Journal of King Saud University - Computer and Information Sciences 36 (2024) 101931

HOSTED BY

Contents lists available at ScienceDirect

Journal of King Saud University - Computer and Information Sciences

journal homepage: www.sciencedirect.com

Link prediction using extended neighborhood based local random walk in multilayer social networks

Xueping Ren

Information Engineering College, Hangzhou Dianzi University, Hangzhou, Zhejiang 311305, China

ARTICLE INFO

Keywords:
Multilayer social networks
Link prediction
Local random walk
Vertex influence
Extended neighborhood

ABSTRACT

One of these challenges in the analysis of social networks is the problem of link prediction. The purpose of this problem is to find links that have not yet been observed, but may exist in the future. There are many solutions for link prediction on monoplex networks. However, many real social networks model communication in multiple layers, which are known as multilayer social networks. A solution for multilayer networks involves taking into account the information of all layers to make predictions for a target layer. Among the existing solutions, local random walk has been confirmed as an efficient technique for link prediction in monoplex networks, but this technique is inefficient for link prediction in multilayer networks due to computational complexity. In order to address this issue, in this paper we propose Extended Neighborhood based Local Random Walk (ENLRW) for link prediction in multilayer networks. ENLRW is an extended version of the classical local random walk technique in which the nearest neighbors are considered based on the extended neighborhood concept. ENLRW calculates the similarity between vertices by integrating several different metrics through reliable paths that include intra-layer and inter-layer information. Besides, ENLRW considers vertex influence as a similarity metric to provide an effective reliable biased random walk. The results of the simulations show that the use of different inter-layer and intra-layer information as well as the local random walk configuration with extended neighborhood provides a trade-off between precision and complexity. Specifically, ENLRW improves the average precision by 3.1% compared to the best available state-of-the-art method.

1. Introduction

The main entities of a social network are the vertices and the links between them (Liu et al., 2024). Each vertex is assigned to a user in the social network, where the characteristics of the vertices distinguish them from each other. Also, the links between vertices can be represented by one or more types of relationship such as friendship, disease, and cooperation. A link between two vertices represents a connection between two vertices in a network. Therefore, links can communicate between network entities. Social networks contain valuable information for analyzing social structures (Liu et al., 2011). The analysis of these networks is considered as an important research topic in recent decades. Social network analysis is an interdisciplinary topic that has attracted researchers from different disciplines. Some of the most important challenges related to social network analysis are: connection detection, network structure analysis, network visualization and link prediction (Mohamed et al., 2023).

Social networks as dynamic platforms are always increasing members and forming new connections (Gao and Rezaeipanah, 2023). Insights into these networks show that some of these connections may be lost for various reasons. Among these reasons, privacy protection, hidden errors in the search engine crawler, data storage or data transmission issues can be mentioned (Zhang et al., 2022; Cao et al., 2022). In addition to these issues, many links may appear in the future. Therefore, it is important to anticipate missing connections or possible links that may exist in the future. In relation to these connections, the problem of link prediction appears as an important issue in the analysis of social networks (Zhao et al., 2024). This problem is in the sense of predicting the probability of establishing a connection between two vertices, with the knowledge that there is currently no connection between these two vertices. In other words, considering a snapshot of the current state of the network, the existence of new links between members should be predicted in the next snapshot.

Link prediction is one of the most important topics in social network analysis due to its applications in various issues such as social, political, economic, and security (Torabi et al., 2022; Jannesari et al., 2023). The dynamics of social networks have further complicated this issue, where the decision to discover new connections must be made online.

E-mail address: renxueping78@163.com.

https://doi.org/10.1016/j.jksuci.2024.101931
Received 12 November 2023; Received in revised form 25 December 2023; Accepted 13 January 2024
Available online 17 January 2024

Meanwhile, the heterogeneity of social networks complicates the problem of link prediction. Currently, there are many methods for link prediction. However, this issue is still open due to the importance of the topic and wide applications, and efforts to provide more accurate solutions are still ongoing. Among these issues, we can refer to the discussion of correctness, accuracy and time of transplant prediction in different methods (Cheng et al., 2024).

Recent literature on link prediction often includes simplex and monoplex/single-layer networks (**Yin et al. 2020). In these networks, all links are of the same type and the vertices are considered in the same layer/platform. However, real-world networks are often heterogeneous and multilayered. Basically, the links between vertices in multilayer/heterogeneous networks have different types. With a more comprehensive view, multilayer networks can be described as multiplex networks, where the same types of vertices in different layers include links of different types (Mishra et al., 2023). For example, a user is a member of two networks, Facebook and Twitter, and has different connections in each network. Obviously, this description provides intra-layer and inter-layer connections for each vertex in the network. Therefore, inter-layer links in multi-layer networks can only connect vertices where each vertex is an actor in different layers.

Link prediction in multilayer social networks as an example of multiplex networks has attracted the attention of many researchers (Liu et al., 2011). The problem of link prediction in these networks is addressed by collecting intra-layer and inter-layer information and then using them to calculate the similarity between vertices. Fig. 1 (a) shows an example of a monoplex network, and Fig. 1 (b) shows an example of a monoplex network. In multiplex networks, the embedding of vertices obtained from different layers can be used to solve the link prediction problem (Yu et al., 2017).

Most solutions use a similarity-based metric to solve the link prediction problem (Yan et al., 2023; Zhao et al., 2023d). A similarity metric can provide the probability of a link between two vertices in the future with a numerical value as a similarity score, where it is assumed that there is no direct link between two vertices. In this regard, a metric can calculate the similarity score of a target vertex with a set of non-adjacent vertices and then recommend a list of vertices with the highest score as future connections. The existing methods in the link prediction problem are divided into three general categories: vertex-based, structure-based, and social theory-based (Salimian et al., 2021; Berahmand et al., 2022). Vertex-based methods use unique features of vertices to calculate similarity scores. Structure-based methods use the topology structure of the network graph for the link prediction task. Also, social theory-based methods seek to discover similar users based on hypotheses such as belief, involvement, commitment, and attachment.

Among link prediction methods, structure-based methods are more common due to the use of topological information (Lim et al., 2021). These methods also include three categories based on neighbor, based on path and based on random walk (Berahmand et al., 2022). These categories have led to the emergence of similarity metrics with local, semi-local, and global characteristics (Zhao et al., 2023a; Xue et al., 2023). In local metrics, only the information of directly neighboring vertices such as degree are considered to calculate the similarity score. Semi-local metrics use the information of neighboring vertices with different and limited levels to estimate the similarity. Also, global metrics use the entire network information to calculate the similarity score. In general, local similarity metrics have low precision because they use limited information. However, these metrics have low time complexity. On the contrary, global metrics have higher accuracy with more complexity due to considering the information of the entire graph. Due to the high complexity of global metrics and the low accuracy of local metrics, semi-local metrics have been developed that focus on combining both local and global structures. These metrics consider the information of neighbors up to a fixed distance from a vertex to calculate similarity and can create a balance between complexity and accuracy.

The local random walk technique is one of the most well-known link prediction metrics based on the local structure, whose effectiveness has been proven for monoplex networks (Liu et al., 2011). In some studies, local random walk has been generalized for multilayer networks (Mishra et al., 2023). The important challenge for this problem is the high complexity of this technique to calculate inter-layer and intra-layer weights. To address this challenge, we introduce Extended Neighborhood based Local Random Walk (ENLRW) to develop local random walk semi-locally on multilayer networks. ENLRW uses the extended neighborhood concept as a distributed approach to identify nearest neighbors. In extended neighborhood, neighbors with different but limited levels/hops are considered. Therefore, ENLRW uses extended neighborhood concept to develop local random walk. Considering this definition, the walker can jump to neighbors with different but limited levels/hops at each step. In addition, ENLRW includes a set of inter-layer and intra-layer features to measure similarity in the form of reliable paths. Meanwhile, ENLRW uses vertex influence as a centrality index to improve the calculation of similarity between vertices to highlight edge strength in reliable paths.

The main contributions of this paper are as follows:

- The traditional local random walk technique is developed considering the extended neighborhood concept to be applicable to multilayer networks.
- Extended neighborhood includes a distributed approach to identify nearest neighbors. According to this approach, each vertex can independently identify a set of nearest neighbors with different levels.
- Several multimodal features are combined by considering intra-layer and inter-layer information to measure the similarity between

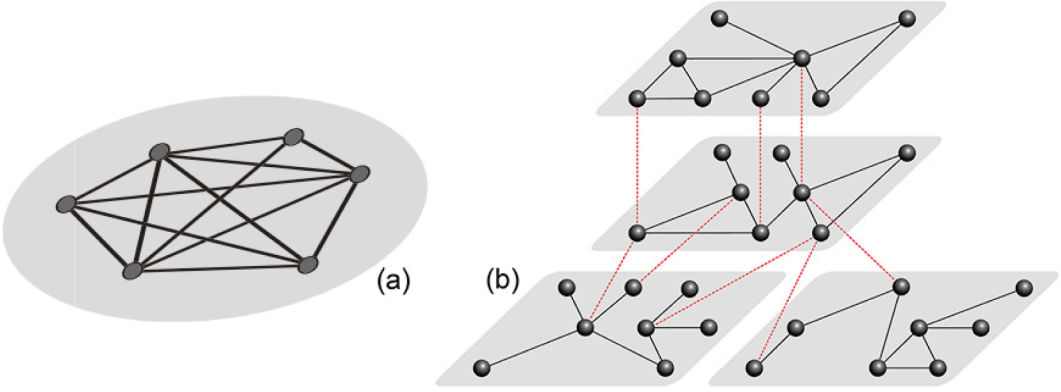

Fig. 1. An example of monoplex (a) and multiplex (b) networks. Black lines show the connection between vertices and red dashed lines refer to the connection of vertices with their counterparts in other layers.

vertices. In addition, we consider influence and reliable paths between vertices to calculate the similarity score.

In the rest of the paper, the preliminaries and classical similarity metrics is presented in Section 2. Section 3 describes the related works while the proposed ENLRW metric for link prediction is discussed in Section 4. In Section 5, we give the performance evaluation and discussion. Finally, the conclusion and future direction is given in Section 6.

