Journal of Network and Computer Applications 166 (2020) 102716

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Applications of link prediction in social networks: A review

Nur Nasuha Daud [a], Siti Hafizah Ab Hamid [a,*], Muntadher Saadoon [a], Firdaus Sahran [b], Nor Badrul Anuar [b]

[a] Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia
[b] Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia

ARTICLE INFO

Keywords:
Link prediction
Social network
Social network analysis

ABSTRACT

Link prediction methods anticipate the likelihood of a future connection between two nodes in a given network. The methods are essential in social networks to infer social interactions or to suggest possible friends to the users. Rapid social network growth trigger link prediction analysis to be more challenging especially with the significant advancement in complex social network modeling. Researchers implement numerous applications related to link prediction analysis in different network contexts such as dynamic network, weighted network, heterogeneous network and cross network. However, link prediction applications namely, recommendation system, anomaly detection, influence analysis and community detection become more strenuous due to network diversity, complex and dynamic network contexts. In the past decade, several reviews on link prediction were published to discuss the algorithms, state-of-the-art, applications, challenges and future directions of link prediction research. However, the discussion was limited to physical domains and had less focus on social network perspectives. To reduce the gap of the existing reviews, this paper aims to provide a comprehensive review and discuss link prediction applications in different social network contexts and analyses, focusing on social networks. In this paper, we also present conventional link prediction measures based on previous researches. Furthermore, we introduce various link prediction approaches and address how researchers combined link prediction as a base method to perform other applications in social networks such as recommender systems, community detection, anomaly detection and influence analysis. Finally, we conclude the review with a discussion on recent researches and highlight several future research directions of link prediction in social networks.

1. Introduction

Social network is one of the favorite means for a modern society to perform social interactions and exchange information via the Internet. It allows people to share their contents without boundaries. People communicate and express their comments, likes and interests via social network as it provides a fast and easy way to share. As a result, it creates a complex connection within the social network and this can be immediately visualized as large graphs. Today, these large and complex graphs represent merely 50% of the world population, or 3.196 billion active social network users, in comparison to the world population of 7.593 billion (Kemp, 2018). With the dynamic and exponential rise in the number of new users, this graph expands over time, with some of the users come from unreliable and untrusted sources. Although connectivity increases the link between users, it also attracts unwanted users and links such as fake users and bots. Consequently, it leads to the following questions; how strong are the connections? What are the factors that drive the links? Also, how does the engagement between nodes help to predict the future nodes?

Link prediction is a study that assumes the upcoming behaviors in social networks. For instance, to predict future collaborators for two researchers in a co-authorship network. Besides, there are some other applications such as recommendation system, anomaly detection, community detection and influence analysis that adopt link prediction to minimize the existing problems and increase the detection process. For instance, Dong et al. (2012) proposed a recommendation system that presented a solution for transfer link prediction problem across heterogeneous networks. Meanwhile, Kagan et al. (2018) introduced anomaly detection that used link prediction algorithms to detect malicious users in the network. Mohan et al. (2017), on the other hand, proposed an advance community detection system that utilized link prediction algorithms and Bulk Synchronous Parallel programming model. They meant

* Corresponding author.
E-mail addresses: nasuha@um.edu.my (N.N. Daud), sitihafizah@um.edu.my (S.H. Ab Hamid), core93@yahoo.com (M. Saadoon), firdaussahran@um.edu.my (F. Sahran), badrul@um.edu.my (N.B. Anuar).

https://doi.org/10.1016/j.jnca.2020.102716
Received 21 May 2019; Received in revised form 8 February 2020; Accepted 15 May 2020
Available online 21 May 2020
1084-8045/

to cater a single system environment and increase the scalability of link prediction algorithms in community detection. Finally, Lu and Szymanski (2017) analyzed information diffusion and how its influence occurred in the network to straightly predict an emergent cascade related to viral events.

