IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 14, 2020, accepted September 30, 2020, date of publication October 6, 2020, date of current version October 19, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3029122

Taxonomy of Link Prediction for Social Network Analysis: A Review

HERMAN YULIANSYAH, (Graduate Student Member, IEEE), ZULAIHA ALI OTHMAN, AND AZURALIZA ABU BAKAR
Center for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia

Corresponding author: Herman Yuliansyah (herman.yuliansyah@tif.uad.ac.id)

This work was supported by the Universiti Kebangsaan Malaysia under Grant FRGS/1/2019/ICT02/UKM/02/7.

ABSTRACT Link prediction is a technique to forecast future new or missing relationships between entities based on the current network information. Graph theory and network science are theoretical concepts that have influenced the link prediction research. Although previous reviews clearly outlined the link prediction research, it was focused on describing prediction approaches only. However, analysis of related studies identified other components that influence link prediction. This review aims to present a continued review and introduce the taxonomy of link prediction using three main components: the prediction approaches, prediction features, and prediction measurements. Each component has been detailed using its own taxonomy available at the present review. Furthermore, this review compares the prediction approaches and prediction features also benchmark algorithms and measurement methods of previous link prediction studies. In conclusion, the previous studies mostly focused on structural features and similarity-based approaches, while measuring the proposed methods using the Area Under the Curve (AUC) score. The proposed link prediction taxonomy can guide the researchers to generate new ideas and innovations that contribute to the link prediction research.

INDEX TERMS Social network analysis, link prediction, prediction approaches, prediction features, prediction measurements.

I. INTRODUCTION

In 1999, Barabási & Albert conducted a study on connectivity between node pairs due to adding new nodes to a growing network, and the new nodes were attached to a node that connected before [1]. This concept is called "Preferential Attachment." Furthermore, in 2001, M.E.J. Newman conducted the study on connectivity between node pairs by using the "triangle" concept, which comprises three nodes (A, B, C) [2]. These nodes were connected to the two edges, namely AB and AC, which led to the emergence of BC based on the rule of triadic closure, which formed a "connected triple." In 2003, Adamic & Adar also examined the similarity of interactions between node pairs for Internet networks by proposing the Adamic/Adar algorithm [3]. Later, Liben-Nowell & Kleinberg, in 2007, formalized this new interaction as a "Link Prediction Problem" and developed an approach known as "Common Neighbors" based on the proximity of nodes in the network [4]. Other studies, such as the Hub Promoted Index (HPI) [5], the Hub Depressed Index (HDI) [5], the Leicht-Holme-Nerman-1 (LHN1) index [6], and Resource Allocation (RA) also use the "triangle" concept to address particular problems by proposing new algorithms [7]. Ravasz *et al.* proposed HPI and HDI to measure the topology in metabolic networks in 2002 [5]. In 2006, Leicht *et al.* proposed the LHN1 index by comparing the degree of similarity of node pairs with each degree [6]. Furthermore, in 2009, T. Zhou *et al.* proposed the RA algorithm, which was driven by the process of resource allocation that occurred in the network [7]. The concept of common neighbors and similarity, namely the Jaccard Similarity Coefficient (JC) [8] and Salton Cosine Similarity (SA) [9], are identical.

The link prediction problems in static unweighted networks are formulated as follows: given a real network $G = (V, E)$ at time t, where G is a graph or network, V is a set of nodes or vertices, and E denotes a set of edges or links at time t. The set of nodes and edges of a network are also denoted by V(G) and E(G), respectively. Furthermore, $E(G) = (u, v)$ or $(u, v) \in E(G)$, where u and v are a pair of nodes with $u \neq v$ and $u, v \in V$. Also, in the experimental research design,

The associate editor coordinating the review of this manuscript and approving it for publication was Santhosh Kumar Gopalan.

set of edges are separated edge test/probe (E^p) and edge train (E^t) where E^t is the current edges of the network at time t, E^p is the predicted edges of the network at time $t+1$, $E^p \cup E^t = E(G)$, and $E^p \cap E^t = \emptyset$ [10]. The candidate node pair (CNP) is generated to predict the future or missing link during interval time t to t+1. CNP is formulated as follows: $(u, v) \notin E^t$. Here, E^p consists of the actual and not actual (fake) edges in the future network. Therefore, to conduct experimental research designs, the observed edges are divided into 10% for E^p and 90% for E^t, usually [11]. Furthermore, the predicted results from E^p are measured to determine the performance.

Graph theory and network science are the theoretical concepts that have influenced the link prediction research with its problem commonly measured based on the proximity of a predicted pair of nodes in a network. The concept of common neighbors served as an inspiration for link prediction proximity measurement. It can be illustrated by two people's have excellent opportunities to meet and form links with many mutual friends [12]. However, network growth, size, and time changing affect the real-world network. Therefore, many factors need to be considered during evolution. The graph-based taxonomy on the location-based social network proposed by Kefalas *et al.* [13] comprises five categories that affect location-based social networks: data factors/features, data representation, methodologies and models, recommendation types, and personalization. Therefore, these five categories can be used to provide the basis for classifying influence factors and combining them according to the results presented in previous studies.

Link prediction has been implemented in several research areas, such as social networks [12], [14], [15], co-authorship networks [16]–[18], marketing and economic networks [19], [20], terrorist networks [21], recommending systems [22], [23], health domain [24], and others. In the real world, social networks (SN) are considered as a new domain for link prediction research. Social networking sites have become famous for interacting and sharing information among users [25], [26]. Here, nodes represent the SN users, and edges denote their relationship or interactions. The Internet's evolution significantly contributed to SN's development because many people add new friends and share their activity regularly. However, the corresponding policy control has not yet been investigated adequately [27]. Therefore, with the addition of new nodes and edges, SN can be characterized as highly dynamic, growing, changing quickly, and massive. Real-world networks are formed by hundreds of thousands or even millions of nodes. Thereby, the techniques to perform link prediction must be highly efficient [28].

The previous reviews have clearly outlined the prediction approaches used in the link prediction research. For example, Linyuan and Zhou [29] presented a minireview that summarizes link prediction algorithms based on physical perspectives and approaches, while Al Hasan and Zaki [30] surveyed a survey of link prediction that focuses on approaches and social network graphs. Furthermore, Wang *et al.* [31] have systematically categorized link prediction based on techniques and problems, and their application to introduction roadmap active research groups. Wang *et al.* [31] also discussed future opportunities for link prediction research on social networks. Martínez *et al.* [28] reviewed link prediction with general-purpose techniques and continued with domain-specific heuristic methods. Later, Haghani and Keyvanpour [32] categorized link prediction based on a technical approach and have discussed its strengths and weaknesses. However, the components capable of influencing the link prediction research are yet to be thoroughly explained.

This review provides a continued review of the link prediction approaches and adds new contribution by analyzing prediction features and prediction measurements to fill in the previous literature review gaps. Prediction approaches are defined as ways/techniques/algorithms proposed by previous studies to address challenges in the prediction link problem, which are classified into similarity-based and learning-based approaches. Furthermore, prediction features are defined as unique attributes or particular aspects of the previous studies that are related to the data and network character being studied. Prediction features are classified into data features, data representations, network types, and network categories. Data features and network types are characteristics of data and network types capable of influencing a prediction. In turn, data representation means composing the data in such a way that it represents the data visualization results. The study also showed that every domain corresponding to link prediction implementation has different network characteristics. Therefore, the network categories are included to denote the type of dataset in the link prediction research. This review also utilized the benchmark algorithms and prediction measurement methods to summarize the most widely used ones. Methods for measuring the proposed method are classified into prediction measurements. This review also outlines the proposed methods, which were mostly measured to evaluate the algorithm's performance using the Area Under the Curve (AUC) score.

Besides the link prediction review, the taxonomy of link prediction, which includes the prediction approaches, prediction features, and prediction measurements, is introduced. The proposed taxonomy includes a hierarchical classification by naming and description. The taxonomy of link prediction needs to be visualized because it helps categorize related studies, making it easy to understand and find new insights in future studies. Several comparisons have also been conducted to achieve this objective by presenting the influence factors and facilitating new ideas and innovations. This present review fills in the previous literature review studies' gaps by providing an overview of the prediction approaches, prediction features, and prediction measurements. Finally, the main objectives of this review are as follows:

1) To present a continued link prediction review.
2) To propose a taxonomy of link prediction with three components, namely the prediction approaches, prediction features, and prediction measurements.

3) To present a comparison between prediction approaches and prediction features.
4) To present a comparison of benchmark algorithms and measurement methods in the existing link prediction studies.

The remaining sections of this paper are organized as follows. Section 2 represents the link prediction literature review, while in section 3, the taxonomy is proposed, which includes the three components that influenced link prediction. Section 4, 5, and 6 provide more detail about the taxonomies of the prediction approaches, prediction features, and prediction measurements. Finally, Section 7 concludes and discusses the main guidelines for future research.

II. RELATED WORK

The prediction approaches are classified based on the initial assumptions to simplify understanding the link prediction approaches. Different terms are used to denote prediction approaches, as shown in Table 1. The concept of the similarity-based method is mentioned in three studies to denote the topological information-based approaches, namely, by Linyuan and Zhou [29], Martínez *et al.* [28], and Pandey *et al.* [33]. There are two groups representing the similarity-based methods. The first group classifies link prediction into the neighbor based, path-based, and random walk-based methods [4], [30]–[32]. The methods based on node neighborhoods adopt the common sense that two nodes x and y are more likely to start a new relationship when having the common neighbors [4] and reflecting the personal interest and social behavior [31]. Furthermore, the methods based on the combination of all paths adopt the idea of the shortest path where the distance implicitly considers the ensemble of all paths between two nodes [4]. In turn, the second group classifies link prediction into the local similarity, global similarity, and quasi-local similarity methods [28], [29], [33]. Therefore, another classification is proposed in terms of similarity-based methods, namely, representing a hybrid of local and global similarity in a weighted network. Pandey *et al.* [33] suggested it. There is no difference in algorithm members between the neighbor based and local similarity methods, but it does not hold for other methods. Most path-based algorithms can be grouped into global similarity, but this can be done not for all random walk-based algorithms. However, the algorithms can also be grouped into quasi-similarity.

In addition to the similarity-based method, the previous reviews also introduce other methods based on non-topological information such as methods based on linear algebraic methods, probabilistic, and statistical models, and classification models, as shown in Table 2. Various names are used to classify prediction approaches. There are only two studies where the authors use the term 'learning-based method' to denote the non-topological information-based approaches, namely, Wang *et al.* [31] and Haghani and Keyvanpour [32].

An in-depth analysis is conducted by mapping the classification differences that have been carried out by previous reviews to determine the relationship between these reviews. Each algorithm and method or model classified by the related reviews is marked with the symbol V to identify classification patterns used by them. Algorithms are grouped in a single column if the reviews mention them repeatedly to show frequency of review.

In this review, we consider using the term 'similarity-based approaches' according to the comprehensive analysis of algorithms, as shown in Table 1, and also conclude that similarity-based approaches can be classified into local similarity, global similarity, and quasi-local similarity methods. Local similarity methods are a neighbor-based method, and global similarity methods are a combination of path-based and random walk-based methods. Furthermore, quasi-local similarity methods are a hybrid of local and global similarity. This classification is deemed to be more dynamic and acceptable for future use, as the local similarity-based methods allow using local information of neighbor proximity, and global similarity-based methods use global information from all networks. In contrast, quasi-local is a combination of local information and network information. The algorithms classified as local similarity-based ones are CN, JA, AA, PA, RA, SA, SO, LLHN, HPI, HDI, PD, RA-CNI, IA, MI, LNB, CAR, FSW, and LIT. In turn, the global similarity-based algorithms are Katz, SR, GLHN, RWR, PFP, MFI, RPR, SP, CCS, RSS, VCP, NSP, RW, MERW, PLM, RFK, BI, and Path distance. Lastly, quasi-local similarity-based algorithms are LP, HT, FL, LRW, SRW, ORA-CNI, and EM.

In the present review, we also conclude using the term 'learning-based approaches' according to the classification of the several related approaches, as shown in Table 2. Learning-based approaches use a learning model to predict hidden links in a network [30] and consider the features provided by previous basic link prediction metrics, internal attributes, and external information [31]. In turn, the maximum likelihood method uses detailed rules and particular parameters to calculate the possibility of unobserved links assuming several principles for organizing network structures and maximizing the potential for observed structures [29].

