion of XAI and LP is a novel aspect of this survey. A taxonomy of XAI tools for LP is presented together with a case study of its use.
- The challenges that could arise during the adoption of XAI tools and methodologies for LP are discussed.

The charm of this survey is that it makes it simple for the readers to gain insight into the considerations made for LP and XAI.

1.3 Research methodology

We adopted a basic methodology to conduct the survey as represented in Fig. 2. The steps comprise of selecting prime Scopus database libraries like Wiley, Elsevier, Springer, and Blackwell and then searching the research articles related to LP. The literatures were searched from years 2013 to 2023 using keywords: “Link Prediction”, “Similarity Metrics”, “Machine Learning”, and “Explainable Artificial Intelligence”. After obtaining the search results, we tried to filter the results. The filtration and pre-processing of the literatures was purely title restricted. The literatures consisting of “Link Prediction” in their title was then selected manually. Figure 3 represents a graphical overview of number of papers published on LP Strategies stated above in ScienceDirect database.

After the preprocessing was performed, we studied the literatures and summarized them by providing an exhaustive literature review and their gaps. Further, we discussed the phases of LP and proposed a taxonomy of LP comprising of Similarity based and Learning

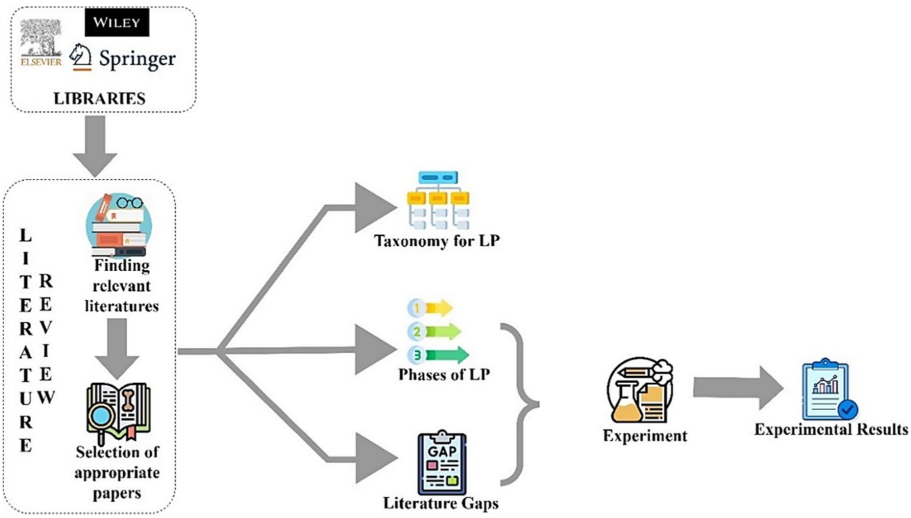


Fig. 2 Pictorial representation of the research methodology conducted for the proposed study

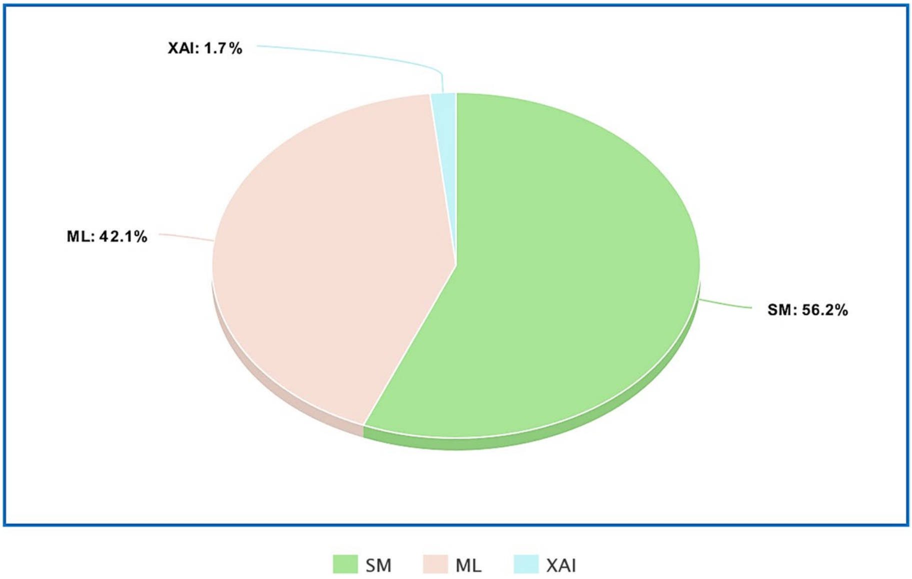


Fig. 3 Percentage of research articles available on ScienceDirect on Link Prediction with Similarity Metrics, Machine Learning and Explainable Artificial Intelligence

based methods. At last, we conducted an experimental exercise on the proposed method and obtained its result.

This paper is organized in following sections: Section 2 provides a Literature Summary; Section 3 provides an overview of the phases to solve a LP problem; Section 4 defines Experiments and Results; Section 5 is about Discussion with Limitations and Open Challenges discussed in Section 6 and Section 7 respectively; followed by Conclusion and Future Work in the last i.e., Section 8.

2 Literature review

After studying various literatures on LP and XAI, an exhaustive literature summary is generated (as shown in Table 2). This summary provides the literature summary of various research works performed from a timespan of 2013 to 2023 providing a general idea of LP methods and XAI tools used. XAI makes use of KG as a tool, which further needs much exposure.

3 Phases of link prediction

The major steps performed in LP problem are: Data collection, network representation (optional), LP method application, performance evaluation and/or model explanation are shown in Fig. 4.

Table 2 Exhaustive Tabular Survey of the related literatures studied

Reference	PM/ML/EV/XAI	Advantages/Drawbacks/Result
Sun et al. [11]	**PM:** LAS **EV:** AUC	Clustering coefficient is directly proportional to clustering degree.
Tang & Wang [12]	**PM:** Multi-nonnegative MF model **EV:** AUC, Precision	Outperformed existing methods.
Liu [13]	**PM:** e-comm recommendation algo based on LP **EV:** MAD, Recommendation Coverage, F1 Score, Precision	Accurate for small dataset.
Zhao et al. [14]	**PM:** efficient sketch based algorithm **EV:** Accuracy	Efficient and cost-effective.
Xu & Yin [15]	**PM:** CRA Index (Improvement of RA) **EV:** AUC	Better than other similarity based methods.
Goswami et al. [16]	**PM:** Movie recommendation algo based on user and movie profiling **ML Model:** RF and SVM **EV:** MAE, Precision, Recall, F1 Score	Better than CF Pearson, CF Cosine and CF Multilevel approaches for both RF and SVM model.
Li & Cai [17]	**PM:** Heirchical Cluster Ensemble Model based on Knowledge Granulation **EV:** AUC	Better than Similarity based methods.
Ahmed & ElKorany [18]	**PM:** Modified FriendTNS **EV:** Precision, Recall, ARHR	Better than existing methods.
Lv et al. [19]	**PM:** Local Path Index **EV:** AUC	Required less CPU time & memory space Poor accuracy than others.
Cheng et al. [20]	**PM:** CD using LP **EV:** NMI	Outperformed baselines.
Jiang et al. [21]	**PM:** CLPE **EV:** NMI, ARI, Avg COND	Better than existing methods. Time consuming.
Zhao et al. [22]	**PM:** MGW based on CORLP **EV:** Precision, Recall, F-Score	Better than CORLP method. Tested on small dataset.
Cui et al. [23]	**PM:** Application of Linkage-Weight in Heterogeneous network **EV:** Precision, Recall, F-Value	Better precision. A trial method.

Table 2 (continued)

Reference	PM/ML/EV/XAI	Advantages/Drawbacks/Result
Berahmand et al. [24]	**PM:** Mutual Influence RW: Modified version of Local RW **EV:** AUC, Precision	Better than other similarity measures.
Papadimitriou et al. [25]	**PM:** Friendlink Algorithm **EV:** Precision, Recall.	Better than KI and RWR. Higher complexity
Papadimitriou et al. [26]	**PM:** EXTENSION OF: Friendlink **EV:** Precision, Recall, MAP, AUC.	Better than various measures. Time consuming.
Shabaz & Garg [27]	**PM:** SULP **EV:** AUROC, Precision, Recall, TPR	Ideal than other LP measures. Unscalable.
Shabaz & Garg [28]	**PM:** SULP **EV:** AUROC, Precision	Poor AUROC and precision.
Yao et al. [29]	**PM:** 3 metric based CN with inclusion of 2 Hop Paths. **EV:** ROC, AUC	Best performance; contains time varied weight info of links.
Li et al. [30]	**PM:** LP Recommendation Algo with Domain Knowledge and Topological Properties. **EV:** Precision, Recall, F1 Score.	Better performance.
Malhotra & Goyal [31]	**PM:** Framework for supervised ML-based future link detection in single-layer and multiplex networks. **ML Model:** For Classification: SVM, KNN, Decision Tree, ANN, Bagging Classifier, ADA Boost, RF Classifier. For Train-test Split: 5 Fold Cross Validation. **EV:** Accuracy, Precision, Recall, F Score, AUC	Poor performance
Thi et al. [32]	**PM:** Transfer AdaBoost with SVM **ML Model:** Naive Bayes Classifier (to initialize latent features matrices) **EV:** Accuracy, Precision, Recall, Performance Time (Speed)	Better Accuracy and PR Curve
Stanhope et al. [33]	**PM:** Group LP **ML Model:** Long-Short Term Based Memory Model **EV:** Hits@ 5, 10, 20	Efficient than other baseline methods.