For the last 10 years, the popularity of link prediction attracts even more studies for further improvement and application of link prediction in other domains. For example, a review by Lü and Zhou (2011) discussed an overview of link prediction algorithms. They restricted the discussion on physical approaches only and focused less on computer perspective. Next, Wang et al. (2015) published an extensive literature on link prediction techniques and discussed the current problems such as temporal link prediction, heterogeneous network and link prediction scalability. They also highlighted the challenges and future directions of link prediction study. Martínez et al. (2016), in their review, underlined the current state-of-the-art in complex networks by explaining link prediction and methods in a set of networks with different properties. Finally, the latest review by Kushwah and Manjhvar (2016) focused on the recent growth of link prediction algorithms and current research on link prediction problem.

To reduce the gap of the existing reviews, this paper aims to present a new link prediction survey which includes the latest approaches and applications and to solve current problems such as large-scale networks, multi-dimensional networks, scalability and network dynamicity. This paper also provides recommendations and new insights for further improvement on link prediction analysis in social networks. Additionally, this survey represents the first attempt to explain link prediction applications in different contexts and analyses with a specific focus on social networks. Finally, the contributions of this paper are as follows:

a) Present existing applications of link prediction in different social network contexts and analyses.
b) Introduce the objectives of link prediction applications in different social network contexts that previous researchers presented.
c) Explain how researchers combined link prediction as a base method to perform other applications in social networks such as recommender systems, community detection, anomaly detection and influence analysis.
d) Highlight limitations and future research directions of link prediction in social networks.

The paper is organized as follows; firstly, we introduce a basic understanding of link prediction and social networks in section 2. In section 3, we provide an overview of link prediction process and its approaches. In section 4, we present existing link prediction applications that focus on social networks such as anomaly detection, community detection, influence analysis and recommender system. Afterward, we present implementations of link prediction in different social network contexts. Lastly, we discuss the restrictions and future research trends of link prediction in section 5 and conclude the review in section 6.

2. Background

In this section, we present the background information of the present study. First, we discuss the supremacy of social networks and their current attractions. Next, we examine the extent of link prediction analyses and finally, we explain link prediction with a focus on social network domains.

2.1. Social networks

Humans form networks among each other since time immemorial. The first network humans ever created was based on relationships with family, friends, clans or tribes to sustain human continuity (Barnett, 2011). The social interaction between humans began face-to-face such that in mother-daughter communication or teacher-students' knowledge sharing. Later, as people started introducing the Internet to the world, the network interaction evolved from offline interaction to online interaction. Social networks are a popular medium that is actively used worldwide. Through social network platforms, people make connections and keep in touch with each other. There are distinct types of social network platforms where each serves their users with different purposes as illustrated in Fig. 1.

Fig. 1 shows that the earliest social network platform, SixDegree was created in 1997 to let people stay connected and share content within their social circles. Then, Friendster came out in 2002 as the first social network platform that attained over 3 million users within the first few months since its launch. It allowed users to discover new friends, as well as to share photos and videos. In 1998, OpenDiary was created as the first blogging platform that brought online journal keepers together to write and share each other's journals and leave comments on them. Next, in 1999, Epinions came out as the pioneer of opinion sharing platform for its users to share and discuss each other's reviews on specific content such as restaurants and places. Then, live news and events platforms were introduced in 2002. For example, Google News, a platform that allowed the users to deliver and stay up to date with current news and stories on the Internet, whereas Meetup updated its users with the latest events, entertainment and movies. In 2006, Twitter became available to the world and up to this day, it becomes one of the most popular live news social network platforms.

Xing came out in 2003 as a social network platform that enabled the users to stay connected with career and business opportunities around them. In 2005, MocoSpace, a gaming networking platform was created. It allowed the users to stream and share a real-time video game to the world. On the same year, Youtube was also founded as an entirely new kind of social network platform that allows the users to share and view videos from all over the world. However, the launch of Badoo in 2006 gave a whole different purpose to the existing social network platforms, where it allowed users to meet their ideal partner and go out on a date. In 2008, an academic social network platform, ResearchGate was founded for researchers to share and stay updated with current research trends in the academic realm. Later, Facebook introduced a technology called Facebook Messenger, an instant messaging application that allowed users to communicate individually or in a group by allowing text, voice, image, video and document sharing. Other popular instant messaging platforms such as Whatsapp, Wechat, Line and Telegram began to surface right after to serve their users with more exciting features.