Classification and latent feature-based models are the key elements of the learning-based approaches [30], [32]. The classification-based model employs features extractions of similarity-based methods, then trains the results using classification models to predict the next missing link [34]. Furthermore, statistical and probabilistic studies have also provided the background for developing link prediction techniques based on statistical analysis and probability theory. Probabilistic models are used to abstract the structure of an observed network and predict the unobserved links using the learnt model [29]. Probabilistic models can describe the relationships in the network structure by utilizing network information [34].

Link prediction based on node attributes, correlation information, influential nodes, and artificial intelligence-based

TABLE 1. Classifying the topological information-based algorithms.

Authors, Year, Ref	Similarity-based approaches	CN	JC	AA	PA	SA	SO	HPI	HDI	LHN1	RA	Katz	HT	RPR	SR	LHN2	ACT/CT	RWR	FL	CST	PFP	LP	LRW	SRW	Others
Liben-Nowell & Kleinberg, 2007, [4]	Node Neighborhoods	V	V	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Ensemble of All Paths	-	-	-	-	-	-	-	-	-	-	V	V	V	V	-	-	-	-	-	-	-	-	-	-
Linyuan & Zhou, 2011, [29]	Similarity-based:																								
	• Local similarity	V	V	V	V	V	V	V	V	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	• Global similarity	-	-	-	-	-	-	-	-	-	-	V	-	-	V	V	V	V	-	V	-	-	-	-	A1
	• Quasi-local	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	V	V	-
Hasan & Zaki, 2011, [30]	Feature based Link Prediction:																								
	• Node Neighborhood based Features	V	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	• Path based Features	-	-	-	-	-	-	-	-	-	-	V	V	V	-	-	-	-	-	-	-	-	-	-	A2
	• Features based on Vertex and Edge Attributes	-	-	-	V	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	A3
T. Wang & Liao, 2014, [34]	Traditional Link Prediction:																								
	• Similarity-based Node-based Metrics	V	V	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P. Wang *et al.*, 2015, [31]	Topology-based Metrics:																								
	• Neighbor-based	V	V	V	V	V	V	V	V	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	A4
	• Path-based	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	V	-	-	V	-	-	A5-A6
	• Random Walk based	-	-	-	-	-	-	-	-	-	-	-	V	V	V	-	V	-	-	V	V	-	-	-	-
Martínez *et al.*, 2016, [28]	Similarity-based methods:																								
	• Local	V	V	V	V	V	V	V	V	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	A7-A13
	• Global	-	-	-	-	-	-	-	-	-	-	V	-	-	V	V	V	V	-	-	V	-	-	-	A14-A19
	• Quasi-local	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	V	V	V	A20
Haghani & Keyvanpour, 2017, [32]	Heuristic based approaches:																								
	• Node neighborhoods	V	V	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	• Ensemble of all paths	-	-	-	-	-	-	-	-	-	-	V	V	V	V	-	-	-	-	-	-	-	-	-	-
Pandey *et al.* 2019, [33]	Similarity based methods:																								
	• Local similarity	V	V	V	V	V	V	V	-	V	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	• Global similarity	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	A21
	• Quasi-local matrix	-	-	-	-	-	-	-	-	-	-	-	V	V	V	-	V	-	-	-	V	-	-	V	A22

Note: CN = Common Neighbors[4], JC = Jaccard Coefficient[8], AA = Adamic/Adar[3], PA = Preferential Attachment[1], SA = Salton Cosine Similarity[9], SO = Sørensen Index[35], HPI = Hub Promoted Index[5], HDI = Hub Depressed Index[5], LHN1 = Leicht-Holme-Nerman-1 Index[6], RA = Resource Allocation[7], Katz = Katz Index[36], HT = Hitting Time, RPR = Rooted Pagerank, SR = simRank[37], LHN2 = Leicht-Holme-Newman-2 Index[6], ACT = Average Commute Time, RWR = Random Walks with Restart[38], FL = FriendLink[39], CT = Commute Time, CST = Cosine Similarity Time, PFP = PropFlow Predictor[40], LP = Local Path Index[41], LRW = Local Random Walk[42], SRW = Superposed Random Walk[42], A1 = Matrix Forest Index (MFI)[43], A2 = Shortest Path Distance (SP), A3 = Clustering Coefficient Score (CCS), A4 = Parameter-Dependent (PD)[44], A5 = Relation Strength Similarity (RSS)[45], A6 = Vertex Collocation Profile (VCP)[46], A7 = Resource Allocation Based on Common Neighbor Interactions (RA-CNI)[47], A8 = Individual Attraction Index (IA)[47], A9 = Mutual Information (MI)[48], A10 = Local Na¨ıve Bayes (LNB)[49], A11 = CAR-Based Indices (CAR)[50], A12 = Functional Similarity Weight (FSW)[51], A13 = Local Interacting Score (LIT)[52], A14 = Negated Shortest Path (NSP), A15 = Random Walks (RW)[42], A16 = Maximal Entropy Random Walk (MERW)[53], A17 = Pseudoinverse of the Laplacian Matrix (PLM)[54], A18 = Random Forest Kernel Index (RFK)[55], A19 = Blondel Index (BI)[56], A20 = Third-Order Resource Allocation Based on Common Neighbor Interactions (ORA-CNI)[47], A21 = Path distance, A22 = Evidential measurement (EM)[57].

methods corresponds to the most recent learning-based approaches classification [33]. Node attributes are used to predict unobserved links and the correlation information method uses information related to the node-to-node, edge-to-edge, or node-to-edge relations [33]. Furthermore, the most influential node can be identified using several methods, such as deanonymization, learning spectral graph transformations, ranking factor graph model, transfer-based model, game theory, diverse node adoption algorithm, and the balanced modularity-maximization model [33]. The artificial intelligence-based methods can be applied to reduce the computation complexity, overhead, cost to predict [33].

TABLE 2. Classifying the non-topological information-based methods.

Learning-based Approaches	Authors							
	[4]	[29]	[30]	[34]	[31]	[28]	[32]	[33]
Linear Algebraic Methods	-	-	V	-	-	-	-	-
Probabilistic and statistical Models	-	V	V	-	V	V	-	V
Clustering models	V	-	-	-	-	-	-	-
Classification models	-	-	V	-	-	V	V	-
Social theory based metrics	-	-	-	-	V	-	-	-
Algorithmic methods:	-	-	-	-	-	V	-	-
Factorization-based methods	-	-	-	V	V	V	V	-
Latent-feature-based model	-	-	-	-	-	-	V	-
Learning Automata	-	-	-	-	-	-	-	V
Deep Learning	-	-	-	-	-	-	-	V
Artificial neural network	-	-	-	-	-	-	-	V
Dempster-Shafer theory	-	-	-	-	-	-	-	V
Network Information's	-	-	-	-	-	-	-	V

Note: [4]=(Liben-Nowell and Kleinberg 2007), [29]=(Linyuan and Zhou 2011), [30]=(Hasan and Zaki 2011), [34]=(T. Wang and Liao 2014), [31]=(P. Wang *et al.* 2015), [28]=(Martínez *et al.* 2016), [32]=(Haghani and Keyvanpour 2017), [33]=(Pandey *et al.* 2019).

In summary, probability and statistical, and factorization-based and supervised learning methods become the most explored ones among the learning-based approaches for link prediction. Artificial intelligence and information networks' adoption corresponds to the latest innovation for link prediction based on learning-based approaches. Besides the link prediction approaches, analysis of prediction features and prediction measurements are also considered, even though it was not investigated in the previous reviews in detail. The variant of the dataset, measurement methods, and the network features are provided in Table 3.

All the previous reviews focus on prediction approaches, and only Pandey *et al.* [33] explored the network features, as shown in Table 3. However, the network features were explored only on static and dynamic networks. Later, link prediction analysis was conducted to find out more detail about network and data features considered as prediction features. The analysis results are summarized to show the prediction comparison based on prediction features and prediction approaches, as shown in Table 4. Here, the feature is a characteristic of an object to be observed. The data features are an object of study related to the observed data, and the network types are related to the observed network. Later, prediction approaches denoted a method or a technique used to predict the links.

Furthermore, an in-depth analysis is conducted to compare prediction features and prediction approaches based on credible sources, IEEExplore, ACM digital library, ScienceDirect, Scopus, and Springer. A total of 836 articles have been published in IEEExplore, 1606 articles in ACM digital library, 1043 articles in ScienceDirect, 3000 articles in Scopus, and 2486 in Springer related to link prediction. Later, 60 articles were selected that fit the prediction link and were published from 2017 to 2020.

Table 4 shows that the previous studies mostly focused on structural features and similarity-based with 52.5% and 50.8%, respectively. The previous studies have conducted studies on various network types, as link prediction is highly dependent on the network scope. In turn, link prediction research is an area of research that is still new and has many open challenges. Adapt to many network types is an open challenge in the development of a link prediction solution. Therefore, it can be concluded that the similarity-based approaches are the most proposed and enhanced to solve the problem of link prediction, in addition to the learning-based approaches. However, in some instances, the similarity-based and learning-based approaches can be combined to achieve better performance. Moreover, a new classification was found for the prediction approach based on similarity, namely community similarity. This community similarity is termed "community similarity-based" in the taxonomy of link prediction.

III. PROPOSED LINK PREDICTION TAXONOMY

Taxonomy is a systematic work associated with the classification. The proposed taxonomy consists of three components that influence link prediction: the prediction approaches, prediction features, and prediction measurements, as shown in Fig. 1. The taxonomy is considered according to previous related studies inline with the most recent state-of-the-art research and the data from real-world network dataset providers.

New community-based methods are included in the similarity-based approach following the latest research trends on prediction links that utilize community information as predictive parameters [20], [87], [102], [113]. The similarity-based approaches include the local similarity-based, global similarity-based, quasi-local similarity-based, and community similarity-based methods, as shown in Fig. 1. In addition, the prediction features are classified into data features, data representation, network features, and network categories. In turn, the measurement methods include the AUC score,

TABLE 3. Comparison of the components.

Authors, Year	Ref	Dataset/s	Measurement/s
Liben-Nowell & Kleinberg, 2007	[4]	astro-ph, cond-mat, gr-qc, hep-ph, hep-th	Prediction performance
Linyuan & Zhou, 2011	[29]	PPI, NS, Grid, PB, INT, USAir.	AUC values, precision
P. Wang *et al.*, 2015	[31]	DBLP, Arxiv, NIPS 1-17, Enron email, Patents citation, Facebook, Twitter, Foursquare, MovieLens, Book- Crossing Book, Wikipedia, Epinions, Slashdot, Plurk.	NN
Martínez *et al.*, 2016	[28]	Yeast, C.elegan, INF, US politics, hamsterster, NORTAD, NSC.	AUC values, precision
Haghani & Keyvanpour, 2017	[32]	NN	Quantitative evaluation metrics Fixed-threshold metrics: Accuracy, Precision, Recall, F1-score, NDCG, Mean Rank (MR), Hit@n The Threshold curves: Receive Operation characteristic (ROC), Precision-Recall (PR) Qualitative evaluation metrics Cost, Scalability, Generalization, Exploring evolutionary patterns, Knowledge Representation
Pandey *et al.*, 2019	[33]	Facebook data	Area under the ROC curve (AUC), Precision, Self-predictability

accuracy, precision, Area Under the Precision-Recall Curve (AUPRC), recall, and the F1-score.

IV. PREDICTION APPROACHES

Based on the related work section, prediction approaches are classified into the similarity-based and learning-based approaches. The similarity-based approach is a technique that focuses on the topological information of an observed network. Meanwhile, the learning-based approach is a technique that focuses on the non-topological information of an observed network. The prediction is determined based on features that provide previous link prediction metrics, internal attributes, and external information.

A. SIMILARITY-BASED APPROACHES

The similarity-based approaches are the simplest one among the link prediction approaches, as it is given as score ranking for each unobserved pair of nodes. The similarity-based approaches can be applied successfully for some networks, but can also fail for some other networks [29]. The proximity of similarities over unconnected node pairs is the basis of this approach [28], [31]. The similarity is calculated based on the determination of potential pairing node candidate that is defined as E(u,v), where u and v of unconnected node pairs are calculated as the index similarity score by using the selected similarity-based methods. The index scores are sorted from the highest to lowest ones. The highest score corresponds to the pair of nodes with the highest possibility of generating new links or missing links. The taxonomy of similarity-based approaches is explained in more detail in Fig. 2.