Table 2 (continued)

Reference	PM/ML/EV/XAI	Advantages/Drawbacks/Result
Nassar et al. [34]	**PM:** Pairseeded PageRank, Triangle Reinforced PageRank **EV:** AUROC	Multiple-seeding strategies were better than others.
Zhang et al. [35]	**PM:** Contextualized Self-Supervised Learning framework **EV:** AUC	Better AUC and performance.
Xu et al. [36]	**PM:** Subgraph Neighboring Relations Infomax **EV:** AUPRC with Hits@10	Better than 4 state-of-the-art methods.
Naravani et al. [37]	**PM:** NA **ML Model:** support vector regression, multiple linear regression, and Gaussian regression. **EV:** RMSEs	Multiple LR was better than non-prediction model.
Agibetov [38]	**PM:** J and NetMF **EV:** AUROC	Better than original NetMF.
Zulaika et al. [39]	**PM:** Link Weight Prediction Weisfeiler–Lehman method **EV:** MSE	Average results.
Weinzierl & Harabagiu [40]	**PM:** Automatic detection of known Misinformation **EV:** Micro Precision, Recall, F1 Score	Better performance.
Nasiri et al. [41]	**PM:** MLRW **EV:** AUC, Precision	Better than several LP methods.
Yasami & Safaei [42]	**PM:** Multilayer model of dynamic complex networks **EV:** Sensitivity, Specificity, LR+, LR−, PV+,PV−, F1-score, MCC, Accuracy	Outperformed traditional single-layer approaches.
Aghabozorgi & Khayyambashi [43]	**PM:** Tridiac Similarity **ML Model:** Linear Discriminant Analysis classifier, Gradient Boosting Machine classifier **EV:** Accuracy, AUC	Outperformed CN, JC, AA, PA.
Muniz et al. [44]	**PM:** CTT **EV:** Improvement Factor	Weighted AA had better results.
Chamberlain et al. [45]	**PM:** Efficient LP with Hashing, BUDDY **ML Model:** GNN	BUDDY is more scalable and outperformed various LP methods. Limited only to undirected graph

Table 2 (continued)

Reference	PM/ML/EV/XAI	Advantages/Drawbacks/Result
Bastami et al. [46]	**PM:** (Concurrent) Gravitation based LP **EV:** Accuracy, ROC	Better accuracy than other methods. Less time consuming.
Jiang et al. [47]	**PM:** GreedyAdd and GreedyAdd with HeuristicAdd **EV:** AUC	Better than RandomAdd and AdjacentAdd
Ghorbanzadeh et al. [48]	**PM:** CN-HA, CN-AH, SCNHA **ML Model:** Logistic Regression, Gradient Boosting, Linear Discriminant, RF, Decision Tree **EV:** mean & std. deviation of AUC, t-test, ROC	Better than other Neighbor-based methods.
Li et al. [49]	**PM:** Meta-Path feature-based Back Propagation neural n/w model **EV:** AUC, Accuracy, Recall	Better than other similarity measures. Unscalable, high cost and time
Wang et al. [50]	**PM:** Multidimensional network model**EV:** AUC	Better accuracy than other methods. Noisy data obtained from processing, limited data usage.
Shakibian et al. [51]	**PM:** Multilayered model based Link Predictor **ML Model:** Least Square Twin SVM **EV:** prediction Accuracy, AUC	Highest prediction accuracy and AUC. Computationally Complex.
Zhao et al. [52]	**PM:** ICP **EV:** Precision, AUC	Better performance High computational complexity
Nasiri et al. [53]	**PM:** Fusing Structure and Feature in Deepwalk **EV:** AUC, F-measure, Avg PR, RMSE, PCC	Better results than other methods.
Bütün & Kaya [54]	**PM:** Improvement of Tridiac closeness **ML Model:** C4.5, KNN, multilayer perceptron, RF, Random Tree **EV:** ROC, AUC	Better than other similarity measures. Node features not utilized.
Zhou et al. [55]	**PM:** Graph Embedding Biased RWR **EV:** AUC	Best results with node2vec. Info embeddings not utilised.
Kou et al. [56]	**PM:** Sinhash based LP **EV:** Precision, Recall, F1-Score, G-measure, Specificity, Accuracy	Better than existing methods.
Lee & Zhou [57]	**PM:** self-included Collaborative Filtering**EV:** AUC	Better than various similarity measures.
Karimi et al. [58]	**PM:** Community-guided LP based on External Similarity **EV:** AUC	Better than other measures.

Table 2 (continued)

Reference	PM/ML/EV/XAI	Advantages/Drawbacks/Result
Gao et al. [59]	**PM:** Context of Social Influence Propagation **ML Model:** Graph Neural N/w **EV:** MAP, AUC, Hit Score, F1-Score	Best performance. Less accurate.
Han et al. [60]	**PM:** Multilayer knowledge n/w method Knowledge transfer model based on trust transitivity **EV:** AUC	Static n/w used.
Bai et al. [61]	**PM:** TOPSIS **EV:** AUC, Precision	Better Precision than other multiplex LP methods.
Wu et al. [62]	**PM:** Serial ensemble strategy via n/w reconstruction **EV:** Precision, AUC, AUPRC	Better performance.
Chen et al. [63]	**PM:** Low Rank Representaion with Non-negative MF **EV:** Precision, AUC	Better AUC and Precision.
Li et al. [64]	**PM:** Ensemble model-based LP **ML Model:** LR1, XGBoost **EV:** AUC, Recall	Better Accuracy.
Mallek et al. [65]	**PM:** Evidential LP **EV:** AUC, Precision	Better than various similarity measures. Only Simple n/w was used.
Wang et al. [66]	**PM:** Fusion Probabilistic MF **EV:** AUC, Accuracy	Better than other LP methods. Used only static n/w
Zhang et al. [67]	**PM:** EdgeConvNorm **ML Model:** GNN **EV:** AUC	Better accuracy. Small-scale static data only used.
Liu et al. [68]	**PM:** Extended RA index **EV:** AUC, Precision	Better than other LP methods.
Bütün et al. [69]	**PM:** Extended NLPMs with weight **ML Model:** KNN, RF, random tree, multilayer perceptron and C4.5 classifiers **EV:** AUC, ROC	Better than Traditional NLPM. Only 2 hop paths were considered.

Table 2 (continued)

Reference	PM/ML/EV/XAI	Advantages/Drawbacks/Result
Kumar et al. [70]	**PM:** Local Global Quasi model **ML Model:** Logistic Regression, Neural n/w, XGBoost **EV:** AUPR, AUC, F1 Score, Balanced Accuracy Score	Better performance than other measures.
Chen et al. [71]	**PM:** Graph Regularization Weighted Nonnegative MF model **EV:** AUC, Precision, RMSE, PCC	Better performance. Static unweighted and undirected n/w used.
Nasiri et al. [72]	**PM:** WCN **EV:** ROC, Precision	Better than existing methods.
Nasiri et al. [73]	**PM:** RGNMF-AN **ML Model:** **EV:** AUC, F-measure, RMSE, PCC	Enhanced prediction performance.
Niranjan et al. [74]	**XAI:** Guided Gradcam based Explainable Classification and Segmentation system	Much accurate decision support system.
Le et al. [75]	**PM:** RotatPRH **EV:** Mean Rank, MRR, Hits@k **XAI:** via KG	Better performance.
Shi et al. [76]	**PM:** KG embedding model based on a relational memory network and CNN. **EV:** MRR, Hits@k **XAI:** via KG	Improved efficiency.
Safavi et al. [77]	**PM:** confidence calibration for KG embeddings **EV:** Expected Calibration Error, Accuracy **XAI:** via KG	Accurate and efficient.
Ma et al. [78]	**PM:** Entity LP for KG **EV:** Hits, MRR **XAI:** via KG	OutperROC,forms baseline methods.
Xiao et al. [79]	**PM:** ManifoldE **EV:** Hits@k **XAI:** via KG	Better performance.