2. Preliminaries

Before elucidating the proposed metric, we first discuss the system model and some classical similarity metrics that are further compared with our metric.

2.1. System model

Link prediction is formalized in $G = (V, E)$ as a monoplex/single-layer network. Here, V is the set of vertices and E is the set of edges, where $N = |V|$ and $M = |E|$. Let $v_i \in V$ and $v_j \in V$ be the i-th and j-th vertices in G, while $e_{ij} \in E$ represents the edge between v_i and v_j. The state of connections in G is formalized by the adjacency matrix A. In general, A is a binary matrix where each $a_{ij} \in A$ represents an edge state between v_i and v_j. Let $a_{ij} = 0$ and $a_{ij} = 1$ indicate no connection and connection between v_i and v_j, respectively. Assuming that t is a time parameter, $G[t]$ and $G[t+1]$ are two snapshots of G. Link prediction in G involves using $G[t]$ to recommend edges that are likely to appear in $G[t+1]$.

Link prediction in a multilayer network is formalized by $G = \{V, E^{(1)}, E^{(2)}, \cdots, E^{(k)}, \cdots, E^{(K)}\}$ find, where K is the number of layers and $E^{(k)}$ is the set of edges in the k-th layer. In fact, the set of vertices is the same in all layers, but each layer can have its own set of edges. The adjacency matrix in these networks is formalized by $A = \{A^{(1)}, A^{(2)}, \cdots, A^{(k)}, \cdots, A^{(K)}\}$, where $A^{(k)}$ is the detail adjacency matrix for the k-th layer. In this regard, $a_{ij}^{(z)} \in A^{(z)}$ shows the edge state between v_i and v_j in the k-th layer. Link prediction in a multilayer network involves using information from all layers to make recommendations at a target layer. Fig. 2 shows a hypothetical example of link prediction in a multiplex network with 2 layers including G_1 and G_2.

2.2. Similarity metrics

Link prediction is usually done by similarity metrics. A similarity metric by considering some topological, demographic or structural features can measure the weight of an edge as a similarity score between two related vertices. Edges with the highest similarity score are considered as future connections (Zhao et al., 2023b). Common Neighbor (CN), Preferential Attachment (PA), Sørensen Index (SI), Jaccard Coefficient (JC), and Adamic Adar (AA) are among the most well-known local similarity metrics, because they only use the information of direct neighbors to measure similarity (Long et al., 2022). Meanwhile, a global similarity metric uses information from the entire network to calculate a similarity score. If neighbors of different and limited levels are used to calculate the similarity score, it is considered as semi-local similarity metrics. Local Path (LP), FriendLink (FL), Katz Index (KI), Common Neighbors Degree Penalization (CNDP), and Local Random Walk (LRW) are some of the most well-known semi-local similarity metrics (Berahmand et al., 2022).

The CN metric (Newman, 2001) is defined as a local similarity metric based on the number of common neighbors between two vertices. This criterion is calculated as follows:

$$CN(v_i, v_j) = \left|\Gamma(v_i) \cap \Gamma(v_j)\right| \tag{1}$$

where $\Gamma(v_i)$ includes all neighbors with direct connection to v_i. Let $|x|$ denote the number of members in set x.

The PA metric (Barabâsi et al., 2002) uses the product of the degrees of two vertices as an index to calculate the similarity score. Therefore, this metric is developed under the assumption that new connections are more likely for higher degree vertices. The PA metric is defined as follows:

$$PA(v_i, v_j) = |\Gamma(v_i)| \cdot |\Gamma(v_j)| \tag{2}$$

The SI metric (Sorensen, 1948) was developed to measure the similarity between data samples on the ecological community. To measure the similarity score, SI takes into account both the factors of lower degrees of vertices and a greater number of common neighbors simultaneously with a higher probability of establishing a link. This metric is defined as follows:

$$SI(v_i, v_j) = \frac{\left|\Gamma(v_i) \cap \Gamma(v_j)\right|}{|\Gamma(v_i)| + |\Gamma(v_j)|} \tag{3}$$

The JC metric (Jaccard, 1901) consists of dividing the number of identical neighbors between two vertices by their total neighbors. In other words, this metric normalizes the number of common neighbors between two vertices. JC assigns a higher similarity score to pairs of vertices with a normalized number of common neighbors. This metric calculates the similarity score as follows:

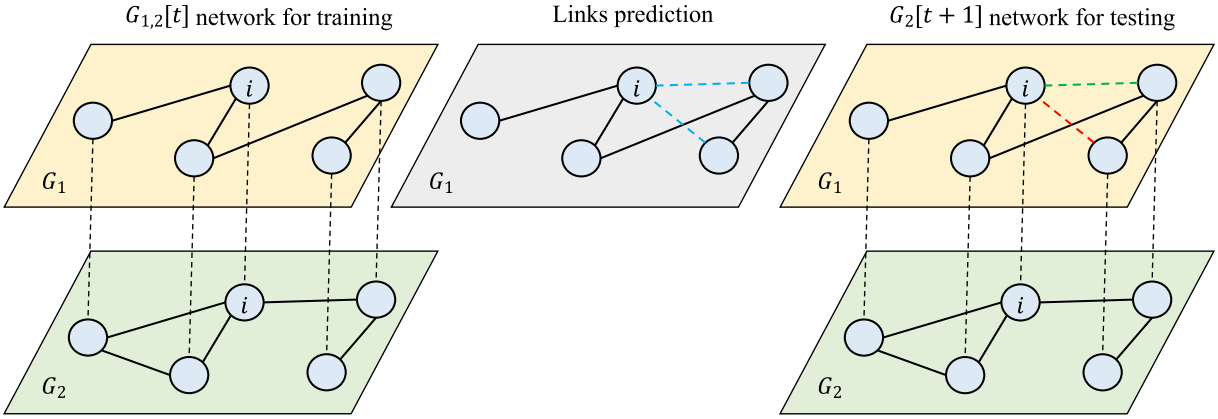

Fig. 2. Hypothetical example of link prediction in a 2-layer multiplex network. The link prediction for vertex v_i of the network G_1 is done as the target layer by considering the information of both networks $G_{1,2}$. Blue dashed lines indicate recommended links. In the two links recommended by the link prediction algorithm, one is correct (green dashed line) and one is incorrect (red dashed line).

$$JC(v_i, v_j) = \frac{|\Gamma(v_i) \cap \Gamma(v_j)|}{|\Gamma(v_i) \cup \Gamma(v_j)|} \tag{4}$$

The AA metric (Adamic and Adar, 2003) was presented as an index to measure the similarity between two web pages. However, this metric is widely used in social networks to calculate similarity. AA estimates the similarity score higher for two vertices whose common neighbors have fewer neighbors. This metric is defined as follows:

$$AA(v_i, v_j) = \sum_{v_k \in \Gamma(v_i) \cap \Gamma(v_j)} \frac{1}{log(k_{v_k})} \tag{5}$$

where k_{v_k} is the degree of vertex v_k which is available according to the adjacency matrix.

The LP metric (Lü et al., 2009) uses local path information including 2-hop and 3-hop to calculate the similarity score. Since the path with length 2 is more important than length 3, LP uses an impact coefficient. LP uses the number of paths with lengths 2 and 3 to calculate the similarity, which is provided by the adjacency matrix. Let A^l be the adjacency matrix to the l power that describes the number of paths of length l. This metric is defined as follows:

$$LP(v_i, v_j) = A^2\{a_{ij}\} + \beta \cdot A^3\{a_{ij}\} \tag{6}$$

where $A^2\{a_{ij}\}$ is the value of a_{ij} in the adjacency matrix to the power of 2 and represents the number of paths of length 2 between v_i and v_j. Also, β is an impact coefficient less than 1. Basically, β can assign higher importance to the number of shorter paths, and therefore is known as a damping factor.

The FL metric (Papadimitriou et al., 2012) can measure the similarity score between two vertices by traversing all paths with limited length. FL assumes that two users in the network can be connected through all paths of finite length. In addition to the number of paths, FL considers the path length coefficient and the total number of possible paths to calculate the similarity. This metric is defined as follows:

$$FL(v_i, v_j) = \sum_{l=2}^{L} \frac{1}{l-1} \cdot \frac{\left|\mathscr{P}_{ij}^{<l>}\right|}{\prod_{k=2}^{l}(N-k)} \tag{7}$$

where N and L are the total number of vertices in the network and the maximum level/length of neighborhood, respectively. Also, $\mathscr{P}_{ij}^{<l>}$ is the set of paths between v_i and v_j with length equal to l.

The KI metric (Katz, 1953) is one of the most well-known path-based techniques, which considers all paths between two vertices with finite lengths to measure similarity. In some studies, KI is considered as a global similarity metric, where the similarity score is calculated using all paths between two vertices in the network. Considering L as the maximum path length, the similarity score in KI is defined as follows:

$$KI(v_i, v_j) = \sum_{l=2}^{L} \beta^l \cdot \left|\mathscr{P}_{ij}^{<l>}\right| \tag{8}$$

where β is a damping factor.

The CNDP metric (Rafiee et al., 2020) uses a combination of some similarity metrics to calculate the similarity score between two vertices. According to the metrics used in the composition, CNDP can be classified as local, semi-local or global. The similarity score in this metric is defined as follows:

$$CNDP(v_i, v_j) = \sum_{v_k \in \Gamma(v_i) \cap \Gamma(v_j)} |\Gamma_{(v_k)}| \cdot |\Gamma(v_k)|^{-\beta C} \tag{9}$$

where $\Gamma_{(v_k)}$ is the set of common neighbors between v_k, v_i, and v_j. β is a constant coefficient and C is the average clustering coefficient in the network.