In conclusion, social networks are significant and convenient for human communication as proven by Facebook that has 2234 million active users in October 2018 (Kemp, 2018). The active users generated massive data from their activities, thus provided the opportunities for researchers to conduct various social network analyses. Recent analyses of social networks include recommending new friends, partner or places to the users in the network (Wu et al., 2013), detecting fake profiles that exist in the network (Kagan et al., 2018), analyzing the influence of information throughout the network (Zhang et al., 2015) and predicting new connection or activities to the users using link prediction (Liu et al., 2018).

2.2. Link prediction

Link prediction is a process of predicting future connections between pairs of unconnected links based on existing connections. Consider the example in Fig. 2, *Node1* is linked to *Node2* and *Node4* simultaneously. However, *Node2* and *Node4* are not linked to each other but *Node1* is a common node of *Node2* and *Node4*. Hence, link prediction should anticipate a connection between *Node2* and *Node4*. Besides analyzing and predicting future connections, link prediction also infers missing links that may appear or disappear in the network during the interval of time, t to a given future time, t' (Liben-Nowell and Kleinberg, 2003). The network is a graph where it represents nodes as entities and edges as interactions or links. The graph, $G = (V, E)$ consists of a set of nodes or

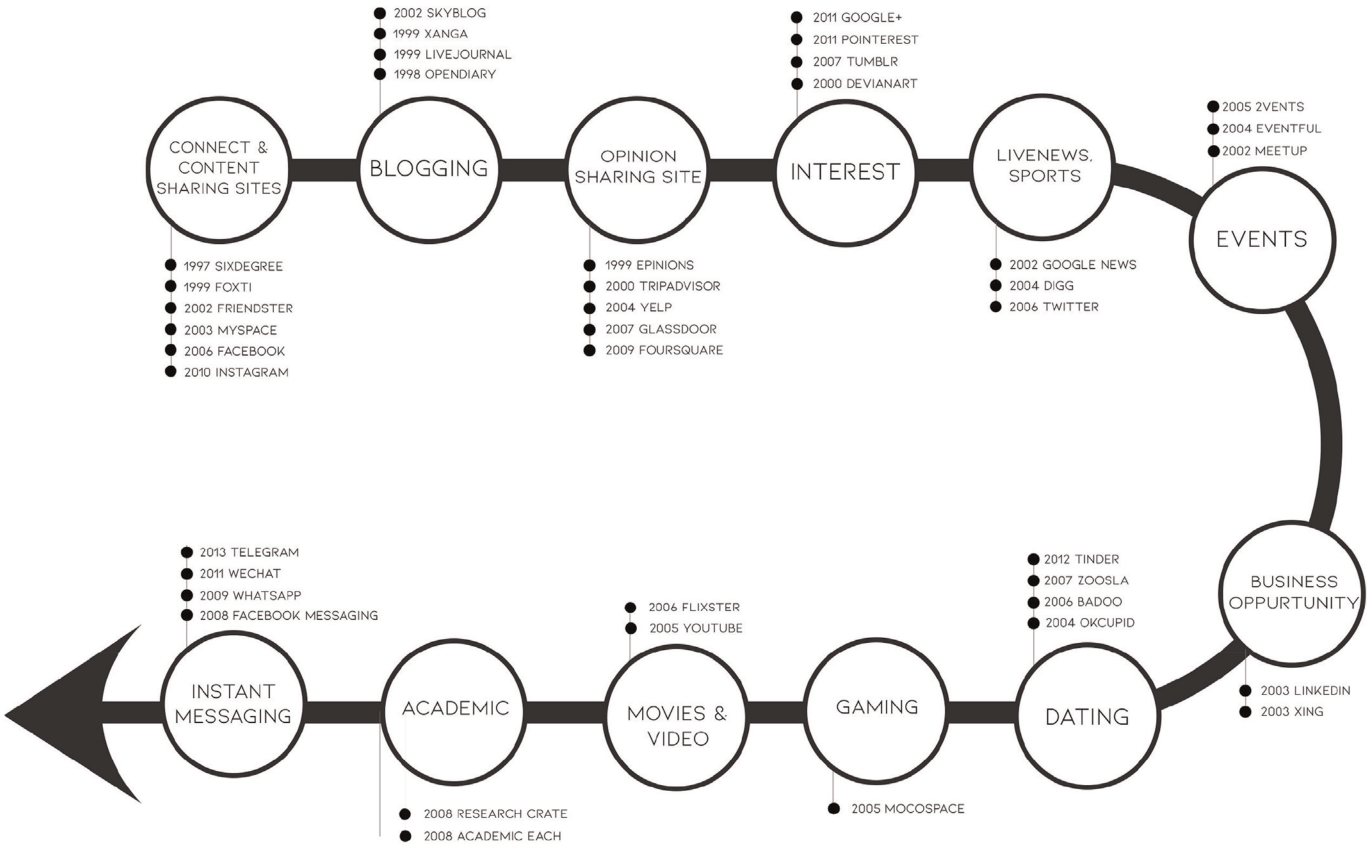

Fig. 1. Timeline of social networks based on category.

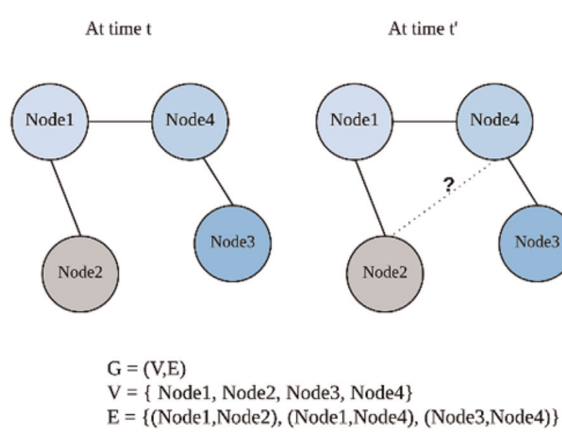

Fig. 2. Example for simple network graph of V and E.