Table 2 (continued)

Reference	PM/ML/EV/XAI	Advantages/Drawbacks/Result
Ranganathan et al. [80]	**PM:** HOPLoP **ML Model:** PGM **EV:** MAP, MRR, Hits@k **XAI:** via KG	Decreased error rates.
Rossi et al. [81]	**PM:** Kelpie**ML Model:** TransE, ComplEx, ConV**EV:** Hits@k, MRR **XAI:** via KG	Outperforms baseline methods.
Stoica et al. [82]	**PM:** CoPER **ML Model:** ConvE, MINERVA **EV:** Hits@k, MRR **XAI:** via KG	Flexible and fast than other approaches.

PM Proposed Method, ***ML*** Machine Learning model, ***EV*** Evaluation Metrics, ***XAI*** Explainable Artificial Intelligence technique

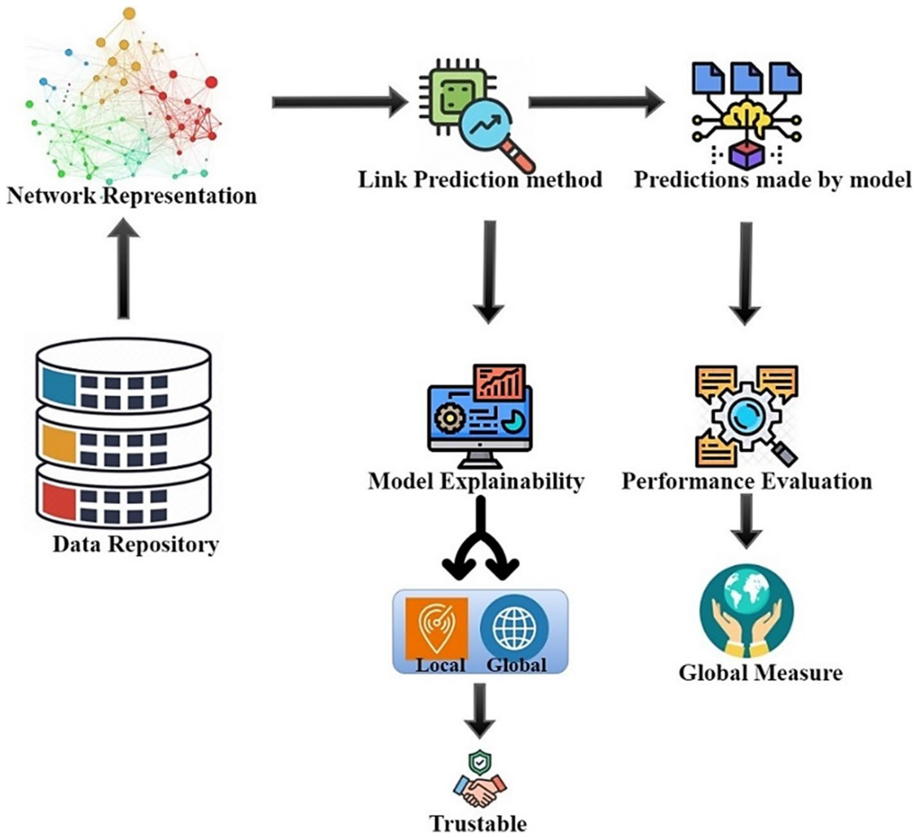


Fig. 4 Phases of a Link Prediction Problem starting from Data Collection followed by Network Representation then application of LP method and evaluation using ML model or its explainability

3.1 Data collection

Data Collection is primarily performed in two ways: 1) downloading existing datasets from data repositories and libraries like SNAP, Kaggle, Github and others; and 2) dataset construction.

Data collection, data cleaning, and data labelling are the three essential steps in the dataset construction process. Data gathering involves finding datasets for ML model training. There are two main approaches: when there is small dataset for training, Data Generation is used whereas Data Augmentation is another approach to obtain data. The procedure involves adding recently acquired external data to existing datasets. Data production involves: crowdsourcing, a business model that connects huge groups of people online to complete activities; and synthetic data, manufactured by a machine, to increase our training data or add future data updates.

3.2 Network representation (NR)

NR encompasses various techniques for representing networks, each with its unique approach. To graph the network adjacency matrix is often used which utilizes similarity measures. Embedding-based methods represent network node properties or linkages and converts nodes, linkages, and their characteristics into a vector space while preserving graph structure. PGMs offer representation of graph probability distributions to show complex probability connections where nodes represent random variables and edges represent probabilistic linkages between variables. Finally, GNNs (also known as KG) are effective at understanding massive, dynamic graph datasets with billions of elements, especially complex network architectures.

3.3 Link prediction methods

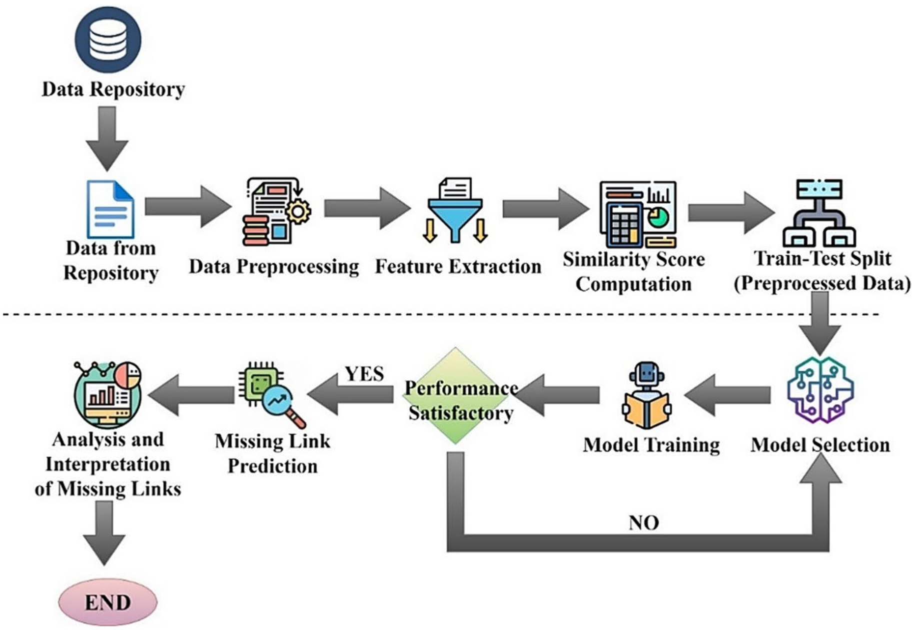
LP methods can be broken down into two primary categories: those that are based on similarities i.e., Similarity Metrics, and those that are based on learning i.e., Learning Based Methods. The first type computes the likelihood of a link existing between two node pairs based on the assigned similarity score. Either the nodes' attributes or the network's topology can play a role in calculating the similarity score. There are three distinct types of learning-based approaches. Figure 5(a) depicts the workflow of LP problem adopting similarity metrics whereas Fig. 5(b) adopts learning-based techniques.

3.3.1 Similarity metrics

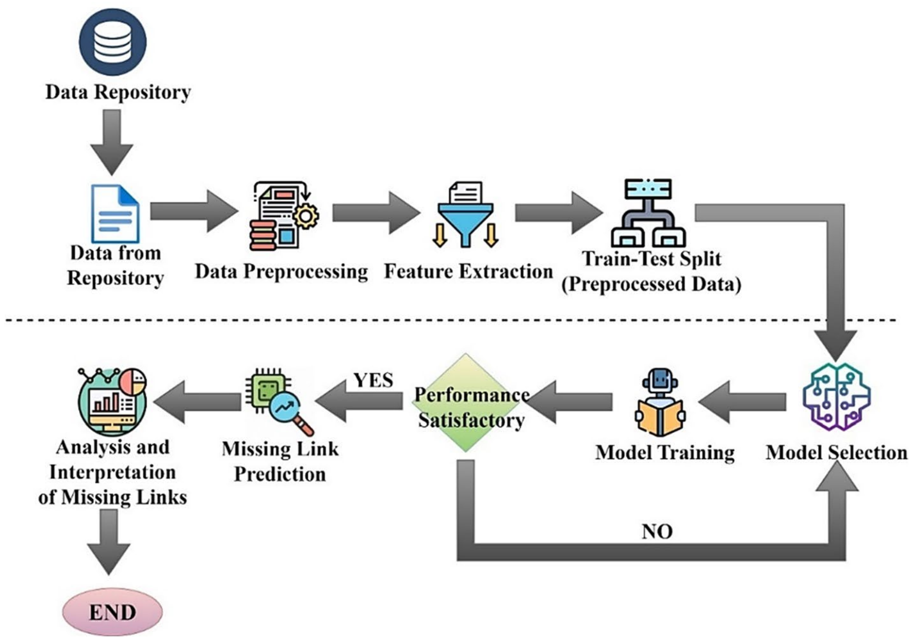
Similarity-based algorithms first determine the probable strength of a connection between two nodes based on their resemblance, then select the "L" linkages with the most similarities. Network topology calculates the similarity score of two non-connected nodes. Local, global, and quasi-local scores can compare nodes. Local-based scores detect similar node pairings using local information. Global node similarity methods consider network architecture. Global information and computational complexity benefit them. Quasi-local similarities balance these techniques. They need more data than local indices but less time than global ones. Researchers use several similarity metrics to tackle LP problems, including neighbors, dataset similarity/dissimilarity, node closeness, and degree of connectivity. Studies carried out in [3, 4] have discussed various Similarity Metrics adopted widely.