The LRW metric (Liu and Lü 2010) is a process that occurs over time on network vertices, and each step includes a walker's probabilistic-random jump on a neighboring vertex. In LRW, different distributions such as uniform and normal can be used. Assume that the walker is first located on the vertex v_i. In step τ, the walker can choose a $v_j \in \Gamma(v_i)$ as the next position to jump. According to Biased LRW, the probability of choosing each $v_j \in \Gamma(v_i)$ by the walker is different and is determined by using factors such as strength and degree. Biased LRW calculates the probability of a walker jumping from v_i to $v_j \in \Gamma(v_i)$ as follows:

$$P_{ij} = \frac{w_j . a_{ij}}{\sum_{v_j \in \Gamma(v_i)} w_j . a_{ij}} \tag{10}$$

where w_j is the weight/importance of v_j for biasing the walker steps.

Let $\pi(\tau)$ be a walker jump in step τ, and $\pi(\tau)$, $\forall(\tau = 0, 1, 2, 3, \cdots)$ be a sequence of walker steps. According to this, the similarity score by LRW can be calculated as follows:

$$LRW(v_i, v_j) = \frac{k_{v_i}}{2M} \cdot \pi_{ij}(\tau) + \frac{k_{v_i}}{2M} \cdot \pi_{ji}(\tau) \tag{11}$$

where $\pi_{ij}(\tau)$ shows the state of the walker jump in step τ between v_i and v_j.

3. Related works

In the real world, the concept of link prediction is used for many social services. Due to the complexity and diversity of social networks, many approaches have been proposed to solve the link prediction problem (Lim et al., 2021; Berahmand et al., 2022). Most of the works that have been done so far in the field of link prediction use the structural features of the network. These features include the interaction pattern and network connection structure. Some works try to reduce the size of the graph while maintaining the topological features in order to provide link prediction with higher accuracy and less complexity (Yan et al., 2023).

Improving the performance of link prediction by integrating local, global and association information can increase the accuracy of link prediction, where in the pessimistic case the complexity of combination-based metrics is high (Yin et al., 2010). There are different categories for link prediction solutions, which can be referred to centralized and decentralized approaches (Zhao et al., 2023c; Wang et al., 2023). These approaches are based on ensemble framework, random step, matrix partitioning, unsupervised techniques, and supervised techniques (Cheng et al., 2024). Insights into the literature show that recently there has been an increased interest in improving link prediction results through unsupervised approaches, where this can be effective in dealing with large-scale and complex networks (Li et al., 2015).

A deep model has been proposed to solve the problem of relational classification in which the relation between words in an assumed sentence is classified (Xue et al., 2022). This method is basically a combination of feature engineering and convolutional neural network (Zhao et al., 2022; Yue et al., 2023). This method is ineffective in dealing with the problem of link prediction in dynamic networks, because they only use the information of the link structure to predict future connections. DeepWalk (Keikha et al., 2021) is another model that solves the link prediction problem with relational data and using deep learning models. Using local information obtained from shortened random paths, DeepWalk can learn hidden profiles for each vertex as sentence equivalents. This method uses random walk and hierarchical approaches to create communities and improve link prediction. Although DeepWalk is related to relational data and deep learning, it is used to learn inappropriate representations for each vertex.

Li et al. (2015) presented the Node-Coupling Clustering (NCC) algorithm for link prediction. NCC combines network clustering information with common neighbors' information and provides a metric to calculate the similarity of two vertices. Aghabozorgi and Khayyambashi (2018) introduced a neighbor-based similarity metric called Triadic Measure (TM) that uses units called motifs. Motifs are various forms of

small oriented networks and consist of three vertices. Here, 13 different types of motifs are introduced. Each motif to calculate the similarity score, for each common neighbor such as v_k between two vertices v_i and v_j, the number of motifs consisting of v_i, v_j, and v_k are counted. This process is done for all 13 defined motifs and finally the average results are considered as the similarity score between v_i and v_j.

Recently, researchers have shown that network clustering contains very useful information for link prediction (Liu et al., 2011; Cao et al., 2023). Feng et al. (2012) clustered the network to discover the relationship between the clusters created from the graph structure. These relationships are used to improve the link prediction process. Bastami et al. (2019) presented a gravity-based link prediction method that uses the information of clusters in the network. The detected clusters are applied in parallel to the link prediction method to increase the speed of the task execution. Here, the authors use the AA metric to calculate the similarity score between vertices.

Some recent studies focus on combining network structural features, where the effectiveness of these features has been proven to predict connectivity (Yu et al., 2017). Yin et al. (2010) designed a link prediction framework that simultaneously uses graph structure and vertex characteristics. This unified framework is introduced as link recommendation (LINKREC). LINKREC is evaluated on co-authorship and co-starring collections. The results of this evaluation showed that combining graph information provides acceptable results compared to using only one feature. Here, the authors applied a tuning parameter to determine the degree of involvement of each feature.

Wang et al. (2011) showed that the similarity between the activities of two vertices is proportional to the distance between them. Therefore, the smaller the distance between two vertices, the greater the similarity between them. Here, several features were extracted from the information of the vertices, including activity and movement path, as well as some structural features of the network. The information of these features is modeled by a supervised classifier to be used for link prediction. This strategy increases the accuracy of forecasting significantly. Long et al. (2022) developed a Graph Neural Network (GNN)-based framework for link prediction in biological data. This method has been tested on data such as protein strands and drug molecular structure.

Kumar et al. (2022) presented a link prediction method based on vertex centrality metrics. Centrality metrics can provide local, semi-local and global structures of any network. Finally, these metrics are modeled as data features by different classifiers and link prediction is applied based on that. Here, the presence and absence of links are described by positive and negative labels for each vertex. The results showed that the use of light gradient boosted machine as a classifier provides better link prediction results. A similar work has been done by Rezaeipanah et al. (2020), where the authors use a Support Vector Machine (SVM) classifier for link prediction in multiplex online ego-social networks. Here, twitter and foursquare are used as a multiplex network with two layers for link prediction. The features used to build the classification model include nodal structure, ego-paths and *meta*-paths (SEM-Path).

Yang et al. (2022) presented the Reliable Paths has been developed for Multilayer Networks (RPMN) algorithm. RPMN is an efficient similarity metric that uses several topological features such as *meta*-paths to calculate link weights. The authors calculate the similarities based on reliable paths by considering the weights of paths between users. In a similar work, reliable paths were used by Gao and Rezaeipanah (2023) to extend the KI (EKI) metric. EKI uses intra-layer and inter-layer information to predict links in multiplex networks. EKI combines several topological features to find link weights and then reliable paths.

Mohamed et al. (2023) presented a method for inductive link prediction in knowledge graphs using locality-aware subgraphs. This method formulates the link prediction problem as a graph classification problem. In this problem, a subgraph is extracted for each vertex considering the neighborhood level, then a GNN is drawn for encoding and learning. The authors apply local clustering based on Personalized PageRank (PPR) to locality-aware subgraphs reasoning. Mishra et al. (2023) presented a method based on Higher Order Paths and layer fusion to improve Link Prediction in MULtiplex networks (HOPLP-MUL). This method seeks to find the regional influence of vertices efficiently by combining the structure of different layers. Here, link weights are calculated by combining the information of different layers to measure the relative density of the layers. After that, HOPLP-MUL can calculate the probability of the emergence of new links by considering longer paths between vertices. In addition, HOPLP-MUL considers the damping effect of long paths and density-based layer ranking for link prediction.

The existing works provide acceptable solutions for the link prediction problem. Most solutions work well on monoplex networks. However, many real social networks involve communication at multiple layers. Some works have shown that considering connections in different layers leads to improved link prediction in a target layer. However, there are many topological and multifaceted features encountered in multilayer networks. This has resulted in the problem of link prediction in multilayer networks still being challenging. For a better understanding of the related works, we have summarized the details of the reviewed study in Table 1.

4. Proposed metric

ENLRW is a similarity metric for link prediction in multilayer networks that combines several inter-layer and intra-layer features to formulate a reliable biased random walk. The flowchart of ENLRW is shown in Fig. 3. Here, the inter-layer features used in ENLRW include CN (Newman, 2001), SI (Sorensen, 1948), and FL (Papadimitriou et al., 2012). Also, Meta-Paths (MP) (Rezaeipanah et al., 2020), Clustering Meta-Path (CMP) (Jalili et al., 2017), and Semi-Local-Influence (SLI) (Shao et al., 2019) are used as intra-layer features in ENLRW. MP, CMP, and SLI are defined as follows:

- *MP metric (*Rezaeipanah et al., 2020*):* A *meta*-path with length 2 in a network with two layers G_1 and G_2 is described as $v_i \overset{G_1}{\rightarrow} v_k \overset{G_2}{\rightarrow} v_j$. $v_i \overset{G_1}{\rightarrow} v_k$ is an edge in G_1 defined by $a_{ik}^{(1)} = 1$, while $v_k \overset{G_2}{\rightarrow} v_j$ is an edge in G_2 defined by $a_{jk}^{(2)} = 1$. Here, v_k as an intermediate vertex allows connecting v_i to v_j with a path of length 2. MP is defined by the number of vertices similar to v_k where at least one edge exists in both layers. Therefore, the *meta*-paths associated with the MP metric include $v_i \overset{G_1}{\rightarrow} v_k \xrightarrow{G_1\&G_2} v_j$, $v_i \xrightarrow{G_1\&G_2} v_k \overset{G_1}{\rightarrow} v_j$, $v_i \overset{G_2}{\rightarrow} v_k \xrightarrow{G_1\&G_2} v_j$, $v_i \xrightarrow{G_1\&G_2} v_k \overset{G_2}{\rightarrow} v_j$, and $v_i \xrightarrow{G_1\&G_2} v_k \xrightarrow{G_1\&G_2} v_j$. In this definition, $G_1\&G_2$ indicates the existence of links in both layers G_1 and G_2.
- *CMP metric (*Jalili et al., 2017*):* A *meta*-path of length 2 based on clustering is defined by $v_i \overset{G_1}{\rightarrow} C \overset{G_2}{\rightarrow} v_j$, where C is considered a cluster instead of a vertex. In fact, C consists of a set of vertices with similar characteristics in the network. In this study, clustering of vertices is done by k-means (Liu, 2021) based on their degree. CMP is calculated by summing the size of clusters in each *meta*-path of length 2 based on clustering.
- *SLI metric (*Shao et al., 2019*):* This metric calculates the importance/influence of vertex v_j in terms of centrality index for possible connection with vertex v_i. SLI is a semi-local metric that calculates the overall influence based on the importance of vertex v_j itself as well as the importance of its nearest neighbors. Here, the set of vertices with maximum length L of v_j are considered as nearest neighbors. The SLI metric is calculated as follows:

$$SLI(v_i, v_j) = \frac{1}{\left|G_{N_L(v_j)}\right|} \cdot \left(\alpha \cdot k_{v_j} + \beta \cdot \sum_{l=1}^{L} \sum_{u \in \Gamma^{+l}(v_j)} \frac{k_u}{l} \right) \tag{12}$$

Table 1
Summary of related works.