1) LOCAL SIMILARITY-BASED METHODS

The local similarity-based methods use the local node search of local topological information to obtain a potential CNP [28], [33]. The local similarity-based methods use a simple calculation formula so that it can complete index calculations with the high speed. However, the local similarity-based methods cannot calculate the node candidate pairs' proximity scores when the node candidate pairs are the neighboring node neighbors or is situated on the distance minimum of two nodes from neighbors' nodes.. This condition is the limitation of the local similarity-based methods.

Rafiee *et al.* [59] proposed a similarity-based prediction method based on common neighbors degree penalization that generates similarity scores based on the similarity of common neighbors' topology characteristics and the average network cluster coefficient. Later, Liu *et al.* [111] proposed an extended resource allocation index (ERA), which adds longer paths to the RA index. The ERA index measures similarity based on a parameter adjusting the number of resources transferred by longer paths in different networks and the number of resources exchanged by common neighbors and non-common neighbors between two endpoints [111].

2) GLOBAL SIMILARITY-BASED METHODS

The global similarity-based methods use the global node search of all path network information to obtain a potential CNP [28], [33]. It should be noted that the computation complexity is the main limitation of the global similarity-based methods. The path-based methods are only based on the path length and ignore interactions between paths [93]. Later, Yao *et al.* [93] proposed the resources-from-short-paths index, which implies that interaction between paths is the process of receiving resources from the neighbors of an intermediary node. Then, more and more intermediary nodes will make many contributions to this interaction path [93].

3) QUASI-LOCAL SIMILARITY-BASED METHODS

The quasi-local similarity-based methods are a combination of local and global similarity-based ones. The quasi-local similarity-based methods are intended to reduce the

TABLE 4. Comparison between prediction features and prediction approaches.

Authors, Year	Ref	Data features	Network types	Prediction approaches
Cai *et al.*, 2019	[58]	Community profiles	Dynamic networks, Network as times-series	Deep Learning (Recurrent neural network)
Rafiee *et al.*, 2020	[59]	Structural features	Homogeneous, undirected networks	Similarity metric
Curado, 2020	[60]	Structural features	Undirected networks	Commute distances, vertex similarity
M. Zhou *et al.*, 2020	[61]	Structural features	Unweighted and undirected networks	Deep learning (DeepWalk)
J. Liu *et al.*, 2020	[62]	Structural features	Heterogeneous networks	Nodes similarity
S. Li *et al.*, 2020	[20]	Community profiles	Sparse and multigraph networks	Community similarity
Aslan & Kaya, 2020	[63]	Time features	Bipartite networks	Similarity
Chen *et al.*, 2020	[64]	Topological structure	Sparse and undirected networks	Matrix factorization
K. Li *et al.*, 2020	[65]	Structural features	Unweighted and undirected networks	Logistic regression and Xgboost algorithm
Zhongying Zhao et al., 2020	[66]	Structural features	Heterogeneous networks	Matrix factorization
Singh *et al.*, 2020	[67]	Community profiles	Undirected and directed networks	Community similarity
Lim *et al.*, 2019	[68]	Structural features	Time-evolving network	Deep reinforcement learning
Chen *et al.*, 2019	[69]	Topological structure	Weighted networks	Weight similarity
Yuan *et al.*, 2019	[70]	Structural features	Simple networks	Supervised learning
Xu *et al.*, 2019	[71]	Structural features	Dynamic and weighted networks	Similarity aggregation
J. Wu *et al.*, 2019	[72]	Structural features, matrix based models	Undirected, unweighted and weighted networks	Influential node
M. Wu *et al.*, 2019	[73]	Structural features	Static and homogeneous networks	Similarity metric
J. Wang *et al.*, 2019	[74]	Activity/tag features	Homogeneous, directed and undirected networks	Nodes similarity
Zhenbao Wang *et al.*, 2019	[75]	User degree	Undirected simple networks	Nodes similarity
Shao *et al.*, 2019	[76]	Structural Features	Undirected and directed networks	Machine Learning, Node2vec
Sharma *et al.*, 2019	[77]	Ego networks, circles and profile features	Undirected and unweighted networks	Deep belief network
Pech *et al.*, 2019	[78]	Network structural, matrix based models	Simple, undirected, unweighted, weighted, and directed networks	Hybrid (score matrix or similarity matrix)
Najari *et al.*, 2019	[79]	Structural features	Multiplex and heterogeneous networks	Interlayer similarity
Kumar *et al.*, 2019	[80]	Structural features	Undirected and directed networks	Path similarity
Mallek *et al.*, 2019	[81]	Structural features and social information	Unweighted and undirected networks	Similarity and probability
Jeong and Kim, 2019	[82]	Structural features	Heterogeneous networks	Networks information
H. Gao *et al.*, 2019	[83]	Structural features, time features	Undirected and dynamics	Dynamical similarity
Chi *et al.*, 2019	[84]	Time features	Simple, dynamic networks	Probability
Bütün & Kaya, 2019	[85]	Structural features	Directed and heterogeneous networks	Nodes similarity
Ai *et al.*, 2019	[86]	Tag-based weight, item attribute based weight, rating-information weight based	Unweighted and undirected networks	Hybrid (probability and ratings)
Bastami *et al.*, 2019	[87]	Local, global, community information	Undirected and dense networks	Community similarity
Gundala & Spezzano, 2019	[88]	Structural features, nodes features	Simple/static network, homogeneous, undirected and unweighted networks	Nodes similarity
Mahmoudi *et al.*, 2019	[89]	Time features, user profiles features, community features	Dynamic networks	Nodes similarity, community similarity
Aslan & Kaya, 2018	[18]	Structural features and Tag features	Heterogeneous and bipartite networks	Nodes similarity
X. Zhang *et al.*, 2018	[90]	Structural features	Undirected and unweighted networks	Probability
Cai *et al.*, 2019	[58]	Time features	Dynamic networks	Recurrent neural network
Yang *et al.*, 2018	[91]	Structural features	Unweighted and undirected networks	Nodes and path similarity
S. Wu *et al.*, 2018	[92]	Structural features	Multi-relational networks	Nodes similarity
Y. Yao *et al.*, 2018	[93]	Structural features	Unweighted and undirected networks	Quasi-local path
Chiu & Zhan, 2018	[94]	Time features	Dynamic networks	Deep learning (Deep neural network)
Zhiqiang Wang *et al.*, 2018	[95]	Structural features	Directed and undirected networks	Probability matrix factorization
J. Wu, 2018	[96]	Structural features	Unweighted and undirected networks	Probability
T. Li *et al.*, 2018	[97]	Structural features	Dynamic networks	Deep learning (Recurrent Neural Network)
Bütün *et al.*, 2018	[98]	Structural features	Directed, weighted and temporal networks	Classification models
Moradabadi & Meybodi, 2018	[99]	Structural features	Weighted networks	Learning automata

TABLE 4. ***(Continued.)*** **Comparison between prediction features and prediction approaches.**

Shang *et al.*, 2017	[100]	Network structure formation	Directed networks	Nodes similarity
Dai *et al.*, 2017	[101]	Structural features	Multi-relational networks	Matrix factorization
Mohan *et al.*, 2017	[102]	Community profiles	Undirected and unweighted networks	Community similarity
M. Gao *et al.*, 2017	[103]	Structural features	Bipartite networks	Similarity-based
W. Wang *et al.*, 2017	[104]	Structural features	Unweighted and undirected networks	Matrix factorization
X. Ma *et al.*, 2017	[105]	Time features	Temporal networks	Matrix factorization
C. Ma *et al.*, 2017	[106]	Structural features	Undirected networks	Adaptive fusion
Yin *et al.*, 2017	[57]	Structural features	Unweighted and weighted networks	Dempster–Shafer theory
Jichao Li *et al.*, 2017	[107]	Structural features	Heterogeneous networks	Probability
Z. Zhang *et al.*, 2017	[108]	Time features	Dynamic networks	Nodes similarity
Y. Li *et al.*, 2017	[109]	User profiles features	Undirected and unweighted networks	Latent variables
Moradabadi & Meybodi, 2017	[110]	Time features	Network as times-series	Learning automata
S. Liu *et al.*, 2017	[111]	Structural features	Unweighted, undirected, dynamic networks	Nodes similarity
Pei *et al.*, 2017	[112]	Structural features	Unweighted and undirected networks	Nodes similarity
J. Wu *et al.*, 2017	[113]	Community features	Unweighted and undirected networks	Community similarity

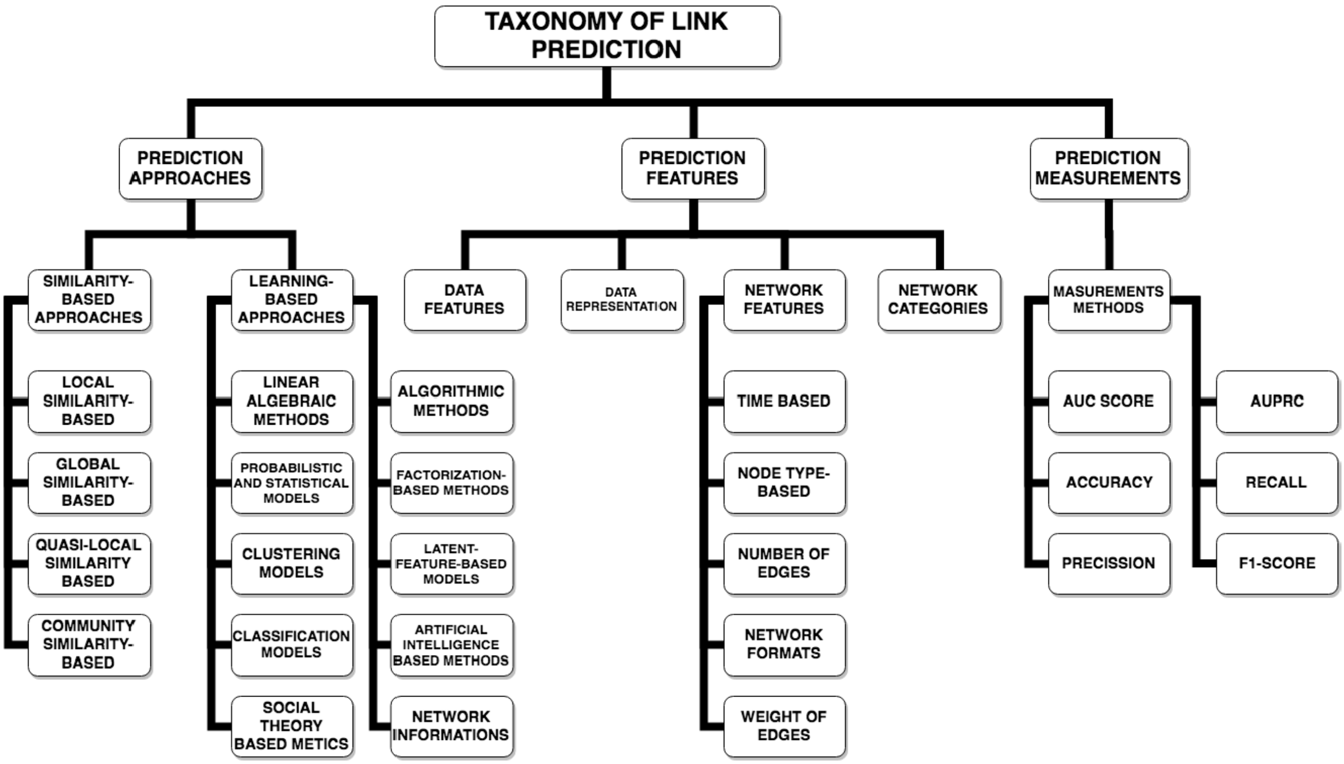

FIGURE 1. **Taxonomy of link prediction.**

computation complexity of the global similarity-based one while expanding topological information usage of the local similarity-based methods [28], [33]. The quasi-local similarity-based methods have a potential open issue related to performance improvement in terms of precision and accuracy. The problem with link prediction is to generate estimates of potential opportunities for new links in the network. This link prediction is related to large amounts of network information, and the problem is the ability to search in the case of the massive data [60]. To overcome this problem, Curado [60] proposed a link prediction approach based on the input graph precondition called Return Random Walk (RRW). RRW is used to minimize the likelihood of a random walk starting and ending at a node that crosses links between classes [60].