3.3.2 Learning based methods

Learning-based techniques incorporate network architectural and non-architectural information into ML frameworks. This lets the techniques determine the likelihood of an edge between two nearby nodes. LP uses supervised learning methods including SVM, RF, KNN, Naive Bayes, Ensemble Learning, Logistic Regression, Radial Basis Function network, and others. Representation learning strategies can be classified as MF-based, Deep Neural Network-based, or Path and Walk-based based on the models' loss function and decoder function (graph similarity metrics) [83].



(a) A generic flowchart depicting process of solving Link Prediction problem using Similarity Metrics.



(b) A generic flowchart depicting process of solving Link Prediction problem using Learning Based approach.

Fig. 5 (**a**): A generic flowchart depicting process of solving Link Prediction problem using Similarity Metrics. (**b**): A generic flowchart depicting process of solving Link Prediction problem using Learning Based approach

3.4 Performance evaluation

Performance of LP methods are commonly evaluated using popular metrics like Accuracy, Precision, F1 Score, Recall, Receiver Operating Curve (ROC), AUC, HR@k, and MRR. Various authors have used some uncommon metrics for performance evaluation which are stated in Table 3.

Table 3 is generated based on various literatures studied in Table 2. The data in Table 3 provides a clear idea of the uses of evaluation metrics in terms of network types and/or system type which might help users in the future to gain knowledge about which metric to use in their work, depending upon network/system type. CD is one among the most well-known problems in LP which uses MAE, NMI, ARI and Average COND whereas for a multilayer complex network, TPR, Sensitivity, Specificity and MCC are used. For DR methods such as MF and embedding, RMSE and PCC are used.

3.5 Explainable artificial intelligence

AI that can be explained by a human expert is referred to as XAI. It contrasts with the idea of the "black box" in ML, where even the inventors of the AI cannot explain why the AI made a particular decision. The social right to explanation is implemented by XAI.

Generally, XAI is classified in two categories: 1) Global: It provides a general explanation of the concept and is based on universal operational principles, 2) Local: It provides an explanation of the rules that produce each individual piece of data.

Figure 6 represents XAI techniques applied Locally, Globally, both Globally and Locally, along with the Explainable tools improve LP result interpretation and user comprehension. Explainability strategies in ML are varied. Permutation Importance compares a model trained on the original data to randomly rearranged feature values to determine feature importance. Partial Dependence Plots help discover key features and understand their interactions by showing the relationship between a target variable and input features. Accumulated Local Effects is used for non-linear models with complicated input-output interactions, while Morris Sensitivity Analysis evaluates input feature superiority. Global Interpretation via Recursive Partitioning explains complex model behavior with decision trees. Anchors explain model workings, while Contrastive Explanation Method compares

Table 3 Some uncommon Evaluation Metrics with their uses in network/usage type and reference

Reference	Evaluation Metric	Use
[16, 30]	MAE	Recommendation systems.
[20, 21]	NMI	CD.
[21]	ARI	CD.
[21]	Average COND	CD.
[27]	TPR	Multilayer complex network.
[37, 41]	RMSE	DR methods (MF, embedding).
[42]	Sensitivity	Multilayer complex network.
[42, 47]	Specificity	Multilayer complex network.
[42]	MCC	Multilayer complex network.
[41, 71]	PCC	DR methods (MF, embedding).
[56]	G-Measure	Signed network.
[59]	MAP	Measures stability of model.

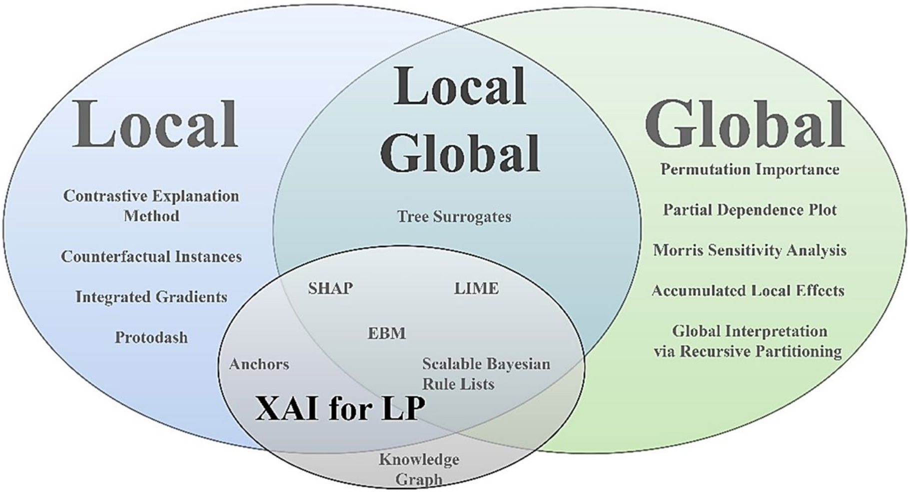


Fig. 6 Classification of XAI technologies based on their Local, Global, both Local and Global applicability and commonly used XAI techniques with Link Prediction to interpret the prediction results of LP model

predictions to similar examples to find minimal input changes that affect predictions. Counterfactual Instances verify model stability and accuracy. Model behavior is explained by Integrated Gradients and LIME. Shapley values determine feature influence, Scalable Bayesian Rule Lists provide interpretable if-then rules, and Explainable Boosting Machine makes accurate, feature-selective predictions using Boosting and Generalised Additive Models.

4 Experiments and results

This section explains the experiment conducted for obtaining an Explainable LP. We have concluded our results based on accuracy and ROC curve. The experiment was conducted on a workstation with Intel Core i7 4770, 2.2 GHz GPU, 16 GB memory and Windows 10 pro operating system. The experiment was implemented in Python along with libraries: pandas, numpy, networkx, scikit learn, and seaborn.

4.1 Dataset and evaluation metrics

The dataset chosen for conducting this research is Facebook-Social-Network-Analysis dataset, which is used to predict future friend recommendation and consists of three entities: Node 1, Node 2 and Connection which represents the "from node", "to node" and Connection type respectively. Connection is a Boolean value: 1 for connected node and 0 for unconnected node. Table 4 provides the general statistics of the dataset. The dataset was taken from https://github.com/abcom-mltutorials/Facebook-Social-Network-Analysis.

In order to evaluate the performance of the method, we opted Accuracy and ROC curve. Accuracy is measured using:

Table 4 General statistics of data

Nodes	3979
Edges	50,591

$$Accuracy = \frac{Correct\ Predictions}{Total\ Predictions} \quad (1)$$

5 Method

The proposed approach incorporates various similarity measures as parameters, utilizes the RF classifier as the ML model, and employs LIME as XAI technique. Figure 7 represents a complete systematic architecture of the proposed approach.

Firstly, the Facebook data was taken from the Github repository, then the data is preprocessed and similarity scores are computed post which we created a dataframe consisting of similarity scores and nodes which was then utilized for computing correlation and splitting into train-test data. After deciding the features and target, RF classifier was trained using train set and then predictions were made using test set. Lastly the performance was evaluated and results were interpreted using LIME.