Reference	Abbreviation	Methodology	Advantages	Weaknesses
Keikha et al. (2021)	DeepWalk	Deep learning based on random walk and hierarchical approaches	Modeling relational data with deep learning and also considering user communities	Increasing complexity due to the use of hierarchical approaches
Li et al. (2015)	NCC	Combining network clustering information with common neighbors	Ensuring scalability by applying local clustering	Complexity in setting the primary parameters
Aghabozorgi and Khayyambashi (2018)	TM	Using motifs units to calculate similarity based on common neighbors	Combination of 13 different motifs from common neighbors	Failure to consider reliability of paths between users and scalability
Yin et al. (2010)	LINKREC	Using graph structure and vertex characteristics	Simultaneous consideration of graph structure and vertex characteristics	High complexity of algorithm in collecting vertex characteristics
Rezaeipanah et al. (2011)	SEM-Path	Using SVM to classify features of nodal structure, ego-paths and *meta*-paths	Combining different features and modeling them with classification methods	Perform simulations only for small scale networks
Yang et al. (2022)	RPMN	Identifying reliable paths by combining several topological features	Combining several topological features to weight the network and find reliable paths	There is no effective strategy for combining topological features
Gao and Rezaeipanah (2023)	EKI	Development of a weighted version of KI metric based on reliable paths	Considering the strength of links in addition to the number of paths between users	Inefficiency for complex scenarios
Mohamed et al. (2023)	PPR	Using locality-aware subgraphs and PageRank-based clustering for graph classification	Formulating the problem as a combined classification problem	Not considering the effect of network scale in finding local subgraphs
Mishra et al. (2023)	HOPLP-MUL	Ranking of density-based layers and higher order paths to find regional influence of vertices	Finding the regional influence of vertices by combining the structure of different layers	Failure to control the criteria related to the estimation of the density of the layers in the influence rate of the vertices

where $G_{N_L(v_j)}$ is the extracted subgraph for v_j of maximum length L, k_{v_j} is the vertex degree of v_j, and $\Gamma^{+l}(v_j)$ is the set of neighbors of length l of v_j. Also, α and β are the impact coefficient for SLI, which shows the influence of vertex v_j and its nearest neighbors, respectively.

According to the framework defined for ENLRW, the weight of each edge in target layer l is calculated by combining intra-layer (i.e., CN, SI, and FL) and inter-layer (i.e., MP, CMP, and SLI) features. Let $w(v_p, v_q)$ be the weight of the edge e_{pq} that represents the strength between vertices v_p and v_q:

$$w(v_p, v_q) = \xi_{Intra} \cdot [CN(v_p, v_q) + SI(v_p, v_q) + FL(v_p, v_q)] + \xi_{Inter} \cdot [[MP(v_p, v_q) + CMP(v_p, v_q) + SLI(v_p, v_q)]] \tag{13}$$

where ξ_{Intra} and ξ_{Inter} are the impact coefficient for intra-layer and inter-layer features, respectively, which are optimally adjusted by Taguchi technique.

Obviously, the similarity score is calculated between vertices v_i and v_j, which are not directly connected. In other words, we assume in the link prediction problem that vertices v_i and v_j are connected through paths. Without loss of generality, the weight of a link is the strength of that link in the path. Hence, we measure the weight between v_i and v_j as a reliable path by multiplying the weight of all links in that path. $w^{(\mathscr{P})}(v_i, v_j)$ is calculated as the weight between v_i and v_j in the path $\mathscr{P}$ as follows:

$$w^{(\mathscr{P})}(v_i, v_j) = \prod_{(v_p, v_q) \in \mathscr{P}} w(v_p, v_q) \tag{14}$$

where $w^{(\mathscr{P})}(v_i, v_j)$ is considered as the similarity score in the ENLRW metric:

$$ENLRW(v_i, v_j) = w^{(\mathscr{P})}(v_i, v_j) \tag{15}$$

In this paper, the path $\mathscr{P}$ is found by LRW. LRW for multiple networks with different strategies such as LRW Classical (LRWC), LRW Maximal Entropy (LRWME), LRW Diffusive (LRWD), LRW Physical (LRWP) and LRW Lévy (LRWL) have already been defined (Berahmand and Bouyer, 2019). Among the existing methods, we use Biased LRW (BLRW) (Nasiri et al., 2023) to find the path $\mathscr{P}$ considering multiple layers. Traditional BLRW defines walker jumps based on layers G_1 and G_2 in a two-layer network. Here, the walker is located on the vertex v_i in layer G_1. According to BLRW policy, the walker has four choices to jump on each step: 1) stay in v_i of G_1, 2) jump to a $v_k \in \Gamma_1(v_i)$ in the same layer G_1, 3) jump from v_i to its counterpart in layer G_2, and 4) jump from v_i to a $v_k \in \Gamma_2(v_i)$ in layer G_2. Each choice to jump has a probability defined by $P_i^{(1)}$, $P_{ik}^{(1)}$, $P_i^{(1\to2)}$, and $P_{ik}^{(1\to2)}$, respectively. A schematic of BLRW for a multiple network with two layers G_1 and G_2 is shown in Fig. 4.

According to BLRW, we define the jumping probabilities from vertex v_i in layer G_1 as follows:

$$P_i^{(1)} = \frac{k_{v_i}^{(1)}}{\sum_{v_k \in \Gamma_1(v_i)} k_{v_k}^{(1)}} \tag{16}$$

$$P_{ik}^{(1)} = \frac{w_1(v_i, v_k).a_{ik}^{(1)}}{\sum_{v_k \in \Gamma_1(v_i)} w_1(v_i, v_k).a_{ik}^{(1)}} \tag{17}$$

$$P_i^{(1\to2)} = \frac{k_{v_i}^{(1)} + k_{v_i}^{(2)}}{\sum_{v_k \in \Gamma_1(v_i)} k_{v_k}^{(1)} + \sum_{v_k \in \Gamma_2(v_i)} k_{v_k}^{(2)}} \tag{18}$$

$$P_{ik}^{(1\to2)} = \frac{w_1(v_i, v_k).a_{ik}^{(1)} + w_2(v_i, v_k).a_{ik}^{(2)}}{\sum_{v_k \in \Gamma_1(v_i)} w_1(v_i, v_k).a_{ik}^{(1)} + \sum_{v_k \in \Gamma_2(v_i)} w_2(v_i, v_k).a_{ik}^{(2)}} \tag{19}$$

where $k_{v_i}^{(1)}$ is the degree of vertex v_i in layer G_1, $w_1(v_i, v_k)$ is the weight between vertices v_i and v_k in layer G_1, $\Gamma_1(v_i)$ is the set of neighbors of v_i in layer G_1, and $a_{ik}^{(1)}$ is the connection status of vertices v_i and v_k in layer G_1.

If $\pi(\tau)$, $\forall(\tau = 0, 1, 2, 3, \cdots)$ is a sequence of walker's steps to form the path $\mathscr{P}$, finally selecting the vertex v_j leads to the completion of the path. Henceforth, $\mathscr{P} = \{v_i, \cdots, v_k, \cdots, v_j\}, \forall v_k \in \pi(\tau)$ formalizes the final path. BLRW is only based on the selection of direct neighbors, while ENLRW develops BLRW based on the extended neighborhood concept by selecting the nearest neighbors with the maximum level L. Therefore, in each step, the walker has the ability to jump from v_i to the set of neighbors up to the maximum level L. The nearest neighbors of the v_i vertex in the layer G_z are defined as follows based on the extended neighborhood concept:

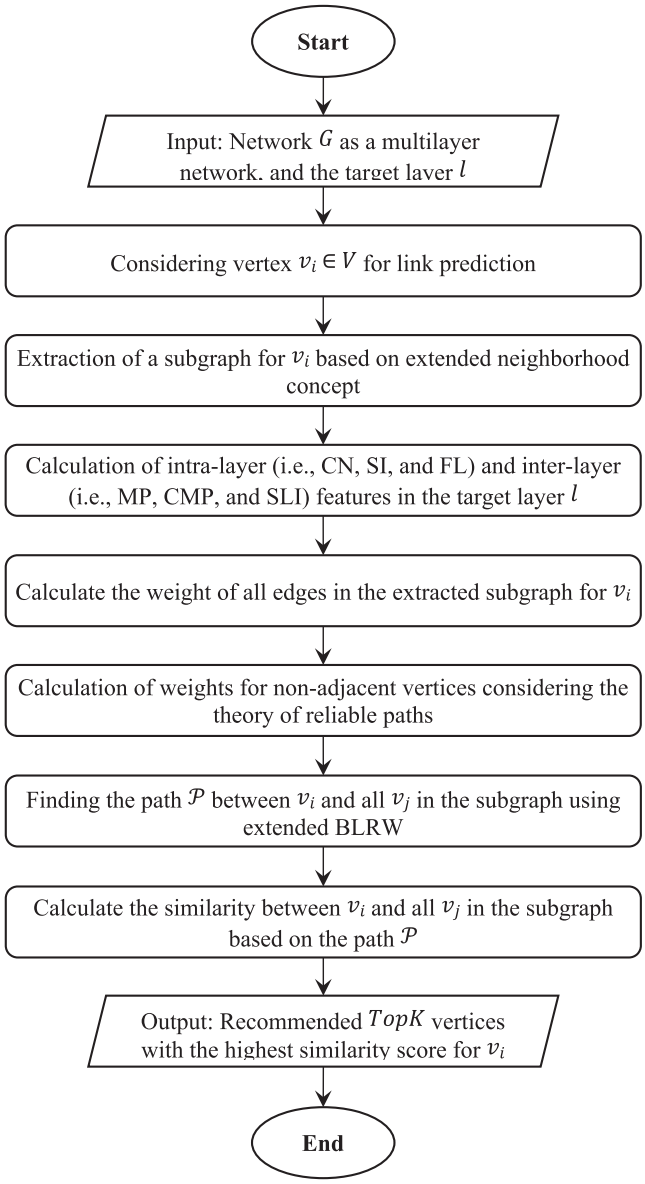

Fig. 3. Flowchart for the proposed ENLRW metric.

$$v_k \in \Gamma_z^l(v_i), \forall l = 1, 2, \cdots, L \tag{20}$$

where $\Gamma_z^l(v_i)$ is the set of all neighbors of vertex v_i with length l in layer G_z. Also, L is the maximum length/hop/level to identify nearest neighbors.