4) COMMUNITY SIMILARITY-BASED METHODS

The community similarity-based methods assume that a node will form a new link not always based on the presence of common neighbors, but because the node enters into a new community or influences the profile similarity of the nodes in a community. The idea of community-based similarity

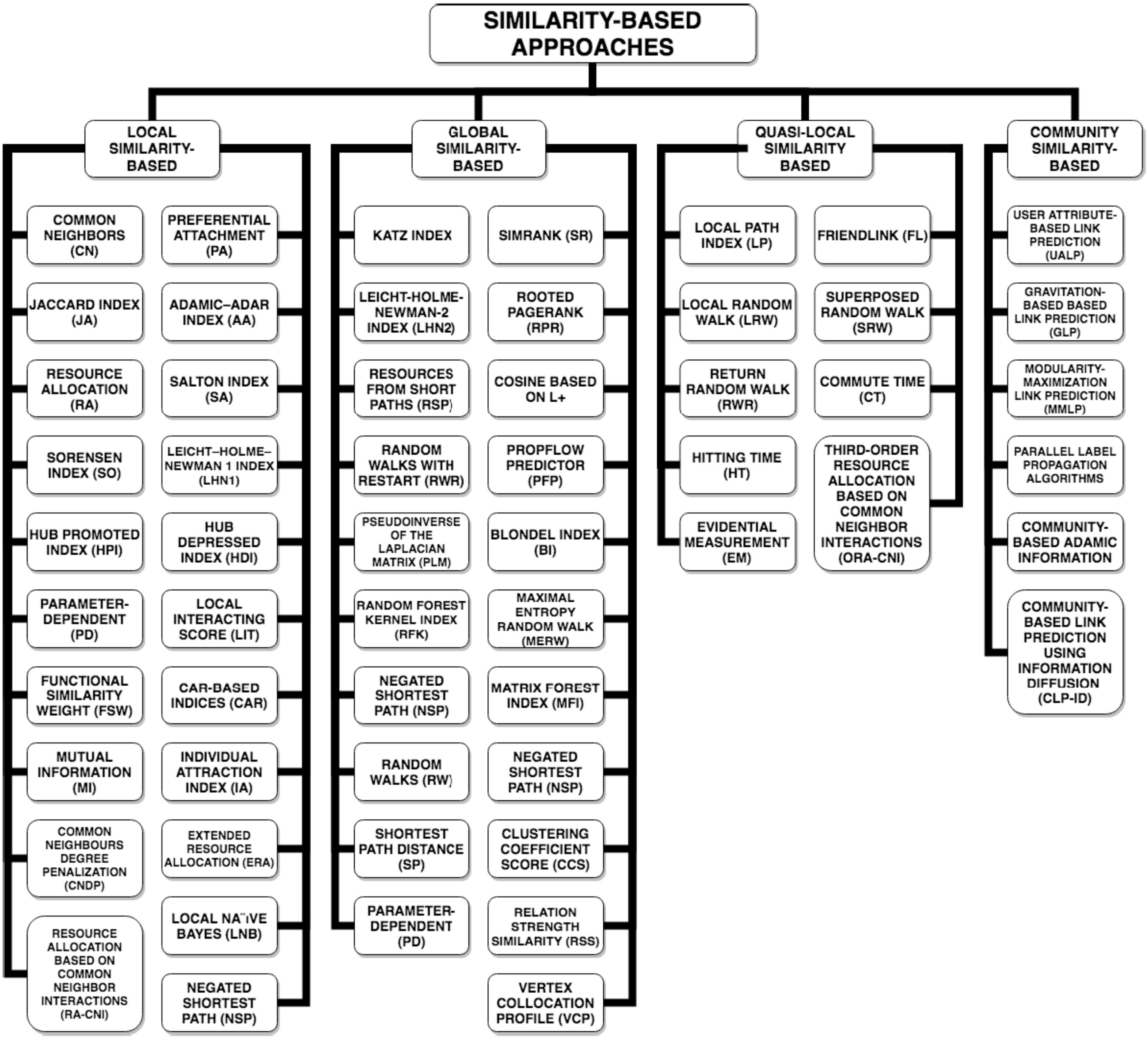

FIGURE 2. Detailed taxonomy of similarity-based approaches in link prediction.

methods is based on the fundamental theory for the group-level analysis of the graph theory, such as cliques, hierarchical clustering, coefficient clustering, community detection, and others. Several researchers have studied community-based similarities. Singh *et al.* [67] described information diffusion and the community structure that divided the network into clusters and denoted the algorithms as CLP-ID. Mahmoudi *et al.* [89] outlined user community changes referred to as User Attribute-based Link Prediction (UALP). Bastami *et al.* [87] proposed the gravitation-based link prediction approach with the integration of node features, community information, and graph properties.

The friend recommendations usually do not consider existing friendships in different SN circles [20]. Later, Li *et al.* [20] proposed an Intelligent Attention Allocation Link Prediction algorithm (IAALPa) that can predict the potential friendship from a different network circle. Furthermore, Mohan *et al.* [102] proposed a measure of hybrid similarity and a scalable method for predicting links based on the community structure on large-scale networks by detects communities based on parallel label propagation algorithms and predicts new links based on community-based Adamic information. Later, Wu *et al.* [113] proposed a balanced Modularity-Maximisation Link Prediction (MMLP) model to solve capturing problem the correlation between link formulation and community evolution by integrating the formulation of two link types into a partitioned network generative model.

B. LEARNING-BASED APPROACHES

The learning-based approaches are based not only on topological information like node and topology from an observed network but also on other network and data features. Classification methods serve as the basis of the learning-based approaches. Adopting probabilistic, statistical, mathematical models, and machine learning models as predictors, as shown in Fig. 3, allows expanding the learning-based approaches.

1) SOCIAL THEORY

Social theory can be used to improve performance by using information about social interaction. These social theories include community, triadic closure, strong and weak ties, homophily, and structural balance [31].

2) FACTORIZATION-BASED METHODS

Chen *et al.* [64] proposed a robust non-negative matrix factorization by combining manifold regularization and sparse learning (MS-RNMF) to address missing links, spurious links, and random noise. Later, Zhiqiang Wang *et al.* [95] proposed an method based on combining adjacent matrix and key topology metrics as the integrated probability symmetric and asymmetric metric factorization framework to obtain a low ranking approach from an adjacent matrix of a network. However, this ranking compared the network with itself, so the matrix's information was not sufficient [95]. Furthermore, Ma *et al.* [105] proposed a new NMF-based algorithm as the equivalence between Eigen decomposition and non-negative matrix factorization (NMF) to overcome the relationship within the matrix-based decomposition algorithm. Zhao *et al.* [66] also proposed a method based on heterogeneous network embedding called HetNERec. The HetNERec constructs a heterogenous co-occurrence network and proposes network embedding for vector representation and extending matrix representation to learn the representation.

3) PROBABILISTIC AND STATISTICAL MODELS

The performance of existing link prediction methods is not always acceptable in all cases, as each network has its unique underlying structural features [106]. However, estimating different features' contribution is a challenging question, as these features are very different. Therefore, an adaptive fusion model is proposed based on the combination of the multi structural features and the use the 'learnt' logistic function predict the likelihood of missing links [106]. Furthermore, Wu [96] proposed the Tree Augmented Naive Bayes (TAN) probabilistic model for link prediction that can mitigate the assumption of strong and independent Local Naive Bayes (LNB).

4) LATENT-FEATURE-BASED MODELS

Zhang *et al.* [108] proposed the two efficient and dynamic increment algorithms based on the improved latent space and resource allocation to predict links dynamically according to the SN structure updates referred to dynamic link prediction algorithms based on improved latent space (DLP-ILS) and dynamic link prediction algorithms based on improved resource allocation (DLP- IRA). The advantage of DLP-IRA and DLP-ILS is that they only need to recalculate the graph when being updated partially. Conversely, the disadvantage is associated with the node adjacency relationship that does not have the common neighbors processed serially, not parallel [108]. Furthermore, Li *et al.* [109] proposed the utility-based link prediction method based on considering that individual preferences are the main reason behind the decision to form links.

5) ARTIFICIAL INTELLIGENCE-BASED METHODS

The concepts of the multilevel deep network-based learning model and user consumption preferences have been introduced by Sharma *et al.* [77] to enhance the prediction accuracy. Later, Pech *et al.* [78] proposed a new link prediction algorithm based on linear optimization to achieve low computational complexity seeking to handle thousands of nodes, even though real SN scale up to millions or billions of nodes [78]. Furthermore, Yin *et al.* [57] proposed a new link prediction method called EM based on the Dempster–Shafer theory and a new method to measure link predictability via local information and Shannon entropy. At the same time, Moradabadi and Meybodi [99], [110] proposed a link prediction method for stochastic SN [99] and the new time series link prediction method based on learning automata applicable to dynamic activities and network changes over time. They concluded that graph modeling is not suitable for analyzing SN [110].

6) NODE ATTRIBUTES (NA) METHODS

Mahmoudi *et al.* [89] proposed a link prediction method based on time and user attributes obtained in real-time called UALP to consider attributes, such as user weight, interaction density and geographic distance. In turn, Zhang *et al.* [90] proposed a model and index using an intermediary process to improve the performance and to measure the positive or negative effects arising from the node attributions or network feature, which was named as the intermediary probability model (IMP).

7) CORRELATION INFORMATION (CI) METHODS

Predicting target interactions using multiple interactions in link prediction aims to exploit handy auxiliary interactions [92]. Moreover, a new way is also required to identify online social networks (OSN) behavioral changes over time [114]. Wang *et al.* [75] proposed a weighted endpoint influence (WSI) with a degree highlighted and H-index defined synthetically to put more emphasis on path information between nodes that are not connected. Besides that, Gundala and Spezzano [88] proposed a solution based on a supervised learning method to predict the emergence of new links by benefitting from node pairs interactions over time. Previously, Pei *et al.* [112] proposed the neighbor set information allocation index based on a set of neighbors obtained from

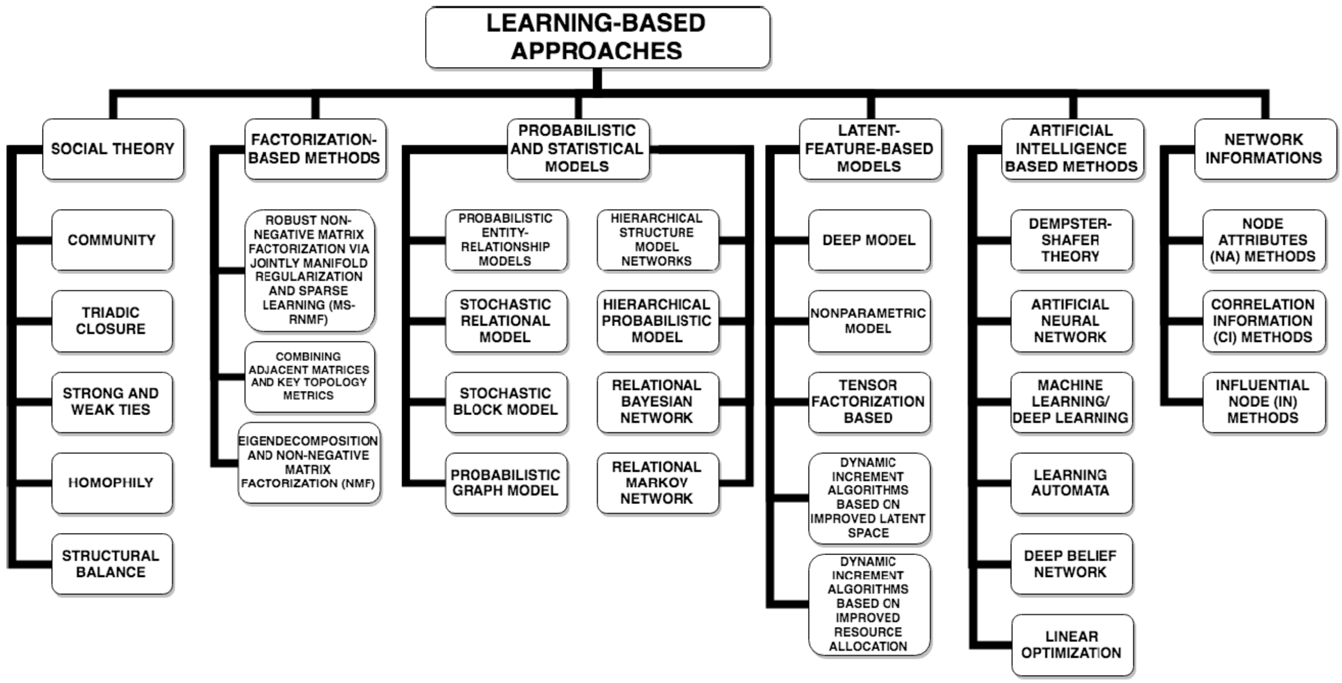

FIGURE 3. Detailed taxonomy of learning-based approaches in link prediction.

the process of virtual information allocation to quantify the possible connection of events corresponding to the two nodes by measuring self-information.