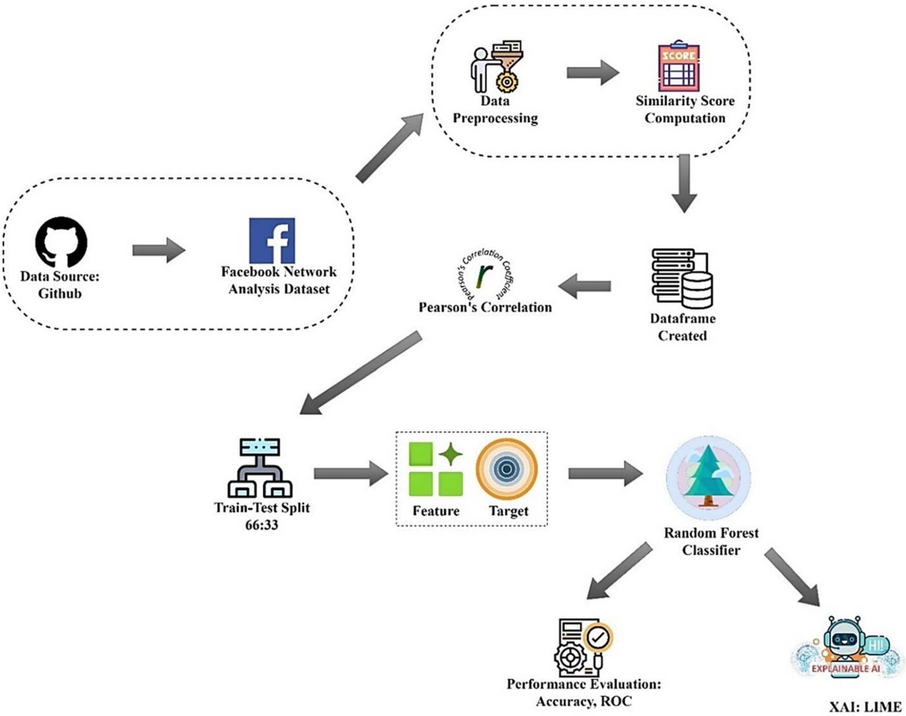


Fig. 7 A systematic architecture of the proposed methodology

Table 5 Computed Similarity Score of first 10 edges (for reference)

from	To	Connection	CN	JC	RA	AA	PA
2	3	0	2	0.016667	0.027106	0.464909	3717
2	10	0	4	0.035398	0.042454	0.876477	3422
2	24	0	4	0.035088	0.054284	0.930221	3481
2	116	1	5	0.04	0.069456	1.167168	4189
2	226	1	2	0.015504	0.029449	0.473451	4248
2	254	0	8	0.068966	0.093967	1.791249	3835
2	265	0	6	0.043478	0.071057	1.345058	5015
2	345	0	4	0.030303	0.050391	0.907075	4543
2	368	0	4	0.028369	0.048267	0.898561	5074
2	446	0	3	0.022388	0.040808	0.697223	4602

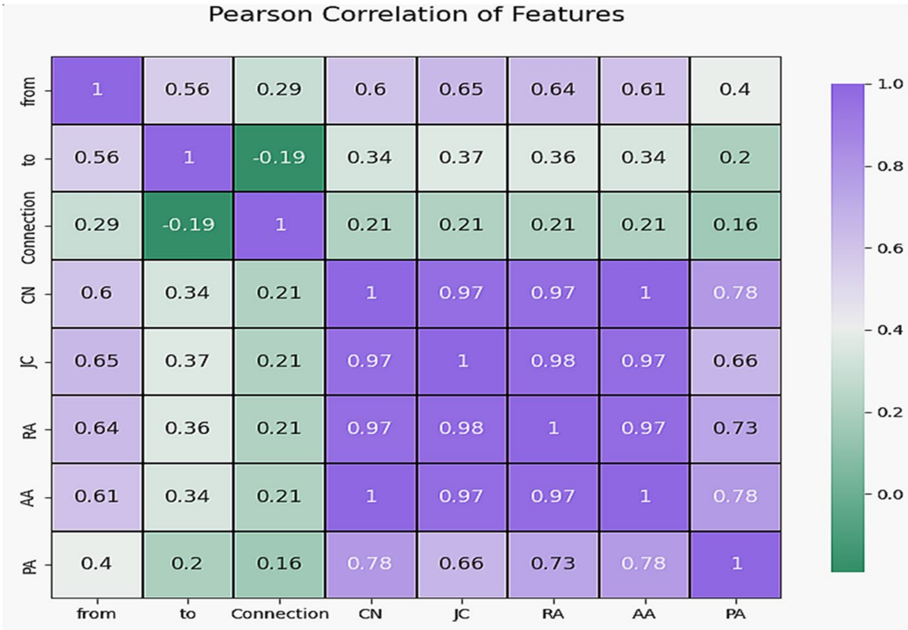


Fig. 8 Pearson Correlation computed of Similarity scores, nodes and Connection (data in Table 2)

6 Results

6.1 Preprocessing and parameters

In order to preprocess the collected data, sorting of the columns was performed to get two nodes as a tuple “edge”. Similarity Metrics CN, AA, RA, JC, and PA were calculated. The first 10 rows of the scores of these are mentioned in Table 5. The data generated in Table 5 is used to Train and Test the classifier (Fig. 8).

The data in Table 5 has a strong linear association between two continuous variables as computed by:

$$\text{Pearson Correlation } (r) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{2}$$

New data consists of 5 feature variables (as shown in Table 5): CN, JC, RA, AA, PA and one target variable 'connectivity': exist (1) or not exist (0).

Data splitting impacts the performance and generalization of the model. We implemented a random splitting technique by randomly shuffling the preprocessed dataframe and subsequently dividing it into a training set and a test set in the ratio of 67:33. Table 6 provides a sample from dataset that is used for training and testing. RF Classifier model was build using ensemble learning.

The computed similarity metrics are treated as primary parameters (features) for feeding input to LIME, no parameters were set empirically. The choice of parameters is completely based on the choice of ML model, XAI tool and type of results to interpret. For conducting this research, similarity metrics are chosen as parameters as they will help in the interpretation of results by supporting the existence or non-existence of links based on the values they generate for a specific node.

7 Experimental analysis

We tested our method via accuracy and ROC curve. Accuracy computed was 0.6678 whereas Fig. 9 shows the plotted ROC curve. Predictions were interpreted using LIME as shown in Fig. 10(a) for data at index 6985. LIME predicts with 91% confidence that the connection does not exist (actual connection not exist as shown in Table 7(row 1, column 2)). Parameters RA and CN increase the probability of not existence. Similarly, Fig. 10(b) and (c) represent LIME results for index values 9864 and 10,256 respectively which predict the inexistence of connection (actual connection also does not exist, shown in Table 7(row 2 and 3, column 2)).

Table 6 Sample from created dataframe for training and testing purpose

Feature					Target
CN	JC	RA	AA	PA	Connection
2	0.016667	0.027106	0.464909	3717	0
4	0.035398	0.042454	0.876477	3422	0
4	0.035088	0.054284	0.930221	3481	0
5	0.04	0.069456	1.167168	4189	1
2	0.015504	0.029449	0.473451	4248	1
8	0.068966	0.093967	1.791249	3835	0
6	0.043478	0.071057	1.345058	5015	0
4	0.030303	0.050391	0.907075	4543	0
4	0.028369	0.048267	0.898561	5074	0
3	0.022388	0.040808	0.697223	4602	0

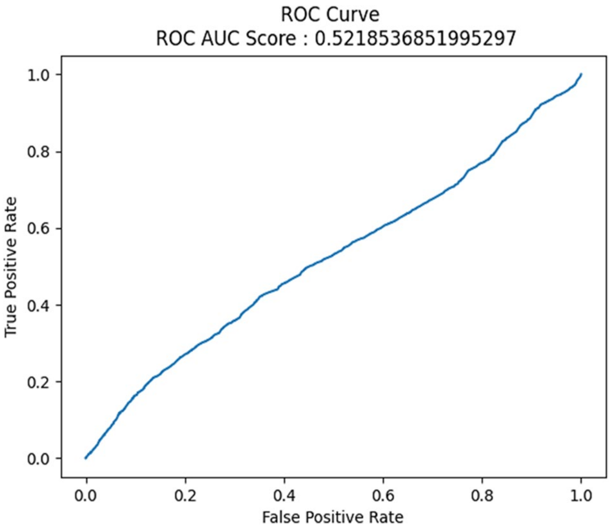


Fig. 9 Plotted ROC curve with ROC AUC Score for the obtained predictions

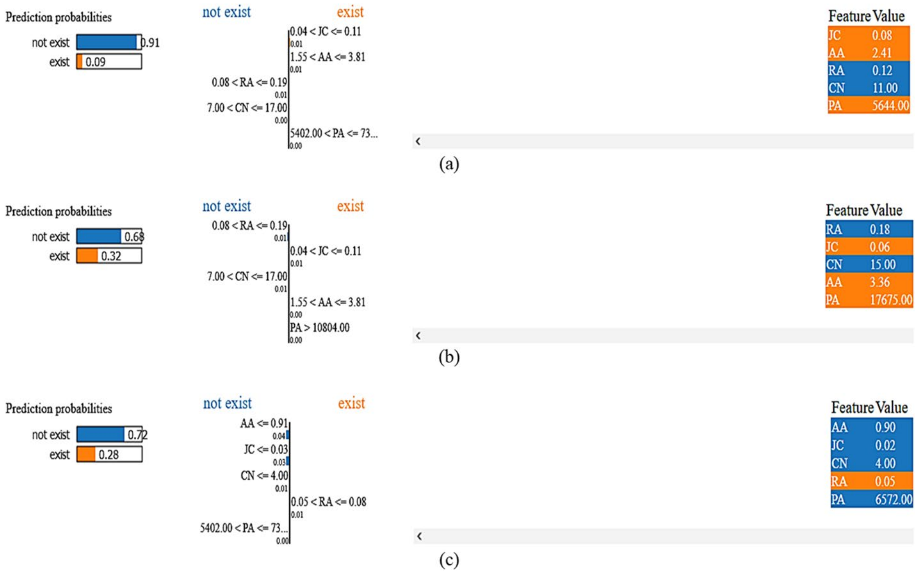


Fig. 10 LIME results for (a) value =6985, (b) value = 9864, and (c) value = 10,256