Direct neighbors are available through the adjacency matrix, but identifying a set of nearest neighbors with different hops is challenging. This problem involves extracting a subgraph for each vertex with a set of nearest neighbors with different levels. Although techniques such as Rubin (Guo et al., 2022) have the ability to find all paths between vertices and extract subgraphs based on it, but they have high complexity. To address this challenge, we use the extended neighborhood concept to extract subgraphs, which includes a distributed technique to identify the nearest neighbors with maximum length L. This technique can maintain high efficiency in facing large-scale networks.

ENLRW uses only one sub-graph including 1-hop, 2-hop, …, and L-hop as nearest neighbors to develop BLRW, where L is the maximum path length for nearest neighbors. ENLRW uses a distributed method to create a neighbor table for each vertex. The process of identifying the nearest neighbors related to the vertex v_k considering the extended

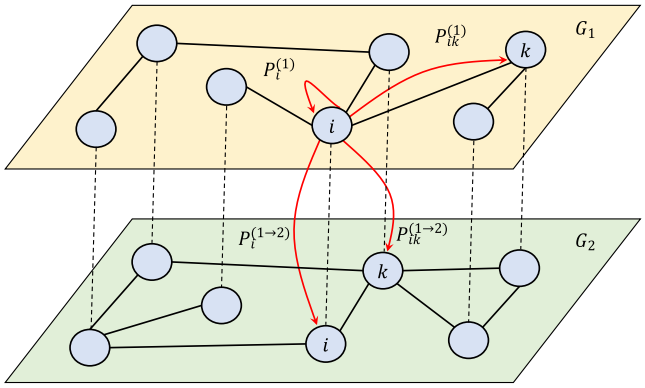

Fig. 4. Schematic of BLRW for a multiple network with two layers G_1 and G_2.

neighborhood with the maximum level L is as follows:

(1) Vertex v_k sends control packet Pk_α to all its neighbors. The details of the packet Pk_α are shown in Fig. 5(a).
(2) If $v_s \in \Gamma(v_k)$, then vertex v_s receives the packet Pk_α.
  1%2 According to Fig. 5(b), vertex v_s updates the packet Pk_α and sends it to all its neighbors. If the "PathLength" field in the updated packet is equal to '0', then the vertex v_s deletes the packet Pk_α and does not send it.
  2%2 Vertex v_s generates control packet Pk_β as response. The details of the packet Pk_β is shown in Fig. 5(c), where v_s sends it to all its neighbors. Here, "$SourceID_{Pk_\alpha}$" indicates the content of the "SourceID" field of the packet Pk_α.
(3) If $v_d \in \Gamma(v_s)$, then vertex v_d receives the packet Pk_β.
  1%2 According to Fig. 5(d), vertex v_d updates the packet Pk_β and sends it to all its neighbors. If the "PathLength" and "DestinationID" fields in the updated packet are equal to '0' and 'v_k', respectively, then vertex v_d deletes the packet Pk_β and does not send it.
  2%2 If the "DestinationID" field in the updated packet Pk_β is equal to 'v_k' (that is, vertex v_d is the same as vertex v_k), then the content of the field "SourceID" (that is, vertex v_s) is stored in the neighbor table of vertex v_k.

According to this process, for each vertex such as v_i, $\Gamma^l(v_i)$ can be found with a distributed approach and low complexity. By accessing the neighbor table, ENLRW can identify the path $\mathscr{P}$ using BLRW. Meanwhile, path $\mathscr{P}$ can lead to the measurement of $w^{(\mathscr{P})}(v_i, v_j)$ and, thus $ENLRW(v_i, v_j)$ for each pair of vertices v_i and v_j. Here, $ENLRW(v_i, v_j)$ calculates the similarity of each $v_i \in E^{te}$ with all $v_j \in E^{tr}$ as the recommendation metric, and then $TopK$ vertices from the E^{tr} set with the highest similarity are presented as final recommendations. Obviously, $a_{ij}^{(l)}, \forall v_i \in E^{te}, v_j \in E^{tr}$ in the target layer l is not available for ENLRW. In other words, the purpose is to calculate the probability of creating a connection associated with $a_{ij}^{(l)}$ in the future, where its accuracy can be evaluated through the E set.

5. Analysis of simulations

This section performs several numerical simulations on a real-world dataset to analyze the rationality of the proposed ENLRW compared to equivalent methods. The ENLRW and all link prediction methods used in the comparisons are simulated using MALTAB R2021a. All evaluations were performed by a computer system with Intel® Core™ i5 Processor (3.3–3.8 GHz), 16 GB RAM, and Windows 11 operating system. For fair comparisons, the results are reported based on the average of 10 runs, while the dataset and experimental setup are the same for all methods. It is worth noting that the optimal values for ξ_{Intra} and ξ_{Inter} are found as

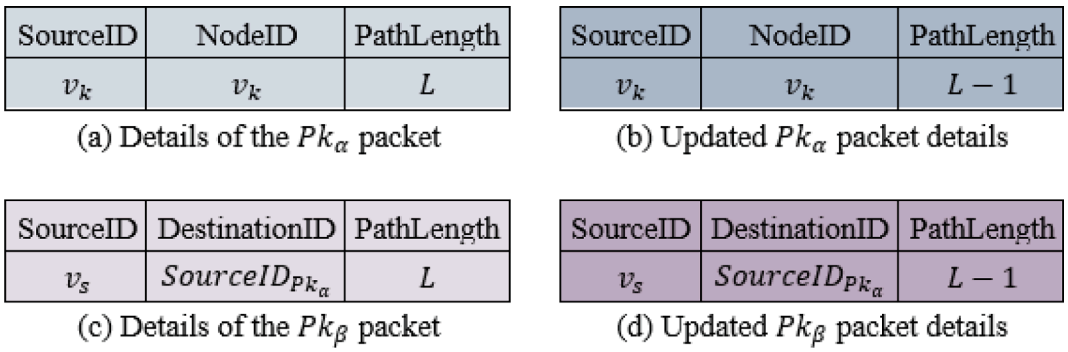

Fig. 5. Details of the control packets to create the neighbor table.

two important coefficients in the metric proposed by the Taguchi technique. According to this, the optimal values for these parameters are 0.35 and 0.65, respectively.

5.1. Dataset

To evaluate different link prediction methods, we use the Twitter-Instagram dataset, which is collected from social networks Instagram and Twitter. This dataset can be downloaded for free from https://data.world/datasets. Since the same users/vertices from both networks are needed to formulate the link prediction problem in multiple networks, we use the available profiles to find the same users. Thus, a Twitter user is identical to an Instagram user if the similarity of their profiles exceeds a threshold value. This leads to the identification of the same users in both networks, but the problem of extracting the subgraph associated with these users remains. To address this issue, we use Breadth First Search (BFS) and Depth First Search (DFS) algorithms to find all paths between users. Applying these algorithms to each network leads to finding the subgraph associated with the same users. The information of these sub-graphs is used for link prediction. In total, we found 13,298 users as identical users on both Twitter and Instagram networks. Table 2 shows a summary of the dataset used as a two-layer network. It is worth noting that the available data from these networks is related to November 2012.

5.2. Performance criteria

To perform link prediction, we divide the set of network links into two sets E^{tr} (training set) and E^{te} (testing set). Link prediction is performed for each vertex in E^{te} set that has at least one link in E^{tr} set. Here, for each vertex of E^{te}, $TopK$ of vertices of E^{tr} are recommended. Every $v_i \in E^{te}$ that has at least one link with the E^{tr} set is considered for the link prediction process. Let Δ_{v_i} be the list of users recommended by a similarity metric for vertex v_i, where $TopK = |\Delta_{v_i}|$. Also, let $TopK = |\Delta_{v_i}|$ be the set of all true relevant users with vertex v_i, where $TrueK_{v_i} = |\vartheta_{v_i}|$. In other words, every ϑ_{v_i} represents $a_{ij} = 1$. According to this definition, the recommended number of related vertices for vertex v_i is as follows:

$$RN_{v_i} = |\Delta_{v_i} \cap \vartheta_{v_i}| \tag{21}$$

Considering RN_{v_i}, various performance criteria such as precision, recall and f-measure can be calculated to evaluate the similarity metrics in the link prediction problem (Liu et al., 2023a, 2023b). Precision represents the ratio of the number of recommended related vertices to the total number of recommended vertices. Recall is the ratio of the number of recommended relevant vertices to the total number of true relevant vertices. Also, f-measure is defined as the weighted harmonic mean of precision and recall. Because the performance criteria are calculated for each user, we report the final performance metrics based on the average for all test users. Eqs. (22)–(24) formulate precision, recall and f-measure, respectively.

$$Precision = \frac{1}{|E^{te}|}\sum_{v_i \in E^{te}} \frac{RN_{v_i}}{TopK} \tag{22}$$

$$Recall = \frac{1}{|E^{te}|}\sum_{v_i \in E^{te}} \frac{RN_j}{TrueK_{v_i}} \tag{23}$$

$$F\_measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{24}$$

Area Under the Curve (AUC) is one of the common measures to quantify the accuracy of link prediction methods (Mishra et al., 2023). Given the recommended links, AUC depicts the probability of recommending an existing link at random relative to the probability of recommending a non-existing link. For each similarity metric, one recommended existing link and one non-existing recommended link are randomly selected for each vertex. After that, an existing link and a non-existing link are randomly selected to compare the scores of each similarity metric. Finally, the AUC measure for each similarity metric is calculated as follows:

$$AUC = \frac{n' + 0.5n''}{n} \tag{25}$$

where n is the total number of comparisons, n' is the number of comparisons in which the recommended link has a higher score, and n'' is the number of comparisons with the same score. Obviously, a higher AUC indicates a better performance of the similarity metric.