8) INFLUENTIAL NODE (IN) METHODS

Nodes are the key features in OSN, and the behavior of nodes is strongly related to the behavior of OSN [115]. Wu *et al.* [72] proposed a novel framework called Influential Node Identification Link Prediction (INILP) to integrate the identification technique for the famous and influential node. Furthermore, Yang *et al.* [91] proposed a significant influence (SI) index for link prediction, which promotes the next relationships and endpoint similarity based on the significant influence of more robust and weaker relationships. In turn, Dai *et al.* [101] proposed the multi-relational networks based on relational similarity and named LPMR to evaluate node belief-by-belief propagation and then to construct the belief for each link type. Then, LPMR uses the similarity between the belief vectors as the influence between different relational networks [101].

9) NETWORK RECONSTRUCTION

Network construction has been used by Ai *et al.* [86] to select a neighbor based on the distance between two objects, and then, to assign the position to an object in a spatial distribution topology to address insufficient rating information or the cold start problem. Previously, Wang *et al.* [104] proposed a kernel framework and reconstructed the network using a different kernel that can obtain global and local network information through kernel mapping. Furthermore, Wu *et al.* [73] proposed a new serial ensemble strategy by using network reconstruction of nine local indices aggregated with the Ordered Weighted Averaging (OWA) operator.

V. PREDICTION FEATURES

Prediction features are a component of the data network that influences link prediction approaches. The taxonomy of link prediction features is explained in more detail in Fig. 4, outlining the feature influencing members, such as data features, data representation, network types, and network categories.

A. DATA FEATURES

Data features are classified into structural features, time features, activity/tag features, user profiles features, group/community profiles pictures, and ego network features, as shown in Fig. 4. Structural features use nodes and edges features to predict the future or missing links and are mostly used in local similarity-based methods. Furthermore, time features, activity/tag features, and user profiles features are the data features that use attribute features of nodes and or edges and depend on the data's variance. Then, group/community profiles use the collection of attribute nodes and edges and cluster them in a group/community. Lastly, ego network features are the features that are influenced by a focal actor or ego as a center in a network, and therefore, network changes depend on certain central actors and involve relationships that are extended from specific individuals.

Similarity influences many people in the social relationship, but rich information available is missing for link prediction [70]. To address this problem, link prediction based on the structural information of signed SN is proposed by comparing user similarity [70]. Later, tag information also has an essential role in creating new links based on the assumption that nodes with similar tags will potentially connect [74]. Furthermore, Wang *et al.* [74] proposed a novel link prediction algorithm based on a tag system homogeneity called tag-aware link prediction to improve the prediction accuracy.

B. DATA REPRESENTATION

There are only two types of data representation, as shown in Fig. 4, i.e., matrix-based and graph-based models. Matrix-based models imply that to predict links in network datasets, it is necessary to convert the network dataset into a matrix, and then calculate link prediction using the similarity-based or learning-based methods. Similarly, graph-based models require a graph method or a specific programming library to calculate link prediction. However, graph-based models are the most commonly used ones in the methods proposed in the previous studies.

The nature of the data influences the SN's structure, so it is sensitive to wrong observations, which can result in distortion [81]. The two main problems are considering all nodes and edges and the risk of the possibility of adding or removing edges, and the problem of missing edges. The graph-based model for social, developed by Mallek *et al.* [81], implements two algorithms with and without prior knowledge on the existing links to handle the edge level uncertainty.

C. NETWORK TYPES

The network types are classified into five viewpoints: time-based, node type-based, number of edges, network formats, and edge type-based, as shown in Fig. 4.

1) TIME-BASED

A time-based network features are classified into the simple/static network, network as times series, and the temporal network. The simple/static network is a network that consists of V and E, where V is a node, and E is an edge that connects two nodes. The network as a times series is a series of multivariate time points related to a network structure [116]. Furthermore, the temporal network is a graph with specific time settings that are modeled mathematically and computationally [117]. The links prediction algorithms ignore the evolution process and show low accuracy and scalability to large-scale networks [71]. In turn, Xu *et al.* [71] proposed a new distributed temporal link prediction algorithm based on label propagation (DTLPLP) that is adjusted according to dynamical interaction among nodes.

Shang *et al.* [118] proposed novel unweighted and weighted link prediction methods in evolving networks and found those common nodes have a higher probability of creating a connection of node pairs for evolving nodes in unweighted networks and reduction of the human factor in weighted networks also influence weight reduction and generally better for static networks. Furthermore, in order to analyze link direction, Shang *et al.* [100] also proposed a phase dynamic algorithm to demonstrate the different roles between bi-directional links and one-directional links in the formation of network structure. The finding of Shang *et al.* [100] shows that a bi-directional link has higher probabilities with another bi-directional link that has common neighbors and vice versa. Later, Shang *et al.* [10] introduced time as a parameter to calculate the future link and modify it to extract information of the existing direct link in evolving networks. The evolving link is stated by time point for past edge set and future edge set. The experiment results show that network structure gives contribution more significant than weight in link prediction.

2) NODE TYPE-BASED

The network features include homogeneous and heterogeneous networks. The homogeneous network is a network that contains nodes and edges of a similar type, and the heterogeneous network is a network that contains those of different types. In heterogeneous networks, link prediction is used to predict missing links with the help of interconnected target and assistive networks. Shang *et al.* [11] proposed an algorithm based on network heterogeneity as an essential structure for information propagation. The proposed algorithm, called Heterogeneity Index (HEI), considers the node's degree and a free heterogeneity exponent. Moreover, Shang *et al.* [11] also proposed a homogeneity index (HOI) and a combination of HEI and HOI algorithms called the Heterogeneity Adaptation Index (HAI). Furthermore, Jeong and Kim [82] proposed a measure of the correlation between the path and link types to enhance link prediction in heterogeneous information networks (HINs) to measure the number of paths, homogeneous neighbors, and adjacent colleagues.

Liu *et al.* [62] have introduced the Collaborative Linear Manifold Learning (CLML) algorithm. It is useful to optimize node similarities' consistency [62]. Furthermore, Li *et al.* [107] proposed a heterogeneous combat network link prediction based on the meta-path approach (HCNMP) to solve the simultaneous prediction problem of multiple link types for a heterogeneous combat network (HCN). HCNMP considers not only related information structures but also semantic information. Later, Zhao *et al.* [66] proposed a recommendation method based on heterogeneous network embedding called HetNERec. HetNERec constructs heterogeneous co-occurrence to identify the latent representation of nodes. The proposed method combines multiple node representation into a single representation and integrates with matrix factorization to generate the recommendation. However, the HetNERec is designed for static scenarios and not consider changing in the dynamic temporal.

3) NUMBER OF EDGES

There are two types of networks based on the number of edges, sparse, and dense networks. A sparse network is a network that the number of edges is close to the

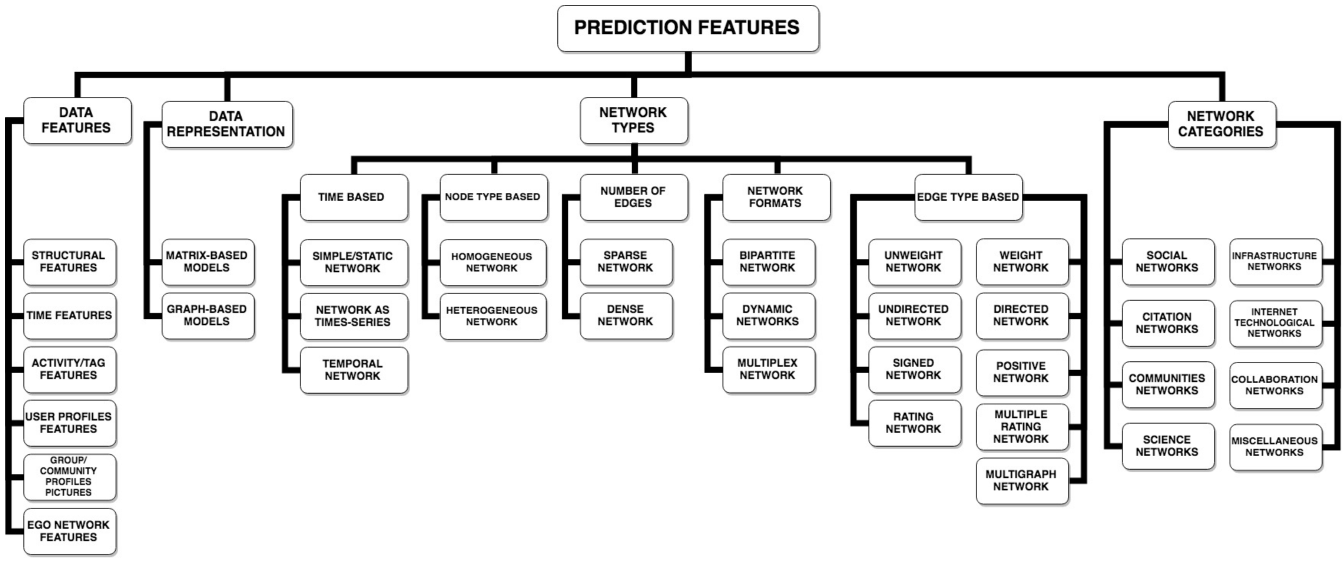

FIGURE 4. **Detailed taxonomy of prediction features in link prediction.**

minimum or zero, and a dense network is a network that the number of edges is close to the maximum number or the square of the number of nodes. Therefore, the link weight and topology structure are the two essential features in link prediction [69]. To integrate the information topology and link weights, Chen *et al.* [69] proposed a model called graph regularization weighted non-negative matrix factorization that uses the weighted cosine similarity method to calculate the weighted similarity between nodes.

4) NETWORK FORMATS

Network formats are classified into the bipartite networks, dynamic networks, and multiplex networks. The bipartite networks consist of two types of nodes, and all edges can only be connected to different node types, but not to nodes of a similar type. Then, the dynamic network is a network in which edges may appear and disappear over time and have a temporal character. Lastly, the multilayer network is a network with similar nodes across layers [79].

There are many weak relationships between node pairs in the bipartite networks. It is due to the classical similarity algorithm that only measures the current network without considering network evolution over time [63]. Furthermore, Aslan and Kaya [63] proposed a projection model called the strengthened projection model and the proximity measure algorithm that considers the network evolution. Aslan and Kaya [18] also proposed a new link prediction algorithm based on strengthening weighted projection to predict the association between authors and topics and to prevent loss of information in the case if a bipartite network is changed into a unimodal network. In turn, multiplex or multilayer networks have a problem to predict new links at the one layer by using structural information from another layer [79]. Later, Najari *et al.* [79] proposed a novel framework by combining inter-layer similarity and proximity-based features, with further extraction of the predicted link from the considered layer.

Li *et al.* [65] proposed the ensemble model-based link prediction algorithm (EMLP) based on the logistic regression algorithm and an Xgboost algorithm based on the learnt models corresponding to the four characteristics of similarity indices in a complex network. Furthermore, Kumar *et al.* [80] proposed link prediction based on the Significance of the Higher Order Path Index (SHOPI) that is used to examine and to enforce the information leakage penalizing common neighbors.

Liu *et al.* [119] proposed a top-n-stability method in the bipartite network for personalized recommendation and deleting unstable similarity and to reduce the recommendation of false information to users. By considering stability similarity also could improve the stability recommendation. Furthermore, Gao *et al.* [103] proposed a projected graph in the bipartite network to map into a unipartite network aiming to define the concept of a CNP and potential link prediction (PLP).