Table 7 Reference for Connection value for LIME Interpreter

from	to	Connection
79	1911	0
936	2583	0
1707	1975	0

8 Discussion

After studying various literatures on LP, it is observed that LP methods are labelled as: Similarity Based and Learning Based methods (shown in Fig. 11). Further learning based methods are classified into: Algorithmic, DR and Probabilistic method.

The algorithmic approach to LP involves employing classification techniques and meta-heuristics. This entails extracting features from network data and utilizing them as inputs for training ML models. By discerning patterns and relationships within the network data, these models strive to predict links between nodes. Concurrently, DR serves as a method to transform larger datasets into more manageable forms, preserving crucial information. Applied to address classification and regression challenges, it aids in obtaining more accurate predictive models for LP. Methods combining DR with LP include MF and embedding-based techniques. Additionally, probabilistic LP utilizes statistical models like ERGM, SBM, or latent space models to estimate the likelihood of node connections. Maximum likelihood-based link prediction assesses the statistical model's parameters for their chance to enhance observed data, encompassing network structures and other attributes.

Figure 12 shows the evolution of Similarity based (left branch) and ML (right branch) based LP approaches used from year 2013 to 2023 with their limitations and utilities. This figure helps novice to select and integrate Similarity based and ML based approach on the basis of their complementary features for increasing the effectiveness of LP methods.

Networks belong to various categories depending upon their structure (multi-layer, multigraph, simple, complex, bipartite), nature (heterogeneity, homogeneity), attributes (node

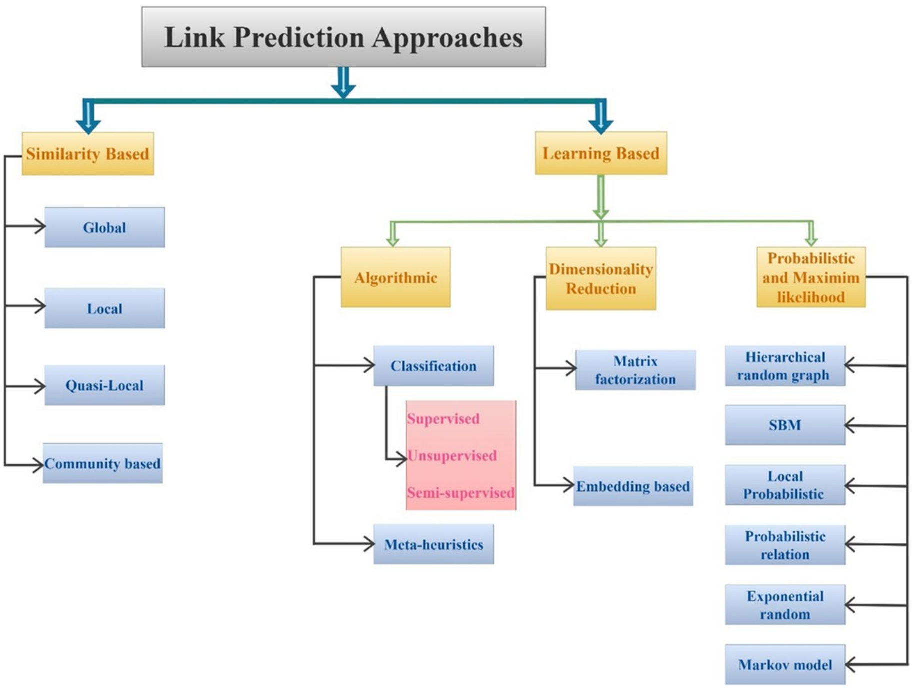


Fig. 11 A generic taxonomy of Link Prediction techniques which involves Similarity Based and Learning Based methods

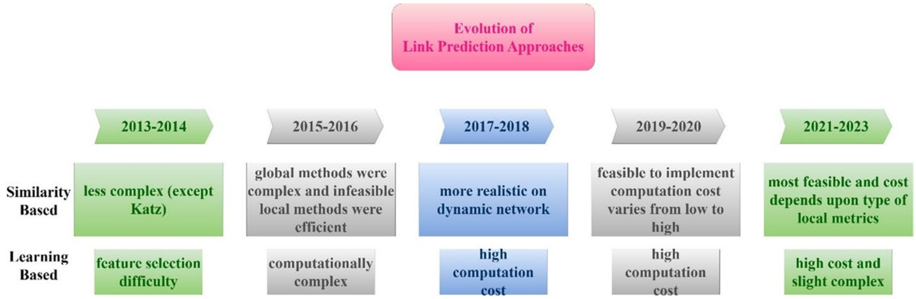


Fig. 12 Major Limitations and utilities of Link Prediction observed from year 2013 to 2023

attributes), and direction (unidirectional, bidirectional, no direction). Due to difference in structure, nature, attributes, direction of graphs, one type of LP method is not applicable to all graphs. Figure 13 shows year-wise the category of graph and LP methods applied on them in chronological order. With the help of this figure, it can be identified which type of graph and the LP method is deployed from the year 2013 to 2023.

The application of XAI to LP is our main innovation. XAI reduces the cost of mistakes, finds their causes, and improves model efficiency by characterizing errors and decreasing biased predictions. The specific requirements for implementing XAI in Python can vary depending on the techniques and libraries. The minimum requirements include Python ML libraries like scikit-learn, TensorFlow, or PyTorch; ML models; Interpretability Libraries like SHAP, LIME, or InterpretML; preprocessed data; and the right XAI approach. Final steps include Documentation and Visualization.

With the complexity of AI technology, algorithms are hard to grasp and analyze. Researchers can create and improve methodologies. The need to minimize the model in many XAI algorithms makes performance prediction challenging. For more complex models, current explainability methods may not account for all factors that influence a choice,

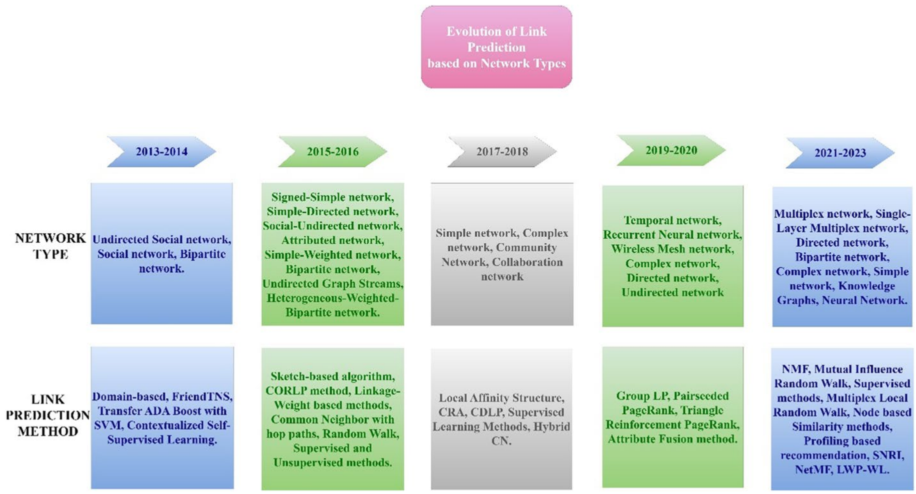


Fig. 13 Evolution of LP methodology with reference to network type from year 2013 to 2023

limiting their usefulness. Creating ethically sound and well explained XAI algorithms is the goal of new research.

The proposed approach preserves some properties that are its effectiveness, robustness, interpretability, and user-friendliness. It excels in effectiveness by making predictions on large dataset, the method is simple and fosters a seamless interaction between the method and its users. Its innovation lies in the incorporation of XAI in LP domain.