vertices, *V* and a set of links or edges, *E*. For a given graph, each edge, $e = (u, v)$ represents an interaction between u and v that takes place in a particular time, $e(t)$. A link prediction measure is used to forecast whether or not there is a future link between pairs of unconnected nodes or vertices.

Generally, measure is defined as a standard metric assigned to an object or event to be compared with other similar objects or events. Link prediction uses conventional measures to determine the likelihood of future association between two nodes and such is dependent on graph-based measures or content-based measures:

a) Graph-based measures

Graph-based measures are the most basic and trivial approach that use the topological structure of networks. Researchers adapted graph-based measures like Common Neighbors (Zeng, 2016), Adamic-Adar (Deylami and Asadpour, 2015) or Katz (Rattigan and Jensen, 2005) in link prediction due to the difficulty of acquiring sufficient content information from a certain network. For example, measuring the graph distance as illustrated in Fig. 3. In this approach, score *(x,y)* for each edge is assigned and ranked based on the shortest path between the *x* and *y*.

b) Content-based measures

Content-based measures often employ attributes of vertices and edges. For example, in a co-authorship network, the nodes are authors while the edges are the authors' collaborations (the authors' co-authoring a paper). The attributes that can be used as content to predict future collaborations between the authors are title, abstract, keywords of papers and published year as shown in Fig. 3. Recently, Zhao et al. (2017) invented a probabilistic approach in a citation network that encompassed various kinds of node attributes such as authors' names, abstract, venue,

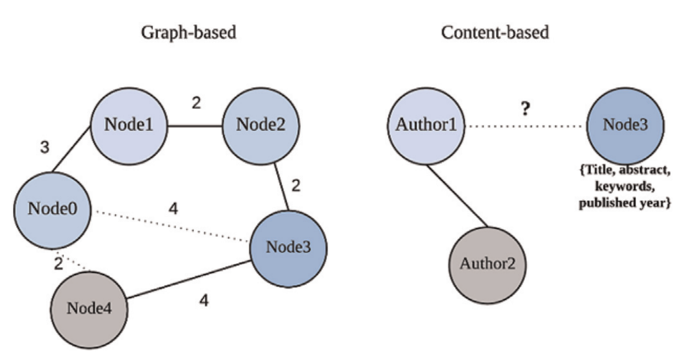

Fig. 3. Graph-based and content-based measures for link prediction.