5) EDGE TYPE-BASED

According to edge type-based, networks are classified into unweighted, weighted, undirected, directed, multigraph, positive, signed, rating, and multiple rating. The unweighted network is a network that has only one edge for two paired nodes. There is no different edge connection between u to v and v to u in the undirected network and vice versa for the directed network. Furthermore, the multigraph is a network with many edges in which edges are unweighted, and several edges or more connects the node pair. The positive network is an edge on a network with a positive marker, and only one edge is allowed for each pair of nodes. Then, the signed network is a network that is given positive or negative weights for

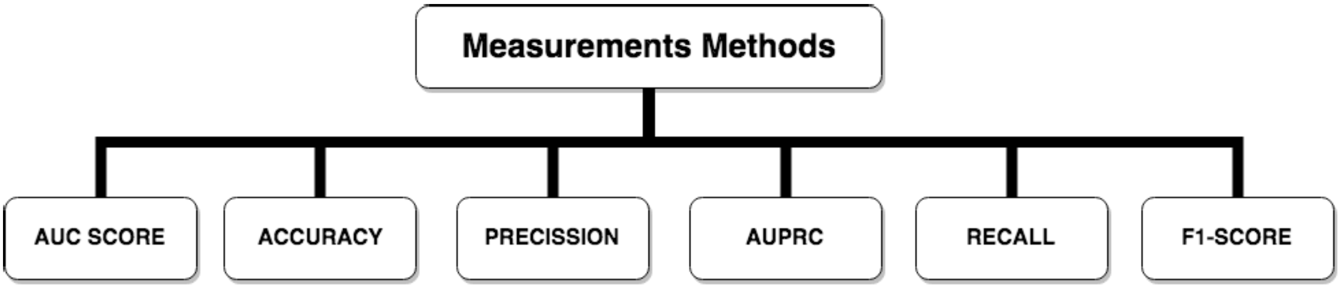

FIGURE 5. Detailed taxonomy of prediction measurements in link prediction.

TABLE 5. Comparison benchmark algorithms and measurement methods in link prediction.

Authors, Year	Ref	Benchmark algorithms													Measurement methods		
		CN	JA	AA	PA	RA	SA	SO	LLHN	HPI	HDI	Katz	LP	Others	AUC	Precision	Others
J. Liu *et al.*, 2020	[20]	-	-	-	V	V	V	V	V	V	V	-	-	B1	V	-	-
Aslan & Kaya, 2020	[63]	V	V	V	V	-	-	V	-	V	V	-	-	B2, B3	-	V	M1
Yuan *et al.*, 2019	[70]	V	-	V	-	V	-	-	-	-	-	-	-	B4	-	-	M2-M3
Xu *et al.*, 2019	[71]	V	-	-	-	V	-	-	-	-	-	-	-	-	V	V	-
M. Wu *et al.*, 2019	[73]	V	V	V	-	V	V	V	V	V	V	-	-	-	V	V	M1
J. Wang *et al.*, 2019	[74]	-	-	-	V	V	-	-	-	-	-	-	-	B5-B8	V	-	-
Zhenbao Wang *et al.*, 2019	[75]	V	-	V	-	V	-	-	-	-	-	-	V	B9-B10	-	V	-
Pech *et al.*, 2019	[78]	V	-	V	-	V	-	-	-	-	-	V	V	B11-B12	V	V	-
Najari *et al.*, 2019	[79]	-	V	V	V	V	-	-	-	-	-	-	V	-	V	-	-
Mallek *et al.*, 2019	[81]	V	V	V	V	V	V	V	-	-	-	-	-	B13-B14	V	V	-
H. Gao *et al.*, 2019	[83]	V	-	V	-	V	-	-	-	-	-	V	V	B15-B16	V	V	-
Chi *et al.*, 2019	[84]	V	-	V	-	V	-	-	-	-	-	-	-	B17-B19	V	V	-
Bütün & Kaya, 2019	[85]	V	V	V	-	V	V	V	V	V	V	-	-	-	V	-	-
Gundala & Spezzano, 2019	[88]	-	V	V	V	-	-	-	-	-	-	-	-	B20-B21	-	-	M5-M6
Mahmoudi *et al.*, 2019	[89]	V	V	V	V	V	V	V	V	V	V	-	-	B22	V	-	-
SUM of V		11	7	12	7	13	5	6	4	5	5	2	4	-	11	8	M1=3

Note: B1 = Resource Allocation Average (RAA), B2 = Cosine Similarity (Cos), B3 = Local Community Paradigm (LCP), B4 = Triad Features (TF), B5 = Common tags (CT), B6 = Jaccard index between users' tags (JT), B7 = Cumulative similarity based on similarity oftags (CSST), B8 = Probability-based (Prob) predictor, B9 = Synthetical Influence Model (SI), B10 = Superposed Random Walk (SRW), B11 = Cannistraci resource allocation (CRA) index, B12 = structural perturbation method (SPM), B13 = Common Neighbors of Groups (CNG), B14 = Common Neighbors Within and Outside of Common Groups (WOCG), B15 = Path Entropy index (PE), B16 = Local Random Walk (LRW), B17 = Knowledge Dissemination Link Prediction (KDLP) algorithm, B19 = Resources from Short Paths (RSP) index, B19 = Random Walk with Restart (RWR) algorithm, B20 = Hits, B21 = Node2Vec, B22 = Com-ST (CRCN), M1= AUPRC, M2 = Accuracy, M3 = F1-score, M4 = ACT, M4 = Time, M5 = AUROC, M6 = MAP

each of its edges and not zeros weights. The rating network is a network that weights a discrete rating, and only one edge is allowed for a pair of nodes. Furthermore, the multiple rating networks are network with a ranking value that has edges for rating, and multiple edges for each node pair are allowed.

A linear dynamical response called LDR was introduced by Gao *et al.* [83] to measure similarity among pairing nodes in dynamic networks. Furthermore, most of the similarity matrices or learning in link prediction failed to consider network changes over time and cannot be implemented into a dynamic network structure [84]. To address this problem, Chi *et al.* [84] proposed a new link prediction for dynamic networks based on the attraction force between nodes called DLPA.

Shang *et al.* [100] proposed a directional prediction algorithm for directed network nodes to analyze the rule in link prediction and to show that bidirectional and one-directional links have different roles; this algorithm was called the phase-dynamic algorithm. Node pairs connected by bidirectional links tend to connect to common neighbors more rather than in the case of unidirectional links [100]. Therefore, Bütün and Kaya [85] proposed a pattern-based supervised link prediction method to improve the triad closeness metric in directed complex networks. Furthermore, the measurement of prediction link directions by expanding neighbor-based steps as a direction-based pattern was introduced by Bütün *et al.* [98] to take into account the role of link directions in the directed network.

D. NETWORK CATEGORIES

Social and citation networks are the most used network categories. Besides these two categories of networks, there are also others, such as community, science, infrastructure, collaboration, and technological Internet networks, as shown in Fig. 4. Miscellaneous networks are networks that cannot be categorized. Each of these network categories has its own characteristics and the potential to be investigated. The network categories are also available in the three real-world network dataset providers available on http://snap.stanford.edu/data/index.html [120], http://networkrepository.com [121], and http://konect.uni-koblenz.de/networks [122].

VI. PREDICTION MEASUREMENTS

The last part of the link prediction taxonomy is prediction measurement methods that are commonly used to evaluate the proposed methods in link prediction, as shown in Fig. 5.

The AUC score of a receiver operating characteristic curve is a standard metric to show the quality of link prediction results [77]. Then, the accuracy is defined as the ratio of the predicted classification results and the correct classification results [70]. Precision is used to evaluate the algorithm performance [73] and to measure the ratio of the selected items relevant to their number [64], [69], [104], [111]. Furthermore, recall is a division between the number of true positive results and the number of positive results returned [70]. F1-score is the weighted average of the precision and recall [70], and the AUPRC score is a metric used to evaluate the quality of algorithms [63]. Moreover, comparison benchmark algorithms and measurement to identify the benchmark algorithms and measurement methods, as shown in Table 5. Based on a sampling of 15 previous studies shown that RA, AA, and CN are sequentially the most used benchmark algorithms related to the local similarity-based methods with percentages of 86.7%, 75%, and 73.3%, respectively. Furthermore, Katz and LP are the most used benchmark algorithms in global similarity-based and quasi-local similarity-based methods. Meanwhile, the AUC score is the most common measurement method to evaluate novel proposed methods with 93.3%. These two results can be used as a basis for determining benchmarking algorithms and measurement methods for evaluating the performance of a novel proposed methods.

VII. CONCLUSION

The three main contributions of this review are as follows:

1) It reviews the studies on link prediction and proposes a taxonomy of link prediction with three components.
2) An extensive examination of relevant link prediction research to validate areas that have been previously conducted by researchers.
3) An extensive comparison to identify and compare the benchmark algorithms and measurement methods.

The previous reviews focused on prediction approaches that are yet to explain the remaining components capable of influencing the link prediction research. In this review, we add different perspectives on prediction features and prediction measurements in addition to prediction approaches to differentiate perspectives and enhance previous reviews. This review also presented a link prediction review with a proposed taxonomy based on the three main components: prediction approaches, prediction features, and prediction measurements. The prediction approaches are classified into the two approaches, namely, the similarity-based and learning-based approaches. The similarity-based approaches allow calculating score ranking for each unobserved pair of nodes, including the local similarity-based, global similarity-based, quasi-local similarity-based, and community similarity-based methods. Furthermore, the learning-based approaches are not only represented by the topological information, such as nodes and topology from an observed network, but also consider other data and network features and include the social theory-based, probabilistic and statistical models, artificial intelligence-based methods, network information, and classification methods. The previous studies mostly focused on structural and none dominate features for other features in link prediction. All studies have their motivation and consideration in determining prediction feature. It shows that link prediction still appears as an open research problem and can be implemented in many domain works.

Furthermore, the AUC score is the most commonly used method for measuring the novel proposed methods. Therefore, future research needs to consider measuring novel proposed methods unless other conditions can be addressed.

Some of the most common challenges associated with link prediction are rapid growth, sparse, and network features. Many real-networks data, such as online social networks, are highly dynamic, with an evolving and sparse network. Avoiding networks with isolated nodes cause cold-start problems is a challenge that needs to be addressed in link prediction [92], [123]. Besides, exploiting features other than the structural features is also a challenge to improve prediction accuracies, such as tags [18], [74], [86] and time [63], [89], [94]. Therefore, these challenges are also in the direction of future research that needs to implement link prediction in other domains such as co-authorship networks [18], [124] and economic networks [20] instead of online social networks. The co-authorship networks are supported by scientific publications' growth, forming a network of relationships between authors.

REFERENCES

[1] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
[2] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 2, Jul. 2001, Art. no. 025102.
[3] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Social Netw.*, vol. 25, no. 3, pp. 211–230, Jul. 2003.
[4] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007.