Our approach incorporates a diverse array of similarity measures, ensuring its adaptability to various graph types and guaranteeing its robustness across diverse datasets.

9 Limitations

While studying available literatures, some limitations were found which are specified as:

- Searching of literatures with keywords provided some irrelevant literatures which required manual filtering and excessive time consumption.
- Various articles did not explicitly mention the ML models used by them. Wherever the models were not specified, we did not mention the model's name in our literature survey (Table 2).
- Many researchers did not specify the drawbacks of their methods clearly. Mostly the drawbacks were drawn on the basis of results only.
- Wherever the factors like node attributes, weights, network properties might be used, they were not utilized. We did not mention them explicitly in our article.

10 Open challenges for research

10.1 Challenges in LP

Scalability, complexity and computational expenses are a few problems faced by LP that have been quoted by other authors. However, some problems continue to go unreported:

- **Dynamicity:** Different types of network dynamism exist, including nodes and edges being added and deleted at next timestamp. LP only handles one or two types of dynamicity; no existing technique covers all dynamicity**.**
- **Generalization of network:** Each network has unique nodes and linkages; thus, they should be structured accordingly. Currently, there is no comprehensive and universal LP solution available for networks.
- **Timestamp missing:** The dataset lacks timestamps for network-wide link or node formation for time period 't'. In such a network, separating training and testing datasets is difficult. Because some linked node pairs may be randomly assigned to the training set and others to the testing set. In this scenario, CN-based methods are unreliable.
- **Imbalance in dataset:** The SN dataset includes mostly bad and some outstanding class. Unsupervised learning algorithms are indifferent to class distributions; therefore, they cannot balance data and focus on class boundaries. This problem can be solved with ensemble methods and data sampling.

10.2 Challenges in XAI

- **Blackbox resemblance:** Experts have trouble understanding many ML algorithms' decisions. Black box solutions for incomprehensible judgements may cause legal, ethical, and operational difficulties. Before implementation, black-box machines cannot be checked or audited, making behaviour assurances problematic. Why or how to rectify a ML system's bad judgement is difficult.
- **Biasness:** Keeping an AI programme from learning biases or unfair perspectives is difficult. Possible gaps in the training data, model, and objective function cause this challenge. For ethical and fair AI systems, these biases must be addressed and mitigated.
- **Fairness of results:** XAI struggles to assess AI system fairness. This difficulty occurs because fairness perceptions vary depending on context and ML algorithm input.
- **Safety issues:** AI reliability is assessed by examining its decision-making process. The fundamental generalisation in statistical learning theory requires organisations to make assumptions from unseen data to fill gaps, making this task difficult.

11 Conclusion and future work

This paper offers an exhaustive literature review on LP problem and XAI, accompanied by a thorough analysis and understanding of LP, its distinct phases, and the problem-solving techniques employed. The prime objective of this study is to establish a generalized concept of XAI and explore its applicability in LP. Among the myriad XAI tools and methods available, the experimental exercise focuses on LIME as it sheds light on the interpretation of link existence or absence between pairs of nodes. The experimental exercise conducted on Facebook, a real-world SN, demonstrates the potential for significant accuracy improvements using various similarity measures and interpretation of results using LIME.

As of our future perspective we will figure out more emerging techniques based on ML, DL, and ANN based LP methodologies. We plan to extend our study by incorporating various datasets to broaden the scope of our analysis. Additionally, we aim to enhance our method's effectiveness by considering node attributes and conducting comparisons with existing methods for a more comprehensive evaluation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11042-024-18287-9.

Funding This research received no external funding.

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

Declarations

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

References

1. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang GZ (2019) XAI-Explainable artificial intelligence. Sci Robot 4(37):7120
2. Martínez V, Berzal F, Cubero JC (2016) A survey of link prediction in complex networks. ACM Comput Surv 49(4):1–33
3. Vital A, Amancio DR (2022) A comparative analysis of local similarity metrics and machine learning approaches: application to link prediction in author citation networks. Scientometrics 127(10):6011–6028
4. Zhou T (2021) Progresses and challenges in link prediction. iScience 24(11):103217
5. Borys K, Schmitt YA, Nauta M, Seifert C, Krämer N, Friedrich CM et al (2023) Explainable AI in medical imaging: an overview for clinical practitioners – saliency-based XAI approaches. Eur J Radiol 162:110787
6. Saeed W, Omlin C (2023) Explainable AI (XAI): a systematic meta-survey of current challenges and future opportunities. Knowl Based Syst 263:110273
7. Pandey B, Bhanodia PK, Khamparia A, Pandey DK (2019) A comprehensive survey of edge prediction in social networks: techniques, parameters and challenges. Expert Syst Appl 124:164–181
8. Bhanodia PK, Khamparia A, Pandey B (2021) Supervised shift k-means based machine learning approach for link prediction using inherent structural properties of large online social network. Comput Intell 37(2):660–677
9. Bhanodia PK, Khamparia A, Pandey B (2021) An efficient link prediction model using supervised machine learning. In: Recent studies on computational intelligence: doctoral symposium on computational intelligence (DoSCI 2020). Springer, Singapore, pp 19–27
10. Bhanodia PK, Khamparia A, Pandey B (2021) An approach to predict potential edges in online social networks. In: Data science and security: proceedings of IDSCS 2020. Springer, Singapore, pp 1–6
11. Sun Q, Hu R, Yang Z, Yao Y, Yang F (2017) An improved link prediction algorithm based on degrees and similarities of nodes. In: Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017 13–8
12. Tang M, Wang W (2022) Cold-start link prediction integrating community information via multi-nonnegative matrix factorization. Chaos Solitons Fractals 162(112421):1–13
13. Liu G (2022) An ecommerce recommendation algorithm based on link prediction. Alex Eng J 61(1):905–910
14. Zhao P, Aggarwal C, He G (2016) Link prediction in graph streams. 2016 IEEE 32nd International Conference on Data Engineering, ICDE 2016. 2016 Jun 22 553–64
15. Xu M, Yin Y (2017) A similarity index algorithm for link prediction. In: Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2017
16. Goswami S, Roy S, Banerjee S et al (2022) A profiling-based movie recommendation approach using link prediction. Innovations Syst Softw Eng. https://doi.org/10.1007/s11334-022-00472-4
17. Li S, Cai N (2018) Construction of brand community overlap based on ensemble link prediction algorithm. In: Proceedings - 2018 3rd international conference on mechanical, control and computer engineering, ICMCCE 2018. Institute of Electrical and Electronics Engineers Inc., pp 438–441
18. Ahmed C, ElKorany A (2015) Enhancing link prediction in Twitter using semantic user attributes. In: Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 1155–61
19. Lv L, Jin CH, Zhou T (2009) Effective and efficient similarity index for link prediction of complex networks. Phys Rev E Stat Nonlinear Soft Matter Phys 80(4). arXiv preprint arXiv:0905.3558
20. Cheng HM, Ning, YZ, Yin Z, Yan C, Liu X, Zhang ZY (2018) Community detection in complex networks using link prediction. Mod Phys Lett B 32(01):1850004
21. Jiang H, Liu Z, Liu C, Su Y, Zhang X (2020) Community detection in complex networks with an ambiguous structure using central node based link prediction. Knowl-Based Syst 195:105626
22. Zhao D, Zhang L, Zhao W (2016) Genre-based link prediction in bipartite graph for music recommendation. Procedia Computer Science 91:959–965
23. Cui Y, Zhang L, Wang Q, Chen P, Xie C (2016) Heterogeneous network linkage-weight based link prediction in bipartite graph for personalized recommendation. Procedia Computer Science 91:953–958
24. 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
25. Papadimitriou A, Symeonidis P, Manolopoulos Y (2011) Friendlink: Link prediction in social networks via bounded local path traversal. In: Proceedings of the 2011 International Conference on Computational Aspects of Social Networks, CASoN'11 66–71