5.3. Results and discussion

The proposed metric for link prediction is evaluated and compared with classical similarity metrics as well as equivalent similarity metrics. Here, CN (Newman, 2001), KI (Katz, 1953), and LRW (Liu and Lü 2010) are used as classical similarity metrics as well as RPMN (Yang et al., 2022) and EKI (Gao and Rezaeipanah, 2023) as equivalent similarity metrics for comparison. We evaluate ENLRW in comparison with LRW to show the effectiveness of extended neighborhood in local random walk. Also, we chose the RPMN method for comparison because it uses BLRW for link prediction and has a similar structure to the proposed metric. Meanwhile, EKI is another state-of-the-art method to compare with ENLRW, where, similar to the proposed metric, it uses reliable paths as link weights. In the evaluations, we consider two scenarios, monoplex and multiplex, on the Twitter-Instagram network. In the monoplex

Table 2
A summary of the Twitter-Instagram dataset extracted from the Twitter and Instagram networks.

Networks	Number of vertices	Number of links	Average degree	Density
Twitter	13,298	52,668	28.3	0.0214
Instagram	13,298	227,794	34.6	0.0266

scenario, information extracted from the Instagram network is used to predict links in the same network, while the multiplex scenario includes using information from both Instagram and Twitter networks to predict links in Instagram. Therefore, in both scenarios, Instagram is set as the target layer for link prediction. In the simulations, the recommended number of vertices as *TopK* is set to 10. A *TopK* of 10 has been used as a reference value in various studies (Jalili et al., 2017; Gao and Rezaeipanah, 2023; Li et al., 2023).

According to Li et al. (2023), we set β to 0.05 in the simulations. The maximum neighborhood level (i.e., L) is another important and adjustable parameter in the proposed ENLRW metric. Insights into the literature confirm that a suitable range of L is between 1 and 5 (Jalili et al., 2017; Yang et al., 2022; Gao and Rezaeipanah, 2023). The results of setting different values of L on ENLRW are reported in Table 3 for both monoplex and multiplex scenarios. The results are given based on the precision, recall and f-measure criteria, where the better values are in bold. According to the average results, we found the best value for L equal to 3.

According to the analysis, the best performance is obtained for ENLRW with $L = 3$ and $\beta = 0.05$. In general, the effect of the L parameter is greater than the β parameter in the results, because the change of L leads to the change of the information considered for link prediction. In general, $L = 1$ considers only first-level neighbors for link prediction, which is limited information. When L increases to 2 and 3, obviously more relevant information is available for link prediction. As depicted, the increase of L has led to an increase in the performance of ENLRW in link prediction. However, $L = 4$ and $L = 5$ lead to performance degradation, because most of the information added from level 4 and 5 neighbors for link prediction is irrelevant and redundant. On the other hand, link prediction in multiplex scenario has better performance than monoplex, because more relevant information from two layers is used for link prediction. On average, f-measure in multiplex scenario has increased by 7.8 % compared to monoplex.

In another experiment, we evaluate ENLRW against some equivalent similarity metrics such as CN, KI, LRW, RPMN, and EKI. The evaluation is based on precision, recall and f-measure with different values of *TopK*. Here, we analyze *TopK* with different values from 1 to 30. Fig. 6 shows the comparison results for the precision criterion. Fig. 6(a) relates to the monoplex scenario, where link prediction is applied to the Instagram network. Also, Fig. 6(b) shows the precision criterion results for the multiplex scenario, where the link prediction is done considering the Twitter-Instagram network for the Instagram layer. Meanwhile, Figs. 7 and 8 report similar results for the recall and f-measure criteria.

As illustrated, the proposed ENLRW metric has obtained acceptable results in both scenarios and outperforms all existing methods. Since more information is available for link prediction in multiplex networks, it was expected that the results of the performance criteria for the multiplex scenario would be better than the monoplex scenario. The precision results show that the increase of *TopK* leads to a decrease in the value of this criterion. Meanwhile, the increase of *TopK* leads to the increase of the recall criterion. The reason for this process is related to the way these criteria are calculated, where the denominator of the fraction in precision is equal to *TopK* and the denominator of the fraction in recall is equal to *TrueK*. According to these concepts, the best precision is always obtained with $TopK = 1$ and the best recall with $TopK = 30$. Meanwhile, f-measure provides the average precision and recall and is suitable for evaluating the performance of link prediction methods.

Considering the monoplex scenario, ENLRW with $TopK = 8$ provided the best value for f-measure. Considering $TopK = 8$, the f-measure for the proposed metric is equal to 52.46 %. RPMN and EKI methods are ranked second and third with 49.88 % and 49.19 %, respectively. According to these results, the proposed metric performs more than 5.17 % better than the best existing method in the monoplex scenario in terms of f-measure. Moreover, ENLRW significantly outperforms all classical similarity metrics (i.e., CN, KI and LRW). This superiority is expected, since classical similarity metrics only use first-level neighborhood information. Considering the multiplex scenario, the proposed metric has provided the highest f-measure value of 65.31 % with $TopK = 7$. With this *TopK*, EKI with 63.28 % and RPMN with 62.42 % are in the next ranks in terms of f-measure, respectively. ENLRW outperforms EKI by 3.21 % and RPMN by 4.63 % in multiplex scenario. For better clarity of evaluations, we have summarized the results of comparisons for both monoplex and multiplex scenarios in Table 4. In these tables, better values are highlighted in bold.

In the last experiment, we compared different link prediction methods in terms of AUC score in Table 5. In these tables, better values are highlighted in bold. In the evaluations, we achieved an AUC score of 77.84 % for link prediction in the monoplex scenario, where multimodal features of Instagram are used for link prediction in this network. The AUC score for the multiplex scenario is higher than that of Instagram internal link prediction, because the multiplex features from both Instagram and Twitter are used for Instagram link prediction. In link prediction for Instagram, which includes information from both networks, the AUC score is 86.63 %. The average results also show the significant superiority of the proposed ENLRW metric with an AUC score of 82.24 %. According to the average results obtained, ENLRW is superior in terms of AUC score compared to CN, KI, LRW, RPMN, and EKI by 23.6 %, 11.9 %, 10.5 %, 3.0 %, and 2.3 %, respectively.

ENLRW has performed better than all classical link prediction metrics including local, semi-local, and global in almost both defined scenarios and has achieved the best precision, recall, f-measure, and AUC. From these results, it can be understood that using the centrality index and multimodal features as reliable paths to calculate the weight of edges can greatly improve the link prediction performance. It is clear that in some cases global metrics have performed better than all methods in terms of precision, especially with the number of recommendations less than 5. Compared with local and semi-local methods, it can be shown that the proposed metric has a competitive and yet efficient performance compared to global metrics. More precisely, in all comparisons, ENLRW has a better average precision than KI and LRW. Also, ENLRW has higher precision than equivalent methods such as RPMN and EKI in most cases. For example, ENLRW outperforms RPMN and EKI in all *TopKs*. This issue proves the efficiency of using the mutual influence of vertices as a centrality index in calculating the similarity between vertices.

Table 3
Evaluation of ENLRW with different values of L parameter.

Scenario	Performance criteria	$L = 1$	$L = 2$	$L = 3$	$L = 4$	$L = 5$
Monoplex network	Precision	63.87	77.25	**78.69**	72.85	68.74
	Recall	49.46	57.11	**58.19**	55.96	53.35
	F-measure	55.75	65.67	**66.90**	63.30	60.08
Multiplex network	Precision	69.98	82.32	**83.52**	77.45	72.35
	Recall	52.63	62.11	**63.35**	59.68	56.33
	F-measure	60.08	70.80	**72.05**	67.41	63.34
Average	Precision	66.92	79.79	**81.11**	75.15	70.54
	Recall	51.05	59.61	**60.77**	57.82	54.84
	F-measure	57.91	68.24	**69.48**	65.35	61.71

6. Conclusion

The expansion of online social networks has led to attracting a significant number of users. Meanwhile, the analysis of these networks has raised new perspectives and challenges due to the large amount of information. The problem of link prediction is one of the basic issues and an important prerequisite of social network analysis. Link prediction has diverse applications in fields such as bioinformatics, dating, information retrieval, and e-commerce. In this problem, an overview of a network is given and a link prediction algorithm must estimate the probabilities associated with creating links between network entities. In recent years,

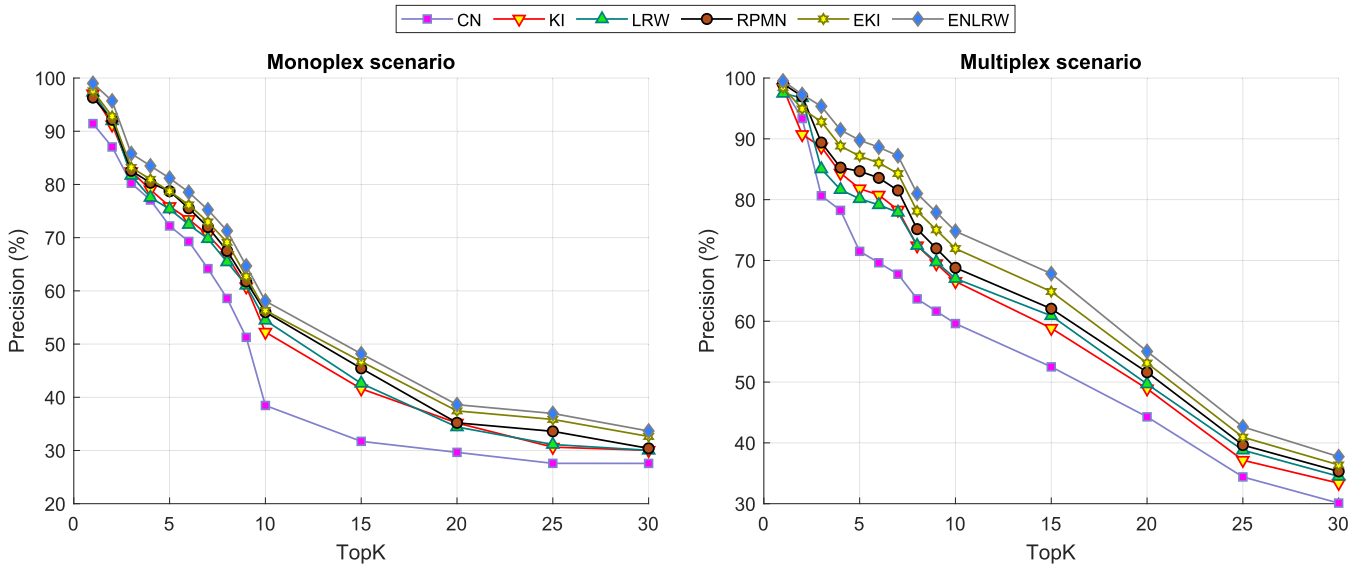

Fig. 6. Precision results in different link prediction methods considering monoplex and multiplex scenarios.