[5] E. Ravasz, "Hierarchical organization of modularity in metabolic networks," *Science*, vol. 297, no. 5586, pp. 1551–1555, Aug. 2002.
[6] E. A. Leicht, P. Holme, and M. E. J. Newman, "Vertex similarity in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 73, no. 2, Feb. 2006, Art. no. 026120.
[7] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," *Eur. Phys. J. B*, vol. 71, no. 4, pp. 623–630, Oct. 2009.
[8] P. Jaccard, "Étude comparative de la distribution florale dans une portion des Alpes et des Jura," *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 547–579, 1901.
[9] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, 1983.
[10] K.-K. Shang, M. Small, X.-K. Xu, and W.-S. Yan, "The role of direct links for link prediction in evolving networks," *EPL (Europhysics Lett.)*, vol. 117, no. 2, p. 28002, Jan. 2017.
[11] K.-K. Shang, T.-C. Li, M. Small, D. Burton, and Y. Wang, "Link prediction for tree-like networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 29, no. 6, Jun. 2019, Art. no. 061103.
[12] L. Yao, L. Wang, L. Pan, and K. Yao, "Link prediction based on common-neighbors for dynamic social network," *Procedia Comput. Sci.*, vol. 83, pp. 82–89, Jan. 2016.
[13] P. Kefalas, P. Symeonidis, and Y. Manolopoulos, "A graph-based taxonomy of recommendation algorithms and systems in LBSNs," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 604–622, Mar. 2016.
[14] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.
[15] S. Huang, J. Zhang, L. Wang, and X.-S. Hua, "Social friend recommendation based on multiple network correlation," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 287–299, Feb. 2016.
[16] Y.-Y. Yeung, T. C.-Y. Liu, and P.-H. Ng, "A social network analysis of research collaboration in physics education," *Amer. J. Phys.*, vol. 73, no. 2, pp. 145–150, Feb. 2005.
[17] I. M. Nadhiroh, R. Hardiyati, M. Amelia, and T. Handayani, "Mathematics and statistics related studies in indonesia using co-authorship network analysis," *Int. J. Adv. Intell. Informat.*, vol. 4, no. 2, p. 142, Jul. 2018.
[18] S. Aslan and M. Kaya, "Topic recommendation for authors as a link prediction problem," *Future Gener. Comput. Syst.*, vol. 89, pp. 249–264, Dec. 2018.
[19] J. Li, L. Zhang, F. Meng, and F. Li, "Recommendation algorithm based on link prediction and domain knowledge in retail transactions," *Procedia Comput. Sci.*, vol. 31, pp. 875–881, 2014.
[20] S. Li, X. Song, H. Lu, L. Zeng, M. Shi, and F. Liu, "Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm," *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112839.
[21] A. Anil, D. Kumar, S. Sharma, R. Singha, R. Sarmah, N. Bhattacharya, and S. R. Singh, "Link prediction using social network analysis over heterogeneous terrorist network," in *Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom (SmartCity)*, Dec. 2015, pp. 267–272.
[22] F. Xie, Z. Chen, J. Shang, X. Feng, and J. Li, "A link prediction approach for item recommendation with complex number," *Knowl.-Based Syst.*, vol. 81, pp. 148–158, Jun. 2015.
[23] H. Zare, M. A. Nikooie Pour, and P. Moradi, "Enhanced recommender system using predictive network approach," *Phys. A, Stat. Mech. Appl.*, vol. 520, pp. 322–337, Apr. 2019.
[24] B. Kaya and E. Gündoğan, "Evaluating reliability of question-disease relations in online health forms: A link prediction approach," *Telematics Informat.*, vol. 35, no. 7, pp. 1799–1808, Oct. 2018.
[25] W. A. Ali, M. Mukhtar, and I. Mohamed, "Validating the factors influencing social commerce adoption in small and medium enterprise in Malaysia," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 1, pp. 440–447, 2019, doi: 10.1142/9781848163324_0012.
[26] R. A. Rahman, K. Omar, S. A. M. Noah, and M. S. N. M. Danuri, "A survey on mental health detection in online social network," *Int. J. Adv. Sci., Eng. Inf. Technol.*, vol. 8, nos. 4–2, p. 1431, Sep. 2018.
[27] M. Rohaninezhad, S. A. Mohd Noah, and S. M. Arif, "Defeasible policy language for online social networks," *Int. J. Adv. Sci., Eng. Inf. Technol.*, vol. 7, no. 3, p. 736, May 2017.
[28] V. Martínez, F. Berzal, and J.-C. Cubero, "A survey of link prediction in complex networks," *ACM Comput. Surv.*, vol. 49, no. 4, pp. 1–33, Feb. 2017.
[29] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Phys. A, Stat. Mech. Appl.*, vol. 390, no. 6, pp. 1150–1170, Mar. 2011.
[30] M. Al Hasan and M. J. Zaki, "A survey of link prediction in social networks," in *Social Network Data Analytics*, C. Aggarwal, Ed. Boston, MA, USA: Springer, 2011, pp. 243–275.
[31] P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: The state-of-the-art," *Sci. China Inf. Sci.*, vol. 58, no. 1, pp. 1–38, Jan. 2015.
[32] S. Haghani and M. R. Keyvanpour, "A systemic analysis of link prediction in social network," *Artif. Intell. Rev.*, vol. 52, no. 3, pp. 1961–1995, 2017.
[33] B. Pandey, P. K. Bhanodia, A. Khamparia, and D. K. Pandey, "A comprehensive survey of edge prediction in social networks: Techniques, parameters and challenges," *Expert Syst. Appl.*, vol. 124, pp. 164–181, Jun. 2019.
[34] T. Wang and G. Liao, "A review of link prediction in social networks," in *Proc. Int. Conf. Manage. e-Commerce e-Government*, Oct. 2014, pp. 147–150.
[35] T. J. Sørensen, "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," I Kommission hos E. Munksgaard, 1948.
[36] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.
[37] G. Jeh and J. Widom, "SimRank: A measure of structural-context similarity," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2002, pp. 538–543.
[38] H. Tong, C. Faloutsos, and J.-Y. Pan, "Fast random walk with restart and its applications," in *Proc. 6th Int. Conf. Data Mining (ICDM)*, Dec. 2006, pp. 613–622.
[39] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "Fast and accurate link prediction in social networking systems," *J. Syst. Softw.*, vol. 85, no. 9, pp. 2119–2132, Sep. 2012.
[40] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New perspectives and methods in link prediction," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2010, p. 243.
[41] L. Lü, C.-H. Jin, and T. Zhou, "Similarity index based on local paths for link prediction of complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 4, Oct. 2009, Art. no. 046122.
[42] W. Liu and L. Lü, "Link prediction based on local random walk," *EPL (Europhysics Lett.)*, vol. 89, no. 5, p. 58007, Mar. 2010.
[43] P. Chebotarev and E. Shamis, "The matrix-forest theorem and measuring relations in small social groups," *Autom. Remote Control.*, vol. 58, pp. 1505–1514, Feb. 2006.
[44] Y.-X. Zhu, L. Lü, Q.-M. Zhang, and T. Zhou, "Uncovering missing links with cold ends," *Phys. A, Stat. Mech. Appl.*, vol. 391, no. 22, pp. 5769–5778, Nov. 2012.
[45] H.-H. Chen, L. Gou, X. Zhang, and C. L. Giles, "Discovering missing links in networks using vertex similarity measures," in *Proc. 27th Annu. ACM Symp. Appl. Comput. (SAC)*, 2012, pp. 138–143.
[46] R. N. Lichtenwalter and N. V. Chawla, "Vertex collocation profiles: Subgraph counting for link analysis and prediction," in *Proc. 21st Int. Conf. World Wide Web (WWW)*, 2012, pp. 1019–1028.
[47] J. Zhang, Y. Zhang, H. Yang, and J. Yang, "A link prediction algorithm based on socialized semi-local information," *J. Comput. Inf. Syst.*, vol. 10, no. 10, pp. 4459–4466, 2014.
[48] F. Tan, Y. Xia, and B. Zhu, "Link prediction in complex networks: A mutual information perspective," *PLoS ONE*, vol. 9, no. 9, pp. 1–8, 2014.
[49] Z. Liu, Q.-M. Zhang, L. Lü, and T. Zhou, "Link prediction in complex networks: A local Naïve Bayes model," *EPL (Europhysics Lett.)*, vol. 96, no. 4, p. 48007, Nov. 2011.
[50] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi, "From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks," *Sci. Rep.*, vol. 3, no. 1, p. 1613, Dec. 2013.
[51] H. N. Chua, W.-K. Sung, and L. Wong, "Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions," *Bioinformatics*, vol. 22, no. 13, pp. 1623–1630, Jul. 2006.
[52] G. Liu, J. Li, and L. Wong, "Assessing and predicting protein interactions using both local and global network topological metrics," *Genome Informat.*, vol. 21, pp. 138–149, 2008.
[53] Z. Burda, J. Duda, J. M. Luck, and B. Waclaw, "Localization of the maximal entropy random walk," *Phys. Rev. Lett.*, vol. 102, no. 16, Apr. 2009, Art. no. 160602.