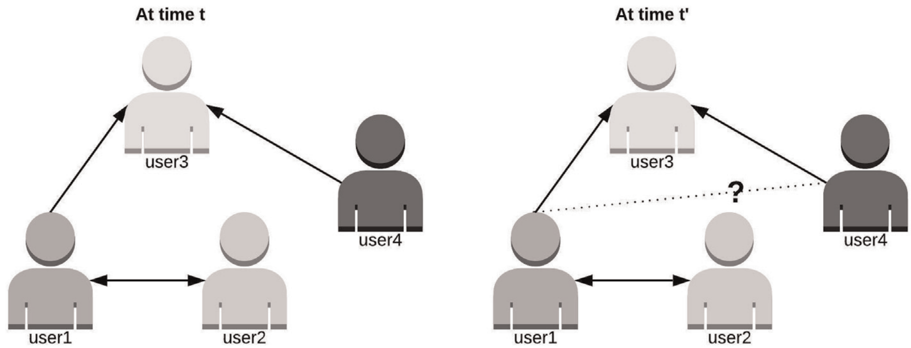

Fig. 4. Twitter follower/following graphs at Interval time *t* and *t.'*

year and number of citations. The work showed that by incorporating more information, higher prediction results could be achieved.

Researchers are actively inventing several distinct approaches that utilize both graph-based and content-based measures to achieve higher prediction performance in a variety of domains. Domains like bibliographic, biological, protein-protein interaction and social networks receive high efficiency in predicting future links.

2.3. Link prediction in social networks

Link prediction is an effective way to picture social interactions among users of a particular social network. Consider the social network platform Twitter as a graph, in which a node represents a Twitter user and a link represents the connection between two nodes or users. The link can be of any type of social link between two users such as tweet, comment, retweet, following and favorite.

Fig. 4 shows the topology of social networks from a period *t* to *t*'. Link prediction predicts the future connection between unconnected nodes, *user1* and *user4* at the period *t*' where *t*' > *t*. Fig. 5 illustrates the taxonomy of link prediction in social networks. The measures used in previous studies of link prediction are shown in Section 2.2. Furthermore, the approaches used can be broadly merged under four main categories: similarity, probabilistic, algorithmic and hybrid, as further explained in Section 3. Finally, applications and social network contexts of link prediction are briefly described in Section 4.

3. Link prediction approaches in social networks

Previous literature demonstrated various successful approaches on link prediction analysis in social networks. Fig. 6 shows the extension of the taxonomy proposed by Martínez et al. (2016) where they extensively studied the art of link prediction approaches and methods. However, due to the evolution of technologies and social networks, there is a need for a new review on the state-of-the-art link prediction approaches. In this work, we analyze 60 papers from highly credible sources such as IEEExplore, ScienceDirect, ACM, Springer and Google Scholar, focusing on link prediction approaches in social networks from the year 2012 until the present. We divide the approaches into a taxonomy which classifies the approaches, namely similarity, probabilistic, algorithmic and hybrid approaches, according to the prediction strategy, the technique used and complexity of the approaches.

3.1. Similarity

Similarity approaches are a conventional link prediction approach that seem to be more promising in comparison to other approaches due to its simplicity and lower computational time. In similarity approaches, for each pair of unconnected nodes (*user1, user4*), the similarity scores are