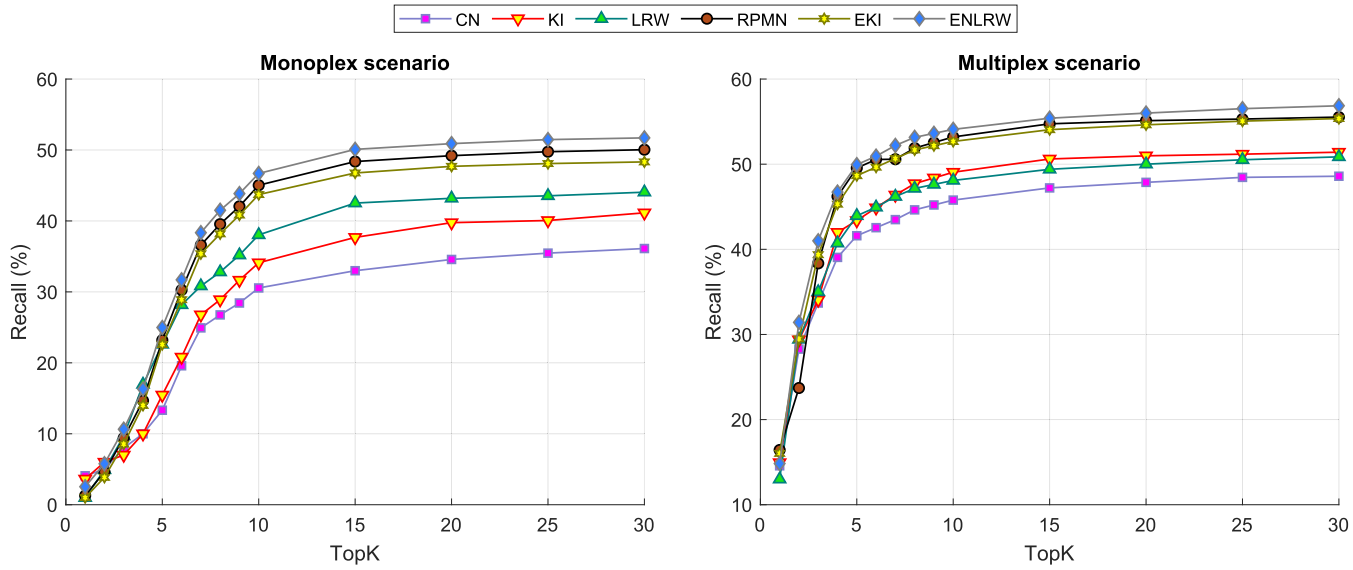

Fig. 7. Recall results in different link prediction methods considering monoplex and multiplex scenarios.

link prediction in multilayer social networks has attracted the attention of many researchers. In this paper, we presented an extended version of the local random walk technique to address the link prediction problem in multilayer social networks. In order to reduce the complexity of the local random walk technique in multilayer networks, we proposed extended neighborhood based local random walk. Considering the extended neighborhood, this technique can identify the nearest neighbors of a vertex in a distributed manner. Also, using the information of different layers in multilayer networks has led to the improvement of the performance of local random walk based on extended neighborhood. Moreover, we calculate similarities based on vertex influence and some topological features through reliable paths to consider the strength of connections in link prediction. Our observations from numerical simulations prove that the use of local random walk based on extended neighborhood as well as vertex influence and topological features for link prediction achieves acceptable complexity and precision on multilayer networks. The proposed metric in the multiplex scenario and compared to EKI as the second-best method in terms of f-measure and AUC has provided 3.2 % and 2.3 % better recommendations, respectively.

Although the proposed metric recommends more suitable vertices considering both monoplex and multiplex scenarios compared to the existing metrics, it ignores the fact that multilayer networks can be a useful tool for social bootstrapping. As a limitation for the proposed metric, interactions between users should be considered for friend recommendations due to their comprehensive perspective on the online social “ecosystem”. Meanwhile, reducing the processing complexity of multilayer networks through parallelization as well as identifying reliable paths between vertices is recommended as part of future work. The generalization of the proposed metric in other similar fields such as anomaly detection, intrusion detection, modeling, text analysis, community detection, and recommender systems is worth future studies. Furthermore, the analysis of the proposed metric on the newer datasets such as TwiBot-22 is suggested as a future work.

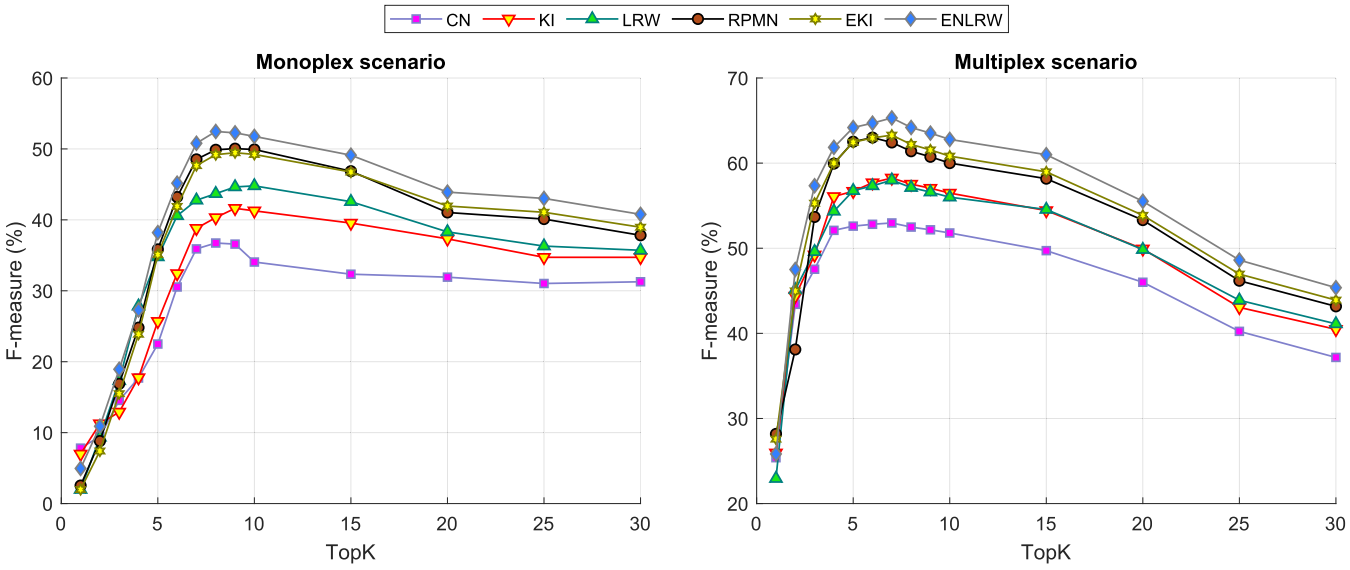

Fig. 8. F-measure results in different link prediction methods considering monoplex and multiplex scenarios.

Table 4
Average results for different methods based on $TopK = 8$ in the monoplex scenario and $TopK = 7$ in the multiplex scenario.

Scenario	Performance criteria	CN	KI	LRW	RPMN	EKI	ENLRW
Monoplex network	Precision	58.57	66.62	65.42	67.52	69.12	**71.28**
	Recall	24.93	26.77	30.83	36.60	35.37	**38.34**
	F-measure	36.73	40.32	43.69	49.88	49.18	**52.46**
Multiplex network	Precision	63.68	72.42	72.45	75.15	78.11	**81.00**
	Recall	43.49	46.38	46.20	50.58	50.66	**52.20**
	F-measure	52.97	58.25	58.00	62.42	63.28	**65.31**

Table 5
AUC results in different link prediction methods considering monoplex and multiplex scenarios.

Scenario	CN	KI	LRW	RPMN	EKI	ENLRW
Monoplex network	65.76	71.28	70.50	75.16	77.00	**77.84**
Multiplex network	67.27	75.65	78.36	84.60	83.78	**86.63**
Average	66.52	73.47	74.43	79.88	80.39	**82.24**

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adamic, L.A., Adar, E., 2003. Friends and neighbors on the web. Soc. Networks 25 (3), 211–230.

Aghabozorgi, F., Khayyambashi, M.R., 2018. A new similarity measure for link prediction based on local structures in social networks. Physica A 501, 12–23.

Barabâsi, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. Physica A 311 (3–4), 590–614.

Bastami, E., Mahabadi, A., Taghizadeh, E., 2019. A gravitation-based link prediction approach in social networks. Swarm Evol. Comput. 44, 176–186.

Berahmand, K., Bouyer, A., 2019. A link-based similarity for improving community detection based on label propagation algorithm. J. Syst. Sci. Complex. 32, 737–758.