[54] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens, "Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 3, pp. 355–369, Mar. 2007.
[55] P. Chebotarev and E. Shamis, "Matrix-forest theorems," 2006, *arXiv:math/0602575*. [Online]. Available: https://arxiv.org/abs/math/0602575
[56] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, "A measure of similarity between graph vertices: Applications to synonym extraction and Web searching," *SIAM Rev.*, vol. 46, no. 4, pp. 647–666, Jan. 2004.
[57] L. Yin, H. Zheng, T. Bian, and Y. Deng, "An evidential link prediction method and link predictability based on Shannon entropy," *Phys. A, Stat. Mech. Appl.*, vol. 482, pp. 699–712, Sep. 2017.
[58] X. Cai, J. Shu, and M. Al-Kali, "Link prediction approach for opportunistic networks based on recurrent neural network," *IEEE Access*, vol. 7, pp. 2017–2025, 2019.
[59] S. Rafiee, C. Salavati, and A. Abdollahpouri, "CNDP: Link prediction based on common neighbors degree penalization," *Phys. A, Stat. Mech. Appl.*, vol. 539, Feb. 2020, Art. no. 122950.
[60] M. Curado, "Return random walks for link prediction," *Inf. Sci.*, vol. 510, pp. 99–107, Feb. 2020.
[61] M. Zhou, Y. Kong, S. Zhang, D. Liu, and H. Jin, "The deep fusion of topological structure and attribute information for link prediction," *IEEE Access*, vol. 8, pp. 34398–34406, 2020.
[62] J. Liu, X. Jin, Y. Hong, F. Liu, Q. Chen, Y. Huang, M. Liu, M. Xie, and F. Sun, "Collaborative linear manifold learning for link prediction in heterogeneous networks," *Inf. Sci.*, vol. 511, pp. 297–308, Feb. 2020.
[63] S. Aslan and B. Kaya, "Time-aware link prediction based on strengthened projection in bipartite networks," *Inf. Sci.*, vol. 506, pp. 217–233, Jan. 2020.
[64] G. Chen, C. Xu, J. Wang, J. Feng, and J. Feng, "Robust non-negative matrix factorization for link prediction in complex networks using manifold regularization and sparse learning," *Phys. A, Stat. Mech. Appl.*, vol. 539, Feb. 2020, Art. no. 122882.
[65] K. Li, L. Tu, and L. Chai, "Ensemble-model-based link prediction of complex networks," *Comput. Netw.*, vol. 166, Jan. 2020, Art. no. 106978.
[66] Z. Zhao, X. Zhang, H. Zhou, C. Li, M. Gong, and Y. Wang, "HetNERec: Heterogeneous network embedding based recommendation," *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106218.
[67] S. S. Singh, S. Mishra, A. Kumar, and B. Biswas, "CLP-ID: Community-based link prediction using information diffusion," *Inf. Sci.*, vol. 514, pp. 402–433, Apr. 2020.
[68] M. Lim, A. Abdullah, N. Z. Jhanjhi, M. K. Khan, and M. Supramaniam, "Link prediction in time-evolving criminal network with deep reinforcement learning technique," *IEEE Access*, vol. 7, pp. 184797–184807, 2019.
[69] G. Chen, C. Xu, J. Wang, J. Feng, and J. Feng, "Graph regularization weighted nonnegative matrix factorization for link prediction in weighted complex network," *Neurocomputing*, vol. 369, pp. 50–60, Dec. 2019.
[70] W. Yuan, K. He, D. Guan, L. Zhou, and C. Li, "Graph kernel based link prediction for signed social networks," *Inf. Fusion*, vol. 46, pp. 1–10, Mar. 2019.
[71] X. Xu, N. Hu, T. Li, M. Trovati, F. Palmieri, G. Kontonatsios, and A. Castiglione, "Distributed temporal link prediction algorithm based on label propagation," *Future Gener. Comput. Syst.*, vol. 93, pp. 627–636, Apr. 2019.
[72] J. Wu, J. Shen, B. Zhou, X. Zhang, and B. Huang, "General link prediction with influential node identification," *Phys. A, Stat. Mech. Appl.*, vol. 523, pp. 996–1007, Jun. 2019.
[73] M. Wu, S. Wu, Q. Zhang, C. Xue, H. Kan, and F. Shao, "Enhancing link prediction via network reconstruction," *Phys. A, Stat. Mech. Appl.*, vol. 534, Nov. 2019, Art. no. 122346.
[74] J. Wang, Q.-M. Zhang, and T. Zhou, "Tag-aware link prediction algorithm in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 523, pp. 105–111, Jun. 2019.
[75] Z. Wang, Y. Wang, J. Ma, W. Li, N. Chen, and X. Zhu, "Link prediction based on weighted synthetical influence of degree and H-index on complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 527, Aug. 2019, Art. no. 121184.
[76] H. Shao, L. Wang, and Y. Ji, "Link prediction algorithms for social networks based on machine learning and HARP," *IEEE Access*, vol. 7, pp. 122722–122729, 2019.
[77] P. K. Sharma, S. Rathore, and J. H. Park, "Multilevel learning based modeling for link prediction and users' consumption preference in online social networks," *Future Gener. Comput. Syst.*, vol. 93, pp. 952–961, Apr. 2019.
[78] R. Pech, D. Hao, Y.-L. Lee, Y. Yuan, and T. Zhou, "Link prediction via linear optimization," *Phys. A, Stat. Mech. Appl.*, vol. 528, Aug. 2019, Art. no. 121319.
[79] S. Najari, M. Salehi, V. Ranjbar, and M. Jalili, "Link prediction in multiplex networks based on interlayer similarity," *Phys. A, Stat. Mech. Appl.*, vol. 536, Dec. 2019, Art. no. 120978.
[80] A. Kumar, S. Mishra, S. S. Singh, K. Singh, and B. Biswas, "Link prediction in complex networks based on significance of higher-order path index (SHOPI)," *Phys. A, Stat. Mech. Appl.*, vol. 545, May 2020, Art. no. 123790.
[81] S. Mallek, I. Boukhris, Z. Elouedi, and E. Lefèvre, "Evidential link prediction in social networks based on structural and social information," *J. Comput. Sci.*, vol. 30, pp. 98–107, Jan. 2019.
[82] H. J. Jeong and M. H. Kim, "Utilizing adjacency of colleagues and type correlations for enhanced link prediction," *Data Knowl. Eng.*, vol. 125, Jan. 2020, Art. no. 101785.
[83] H. Gao, J. Huang, Q. Cheng, H. Sun, B. Wang, and H. Li, "Link prediction based on linear dynamical response," *Phys. A, Stat. Mech. Appl.*, vol. 527, Aug. 2019, Art. no. 121397.
[84] K. Chi, G. Yin, Y. Dong, and H. Dong, "Link prediction in dynamic networks based on the attraction force between nodes," *Knowl.-Based Syst.*, vol. 181, Oct. 2019, Art. no. 104792.
[85] E. Bütün and M. Kaya, "A pattern based supervised link prediction in directed complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 525, pp. 1136–1145, Jul. 2019.
[86] J. Ai, Z. Su, Y. Li, and C. Wu, "Link prediction based on a spatial distribution model with fuzzy link importance," *Phys. A, Stat. Mech. Appl.*, vol. 527, Aug. 2019, Art. no. 121155.
[87] E. Bastami, A. Mahabadi, and E. Taghizadeh, "A gravitation-based link prediction approach in social networks," *Swarm Evol. Comput.*, vol. 44, pp. 176–186, Feb. 2019.
[88] L. A. Gundala and F. Spezzano, "Estimating node indirect interaction duration to enhance link prediction," *Social Netw. Anal. Mining*, vol. 9, no. 1, pp. 1–13, Dec. 2019.
[89] A. Mahmoudi, M. R. Yaakub, and A. Abu Bakar, "A new real-time link prediction method based on user community changes in online social networks," *Comput. J.*, vol. 63, no. 3, pp. 448–459, Mar. 2020.
[90] X. Zhang, W. Pang, and Y. Xia, "An intermediary probability model for link prediction," *Phys. A, Stat. Mech. Appl.*, vol. 512, pp. 902–912, Dec. 2018.
[91] Y. Yang, J. Zhang, X. Zhu, and L. Tian, "Link prediction via significant influence," *Phys. A, Stat. Mech. Appl.*, vol. 492, pp. 1523–1530, Feb. 2018.
[92] S.-Y. Wu, Q. Zhang, C.-Y. Xue, and X.-Y. Liao, "Cold-start link prediction in multi-relational networks based on network dependence analysis," *Phys. A, Stat. Mech. Appl.*, vol. 515, pp. 558–565, Feb. 2019.
[93] Y. Yao, R. Zhang, F. Yang, J. Tang, Y. Yuan, and R. Hu, "Link prediction in complex networks based on the interactions among paths," *Phys. A, Stat. Mech. Appl.*, vol. 510, pp. 52–67, Nov. 2018.
[94] C. Chiu and J. Zhan, "Deep learning for link prediction in dynamic networks using weak estimators," *IEEE Access*, vol. 6, pp. 35937–35945, 2018.
[95] Z. Wang, J. Liang, and R. Li, "A fusion probability matrix factorization framework for link prediction," *Knowl.-Based Syst.*, vol. 159, pp. 72–85, Nov. 2018.
[96] J. Wu, "A generalized tree augmented naive Bayes link prediction model," *J. Comput. Sci.*, vol. 27, pp. 206–217, Jul. 2018.
[97] T. Li, J. Zhang, P. S. Yu, Y. Zhang, and Y. Yan, "Deep dynamic network embedding for link prediction," *IEEE Access*, vol. 6, pp. 29219–29230, 2018.
[98] E. Bütün, M. Kaya, and R. Alhajj, "Extension of neighbor-based link prediction methods for directed, weighted and temporal social networks," *Inf. Sci.*, vols. 463–464, pp. 152–165, Oct. 2018.
[99] B. Moradabadi and M. R. Meybodi, "Link prediction in stochastic social networks: Learning automata approach," *J. Comput. Sci.*, vol. 24, pp. 313–328, Jan. 2018.
[100] K.-K. Shang, M. Small, and W.-S. Yan, "Link direction for link prediction," *Phys. A, Stat. Mech. Appl.*, vol. 469, pp. 767–776, Mar. 2017.

[101] C. Dai, L. Chen, B. Li, and Y. Li, "Link prediction in multi-relational networks based on relational similarity," *Inf. Sci.*, vols. 394–395, pp. 198–216, Jul. 2017.
[102] A. Mohan, R. Venkatesan, and K. V. Pramod, "A scalable method for link prediction in large real world networks," *J. Parallel Distrib. Comput.*, vol. 109, pp. 89–101, Nov. 2017.
[103] M. Gao, L. Chen, B. Li, Y. Li, W. Liu, and Y.-C. Xu, "Projection-based link prediction in a bipartite network," *Inf. Sci.*, vol. 376, pp. 158–171, Jan. 2017.
[104] W. Wang, Y. Feng, P. Jiao, and W. Yu, "Kernel framework based on non-negative matrix factorization for networks reconstruction and link prediction," *Knowl.-Based Syst.*, vol. 137, pp. 104–114, Dec. 2017.
[105] X. Ma, P. Sun, and G. Qin, "Nonnegative matrix factorization algorithms for link prediction in temporal networks using graph communicability," *Pattern Recognit.*, vol. 71, pp. 361–374, Nov. 2017.
[106] C. Ma, Z.-K. Bao, and H.-F. Zhang, "Improving link prediction in complex networks by adaptively exploiting multiple structural features of networks," *Phys. Lett. A*, vol. 381, no. 39, pp. 3369–3376, Oct. 2017.
[107] J. Li, B. Ge, K. Yang, Y. Chen, and Y. Tan, "Meta-path based heterogeneous combat network link prediction," *Phys. A, Stat. Mech. Appl.*, vol. 482, pp. 507–523, Sep. 2017.
[108] Z. Zhang, J. Wen, L. Sun, Q. Deng, S. Su, and P. Yao, "Efficient incremental dynamic link prediction algorithms in social network," *Knowl.-Based Syst.*, vol. 132, pp. 226–235, Sep. 2017.
[109] Y. Li, P. Luo, Z.-P. Fan, K. Chen, and J. Liu, "A utility-based link prediction method in social networks," *Eur. J. Oper. Res.*, vol. 260, no. 2, pp. 693–705, Jul. 2017.
[110] B. Moradabadi and M. R. Meybodi, "A novel time series link prediction method: Learning automata approach," *Phys. A, Stat. Mech. Appl.*, vol. 482, pp. 422–432, Sep. 2017.
[111] S. Liu, X. Ji, C. Liu, and Y. Bai, "Extended resource allocation index for link prediction of complex network," *Phys. A, Stat. Mech. Appl.*, vol. 479, pp. 174–183, Aug. 2017.
[112] P. Pei, B. Liu, and L. Jiao, "Link prediction in complex networks based on an information allocation index," *Phys. A, Stat. Mech. Appl.*, vol. 470, pp. 1–11, Mar. 2017.
[113] J. Wu, G. Zhang, and Y. Ren, "A balanced modularity maximization link prediction model in social networks," *Inf. Process. Manage.*, vol. 53, no. 1, pp. 295–307, Jan. 2017.
[114] A. Mahmoudi, M. R. Yaakub, and A. Abu Bakar, "A new method to discretize time to identify the milestones of online social networks," *Social Netw. Anal. Mining*, vol. 8, no. 1, p. 34, Dec. 2018.
[115] A. Mahmoudi, M. R. Yaakub, and A. Abu Bakar, "New time-based model to identify the influential users in online social networks," *Data Technol. Appl.*, vol. 52, no. 2, pp. 278–290, Apr. 2018.
[116] G. Nason. (2019). *Network Time Series*. Accessed: Nov. 4, 2019. [Online]. Available: https://people.maths.bris.ac.uk/~magpn/networktimeserie.html
[117] P. Holme, "Temporal networks," in *Encyclopedia of Social Network Analysis and Mining*. New York, NY, USA: Springer, 2018, pp. 3053–3062.
[118] K.-K. Shang, W.-S. Yan, and M. Small, "Evolving networks–using past structure to predict the future," *Phys. A, Stat. Mech. Appl.*, vol. 455, pp. 120–135, Aug. 2016.
[119] J.-G. Liu, L. Hou, X. Pan, Q. Guo, and T. Zhou, "Stability of similarity measurements for bipartite networks," *Sci. Rep.*, vol. 6, no. 1, p. 18653, May 2016.
[120] J. Leskovec and A. Krevl. (Jun. 2014). *SNAP Datasets: Stanford Large Network Dataset Collection*. [Online]. Available: http://snap.stanford.edu/data
[121] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. AAAI*, 2015, pp. 4292–4293.
[122] J. Kunegis, "KONECT: The koblenz network collection," in *Proc. 22nd Int. Conf. World Wide Web (WWW) Companion*, 2013, pp. 1343–1350.
[123] L. Xu, X. Wei, J. Cao, and P. S. Yu, "On learning mixed community-specific similarity metrics for cold-start link prediction," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 861–862.
[124] D. Lande, M. Fu, W. Guo, I. Balagura, I. Gorbov, and H. Yang, "Link prediction of scientific collaboration networks based on information retrieval," *World Wide Web*, vol. 23, no. 4, pp. 2239–2257, Jul. 2020.

HERMAN YULIANSYAH (Graduate Student Member, IEEE) received the S.T. degree from the Faculty of Engineering, Universitas Muhammadiyah Yogyakarta (UMY), Indonesia, in 2007, and the M.Eng. degree in information technology from Universitas Gadjah Mada (UGM), Indonesia, in 2011. He is currently pursuing the Ph.D. degree with the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM).

He has been working as a Lecturer with the Department of Informatics, Universitas Ahmad Dahlan (UAD), Indonesia, since 2011. His research primarily focuses on data mining and social network analytics.

ZULAIHA ALI OTHMAN was born in Malaysia. She received the B.Sc. degree (Hons.) in computer science from the National University of Malaysia, in 1990, the M.Sc. degree in computing from The University of Sheffield, U.K., in 1998, and the Ph.D. degree in computing from Sheffield Hallam University, U.K., in 2004. Since 2004, she has been a Lecturer with the Centre of Artificial Intelligence Technology, Faculty Information Science and Technology, The National University of Malaysia, and has become an Associate Professor since 2009. She has published more than 140 articles, in which 90% of publications are related to artificial intelligence, multi agent, and data analytic.

AZURALIZA ABU BAKAR received the Ph.D. degree in artificial intelligence from Universiti Putra Malaysia (UPM), in 2002.

She has been a Professor in Data Mining with the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), since 2011. Her main research interests include data analytics and artificial intelligence specifically in rough set theory, feature selection algorithms, nature inspired computing, and sentiment analysis. She served as an Advisor for the Data Mining and Optimization Laboratory and a member for the Sentiment Analysis Laboratory. She has led 13 research projects (including three in progress) and a member of 42 research projects. She is a member of IEEE Computational Intelligence Society. She has also served on roughly 30 conference and workshop program committees and has served as the Program Chair.