26. Papadimitriou A, Symeonidis P, Manolopoulos Y (2012) Fast and accurate link prediction in social networking systems. J Syst Softw 85(9):2119–2132
27. Shabaz M, Garg U (2021) Shabaz–Urvashi link prediction (SULP): a novel approach to predict future friends in a social network. J Creative Commun 16(1):27–44
28. Shabaz M, Garg U (2022) Predicting future diseases based on existing health status using link prediction. World J Eng 19(1):29–32
29. Yao L, Wang L, Pan L, Yao K (2016) Link prediction based on common-neighbors for dynamic social network. Procedia Computer Science 83:82–89
30. Li J, Zhang L, Meng F, Li F (2014) Recommendation algorithm based on link prediction and domain knowledge in retail transactions. Procedia Computer Science 31:875–881
31. Malhotra D, Goyal R (2021) Supervised-learning link prediction in single layer and multiplex networks. Machine Learning with Applications 6:100086
32. Nguyen-Thi AT, Nguyen PQ, Ngo TD, Nguyen-Hoang TA (2015) Transfer AdaBoost SVM for link prediction in newly signed social networks using explicit and PNR features. Procedia Computer Science 60:332–341
33. Stanhope A, Sha H, Barman D, Hasan M Al, Mohler G (2019) Group Link Prediction. In: Proceedings - 2019 IEEE International Conference on Big Data 3045–52
34. Nassar H, Benson AR, Gleich DF (2019) Pairwise link prediction. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM 2019. Association for Computing Machinery, Inc, pp 386–393
35. Zhang D, Yin J, Yu PS (2022) Link prediction with contextualized self-supervision. IEEE Trans Knowl Data Eng 35(7):7138–7151
36. Xu X, Zhang P, He Y, Chao C, Yan C (2022) Subgraph Neighboring Relations Infomax for Inductive Link Prediction on Knowledge Graphs. IJCAI International Joint Conference on Artificial Intelligence 2341–7
37. Naravani M, Narayan DG, Shinde S, Mulla MM (2020) A cross-layer routing metric with link prediction in wireless mesh networks. Procedia Computer Science 171:2215–2224
38. Agibetov A (2023) Neural graph embeddings as explicit low-rank matrix factorization for link prediction. Pattern Recogn 133
39. Zulaika U, Sánchez-Corcuera R, Almeida A, López-de-Ipiña D (2022) LWP-WL: link weight prediction based on CNNs and the Weisfeiler–Lehman algorithm. Appl Soft Comput 120
40. Weinzierl MA, Harabagiu SM (2021) Automatic detection of COVID-19 vaccine misinformation with graph link prediction. J Biomed Inform 124
41. Nasiri E, Berahmand K, Li Y (2021) A new link prediction in multiplex networks using topologically biased random walks. Chaos Solitons Fractals 151
42. Yasami Y, Safaei F (2018) A novel multilayer model for missing link prediction and future link forecasting in dynamic complex networks. Physica A 492:2166–2197
43. Aghabozorgi F, Khayyambashi MR (2018) A new similarity measure for link prediction based on local structures in social networks. Physica A 501:12–23
44. 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
45. Chamberlain BP, Shirobokov S, Rossi E, Frasca F, Markovich T, Hammerla N, et al (2022) Graph Neural Networks for Link Prediction with Subgraph Sketching arXiv preprint arXiv:2209.15486.
46. Bastami E, Mahabadi A, Taghizadeh E (2019) A gravitation-based link prediction approach in social networks. Swarm Evol Comput 44:176–186
47. Jiang Z, Tang X, Zeng Y, Li J, Ma J (2021) Adversarial link deception against the link prediction in complex networks. Physica A 577
48. Ghorbanzadeh H, Sheikhahmadi A, Jalili M, Sulaimany S (2021) A hybrid method of link prediction in directed graphs. Expert Syst Appl 165
49. Chao LJ, Ling ZD, Ge BF, Yang KW, Chen YW (2018) A link prediction method for heterogeneous networks based on BP neural network. Physica A 495:1–17
50. 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
51. Shakibian H, Charkari NM, Jalili S (2016) A multilayered approach for link prediction in heterogeneous complex networks. J Comput Sci 17:73–82
52. Zhao Z, Gou Z, Du Y, Ma J, Li T, Zhang R (2022) A novel link prediction algorithm based on inductive matrix completion. Expert Syst Appl 188
53. 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

54. Bütün E, Kaya M (2019) A pattern based supervised link prediction in directed complex networks. Physica A 525:1136–1145
55. Zhou Y, Wu C, Tan L (2021) Biased random walk with restart for link prediction with graph embedding method. Physica A 570
56. Kou H, Liu H, Duan Y, Gong W, Xu Y, Xu X et al (2021) Building trust/distrust relationships on signed social service network through privacy-aware link prediction process. Appl Soft Comput 100
57. Lee YL, Zhou T (2021) Collaborative filtering approach to link prediction. Physica A 578
58. Karimi F, Lotfi S, Izadkhah H (2021) Community-guided link prediction in multiplex networks. J Inf Secur 15(4)
59. Gao H, Li B, Xie W, Zhang Y, Guan D, Chen W et al (2021) CSIP: enhanced link prediction with context of social influence propagation. Big Data Res 24
60. Han J, Teng X, Tang X, Cai X, Liang H (2020) Discovering knowledge combinations in multidimensional collaboration network: a method based on trust link prediction and knowledge similarity. Knowl Based Syst 195:105701
61. Bai S, Zhang Y, Li L, Shan N, Chen X (2021) Effective link prediction in multiplex networks: a TOPSIS method. Expert Syst Appl 177
62. Wu M, Wu S, Zhang Q, Xue C, Kan H, Shao F (2019) Enhancing link prediction via network reconstruction. Physica A 534
63. Chen X, Wu T, Xian X, Wang C, Yuan Y, Ming G (2020) Enhancing robustness of link prediction for noisy complex networks. Physica A 555
64. Li K, Tu L, Chai L (2020) Ensemble-model-based link prediction of complex networks. Comput Netw 166:106978
65. Mallek S, Boukhris I, Elouedi Z, Lefèvre E (2019) Evidential link prediction in social networks based on structural and social information. J Comput Sci 30:98–107
66. Wang Z, Liang J, Li R (2018) Exploiting user-to-user topic inclusion degree for link prediction in social-information networks. Expert Syst Appl 108:143–158
67. Zhang Z, Cui L, Wu J (2021) Exploring an edge convolution and normalization based approach for link prediction in complex networks. J Netw Comput Appl 189:103113
68. Liu S, Ji X, Liu C, Bai Y (2017) Extended resource allocation index for link prediction of complex network. Physica A 479:174–183
69. Bütün E, Kaya M, Alhajj R (2018) Extension of neighbor-based link prediction methods for directed, weighted and temporal social networks. Inf Sci 463–464:152–165
70. Kumar M, Mishra S, Biswas B (2022) Features fusion based link prediction in dynamic networks. J Comput Sci 57:101493
71. Chen G, Xu C, Wang J, Feng J, Feng J (2019) Graph regularization weighted nonnegative matrix factorization for link prediction in weighted complex network. Neurocomputing 369:50–60
72. Nasiri E, Berahmand K, Samei Z, Li Y (2022) Impact of centrality measures on the common neighbors in link prediction for multiplex networks. Big Data 10(2):138–150
73. Nasiri E, Berahmand K, Li Y (2023) Robust graph regularization nonnegative matrix factorization for link prediction in attributed networks. Multimed Tools Appl 82(3):3745–3768
74. Niranjan K, Shankar Kumar S, Vedanth S, Chitrakala S (2023) An explainable AI driven decision support system for COVID-19 diagnosis using fused classification and segmentation. Procedia Comput Sci 218:1915–1925
75. Le T, Le N, Le B (2023) Knowledge graph embedding by relational rotation and complex convolution for link prediction. Expert Syst Appl 214:119122
76. Shi M, Zhao J, Wu D (2023) Convolutional neural network knowledge graph link prediction model based on relational memory. Comput Intell Neurosci 1–9
77. Safavi T, Koutra D, Meij Bloomberg E. Evaluating the Calibration of Knowledge Graph Embeddings for Trustworthy Link Prediction In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)
78. Ma J, Qiao Y, Hu G, Wang Y, Zhang C, Huang Y et al (2019) ELPKG: a high-accuracy link prediction approach for knowledge graph completion. Symmetry 11(9):1096
79. Xiao H, Huang M, Zhu X (2016) From One Point to A Manifold: Knowledge Graph Embedding For Precise Link Prediction. IJCAI International Joint Conference on Artificial Intelligence 1315–21
80. Ranganathan V, Barbosa D, Lin X, Qin L, Zhang W, Zhang Y et al HOPLoP: multi-hop link prediction over knowledge graph embeddings. World Wide Web 25(9)
81. Rossi A, Firmani D, Merialdo P, Teofili T (2022) Explaining Link Prediction Systems based on Knowledge Graph Embeddings. In: Proceedings of the ACM SIGMOD International Conference on Management of Data 2062–75

82. Stoica G, Stretcu O, Platanios EA, Mitchell TM, Póczos B (2020) Contextual parameter generation for knowledge graph link prediction. Proc AAAI Conf Artif Intell 34(03):3000–3008
83. Mutlu EC, Oghaz T, Rajabi A, Garibay I (2020) Review on learning and extracting graph features for link prediction. Mach Learn Knowl Extr 2(4):672–704