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. Inform. Sci. 34 (8), 5375–5387.

Cao, C., Wang, J., Kwok, D., Cui, F., Zhang, Z., Zhao, D., Zou, Q., 2022. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic Acids Res. 50 (D1), D1123–D1130.

Cao, Y., Xu, N., Wang, H., Zhao, X., Ahmad, A.M., 2023. Neural networks-based adaptive tracking control for full-state constrained switched nonlinear systems with periodic disturbances and actuator saturation. Int. J. Syst. Sci. 54 (14), 2689–2704.

Cheng, F., Niu, B., Xu, N., Zhao, X., 2024. Resilient distributed secure consensus control for uncertain networked agent systems under hybrid DoS attacks. Commun. Nonlinear Sci. Numer. Simul. 129, 107689.

Feng, X., Zhao, J.C., Xu, K., 2012. Link prediction in complex networks: a clustering perspective. Eur. Phys. J. B 85, 1–9.

Gao, Z., Rezaeipanah, A., 2023. A novel link prediction model in multilayer online social networks using the development of Katz similarity metric. Neural Process. Lett. 55 (4), 4989–5011.

Guo, F., Zhou, W., Lu, Q., Zhang, C., 2022. Path extension similarity link prediction method based on matrix algebra in directed networks. Comput. Commun. 187, 83–92.

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.

Jalili, M., Orouskhani, Y., Asgari, M., Alipourfard, N., Perc, M., 2017. Link prediction in multiplex online social networks. R. Soc. Open Sci. 4 (2), 160863.

Jannesari, V., Keshvari, M., Berahmand, K., 2023. A novel nonnegative matrix factorization-based model for attributed graph clustering by incorporating complementary information. Expert Syst. Appl. 242, 122799.

Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika 18 (1), 39–43.

Keikha, M.M., Rahgozar, M., Asadpour, M., 2021. DeepLink: A novel link prediction framework based on deep learning. J. Inf. Sci. 47 (5), 642–657.

Kumar, S., Mallik, A., Panda, B.S., 2022. Link prediction in complex networks using node centrality and light gradient boosting machine. World Wide Web 25 (6), 2487–2513.

Li, F., He, J., Huang, G., Zhang, Y., Shi, Y., Zhou, R., 2015. Node-coupling clustering approaches for link prediction. Knowl.-Based Syst. 89, 669–680.

Li, W., Li, T., Berahmand, K., 2023. An effective link prediction method in multiplex social networks using local random walk towards dependable pathways. J. Comb. Optim. 45 (1), 31.

Lim, M., Abdullah, A., Jhanjhi, N.Z., 2021. Performance optimization of criminal network hidden link prediction model with deep reinforcement learning. J. King Saud Univ.-Comput. Inform. Sci. 33 (10), 1202–1210.

Liu, G., 2021. Data collection in mi-assisted wireless powered underground sensor networks: directions, recent advances, and challenges. IEEE Commun. Mag. 59 (4), 132–138.

Liu, Z., Feng, J., Uden, L., 2023b. Technology opportunity analysis using hierarchical semantic networks and dual link prediction. Technovation 128, 102872.

Liu, W., Lü, L., 2010. Link prediction based on local random walk. Europhys. Lett. 89 (5), 58007.

Liu, S., Niu, B., Karimi, H.R., Zhao, X., 2024. Self-triggered fixed-time bipartite fault-tolerant consensus for nonlinear multiagent systems with function constraints on states. Chaos Solit. Fractals 178, 114367.
Liu, Z., Zhang, Q.M., Lü, L., Zhou, T., 2011. Link prediction in complex networks: A local naïve Bayes model. Europhys. Lett. 96 (4), 48007.
Liu, X., Zhou, G., Kong, M., Yin, Z., Li, X., Yin, L., Zheng, W., 2023a. Developing multi-labelled corpus of twitter short texts: a semi-automatic method. Systems 11 (8), 390.
Long, Y., Wu, M., Liu, Y., Fang, Y., Kwoh, C.K., Chen, J., Li, X., 2022. Pre-training graph neural networks for link prediction in biomedical networks. Bioinformatics 38 (8), 2254–2262.
Lü, L., Jin, C.H., Zhou, T., 2009. Similarity index based on local paths for link prediction of complex networks. Phys. Rev. E 80 (4), 046122.
Mishra, S., Singh, S.S., Kumar, A., Biswas, B., 2023. HOPLP− MUL: link prediction in multiplex networks based on higher order paths and layer fusion. Appl. Intell. 53 (3), 3415–3443.
Mohamed, H.A., 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.
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.
Newman, M.E., 2001. Clustering and preferential attachment in growing networks. Phys. Rev. E 64 (2), 025102.
Papadimitriou, A., Symeonidis, P., Manolopoulos, Y., 2012. Fast and accurate link prediction in social networking systems. J. Syst. Softw. 85 (9), 2119–2132.
Rafiee, S., Salavati, C., Abdollahpouri, A., 2020. CNDP: Link prediction based on common neighbors degree penalization. Physica A 539, 122950.
Rezaeipanah, A., Ahmadi, G., Sechin Matoori, S., 2020. A classification approach to link prediction in multiplex online ego-social networks. Soc. Netw. Anal. Min. 10 (1), 27.
Salimian, M., Ghobaei-Arani, M., Shahidinejad, A., 2021. Toward an autonomic approach for Internet of Things service placement using gray wolf optimization in the fog computing environment. Software: Pract. Exp. 51 (8), 1745–1772.
Shao, Z., Liu, S., Zhao, Y., Liu, Y., 2019. Identifying influential nodes in complex networks based on Neighbours and edges. Peer-to-Peer Networking and Applications 12, 1528–1537.
Sorensen, T., 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. Biologiske Skrifter 5, 1–34.
Torabi, E., Ghobaei-Arani, M., Shahidinejad, A., 2022. Data replica placement approaches in fog computing: a review. Clust. Comput. 25 (5), 3561–3589.
Wang, D., Pedreschi, D., Song, C., Giannotti, F., Barabasi, A.L., 2011, August. Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1100-1108.
Wang, T., Zhang, L., Xu, N., Alharbi, K.H., 2023. Adaptive critic learning for approximate optimal event-triggered tracking control of nonlinear systems with prescribed performances. Int. J. Control. https://doi.org/10.1080/00207179.2023.2250880.
Xue, B., Yang, Q., Xia, K., Li, Z., Chen, G.Y., Zhang, D., Zhou, X., 2022. An AuNPs/mesoporous NiO/nickel foam nanocomposite as a miniaturized electrode for heavy metal detection in groundwater. Engineering. https://doi.org/10.1016/j.eng.2022.06.005.
Xue, B., Yang, Q., Jin, Y., Zhu, Q., Lan, J., Lin, Y., Zhou, X., 2023. Genotoxicity Assessment of Haloacetaldehyde Disinfection Byproducts via a Simplified Yeast-Based Toxicogenomics Assay. Environ. Sci. Tech. 57 (44), 16823–16833.
Yan, S., Gu, Z., Park, J.H., Xie, X., 2023. A delay-kernel-dependent approach to saturated control of linear systems with mixed delays. Automatica 152, 110984.
Yang, R., Yang, C., Peng, X., Rezaeipanah, A., 2022. A novel similarity measure of link prediction in multi-layer social networks based on reliable paths. Concurrency Comput.: Pract. Exp. 34 (10), e6829.
Yin, Z., Gupta, M., Weninger, T., Han, J., 2010, April. Linkrec: a unified framework for link recommendation with user attributes and graph structure. In: Proceedings of the 19th international conference on World wide web, pp. 1211-1212.
Yu, C., Zhao, X., An, L., Lin, X., 2017. Similarity-based link prediction in social networks: A path and node combined approach. J. Inf. Sci. 43 (5), 683–695.
Yue, S., Niu, B., Wang, H., Zhang, L., Ahmad, A.M., 2023. Hierarchical sliding mode-based adaptive fuzzy control for uncertain switched under-actuated nonlinear systems with input saturation and dead-zone. Robotic Intell. Automation 43 (5), 523–536.
Zhang, H., Zou, Q., Ju, Y., Song, C., Chen, D., 2022. Distance-based support vector machine to predict DNA N6-methyladenine modification. Curr. Bioinform. 17 (5), 473–482.
Zhao, Y., Liang, H., Zong, G., Wang, H., 2023d. Event-Based Distributed Finite-Horizon H∞Consensus Control for Constrained Nonlinear Multiagent Systems. IEEE Syst. J. 17 (4), 5369–5380.
Zhao, J., Sun, M., Pan, Z., Liu, B., Ostadhassan, M., Hu, Q., 2022. Effects of pore connectivity and water saturation on matrix permeability of deep gas shale. Adv. Geo-Energy Res. 6 (1), 54–68.
Zhao, H., Wang, H., Niu, B., Zhao, X., Xu, N., 2024. Adaptive fuzzy decentralized optimal control for interconnected nonlinear systems with unmodeled dynamics via mixed data and event driven method. Fuzzy Sets. Syst. 474, 108735.
Zhao, H., Wang, H., Xu, N., Zhao, X., Sharaf, S., 2023a. Fuzzy approximation-based optimal consensus control for nonlinear multiagent systems via adaptive dynamic programming. Neurocomputing 553, 126529.
Zhao, H., Zong, G., Wang, H., Zhao, X., Xu, N., 2023b. Zero-Sum Game-Based Hierarchical Sliding-Mode Fault-Tolerant Tracking Control for Interconnected Nonlinear Systems via Adaptive Critic Design. IEEE Trans. Autom. Sci. Eng. https://doi.org/10.1109/TASE.2023.3317902.
Zhao, H., Zong, G., Zhao, X., Wang, H., Xu, N., Zhao, N., 2023c. Hierarchical Sliding-Mode Surface-Based Adaptive Critic Tracking Control for Nonlinear Multiplayer Zero-Sum Games Via Generalized Fuzzy Hyperbolic Models. IEEE Trans. Fuzzy Syst. 31 (11), 4010–4023.