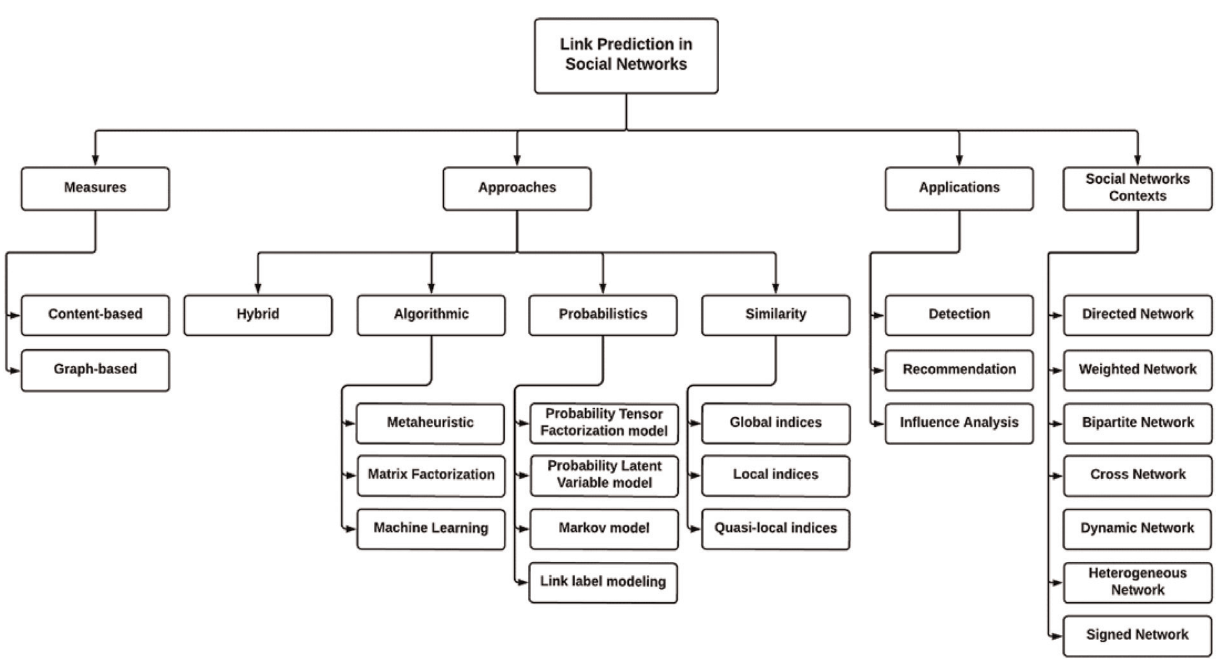

Fig. 5. Taxonomy of link prediction in social networks.

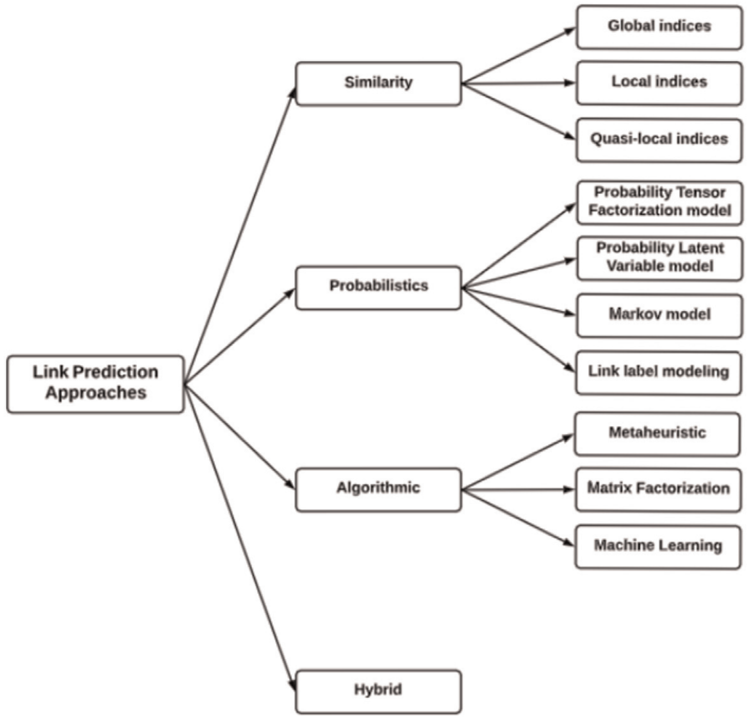

Fig. 6. Link prediction approaches.

computed based on selected link prediction measures, graph-based, content-based or both. A node pair with higher similarity score is assumed to establish a link in the future. Initial researches on similarity approaches focused on utilizing graph-based measure that can be viewed on two levels; local and global level and further classified into three types namely local indices, global indices and quasi-local indices.

a) Local indices

Local indices are one of the most straightforward approaches, considering the number of neighbor nodes and the degree of neighbor to calculate the similarity score in link prediction. A node is often considered as a neighbor node when the path distance is less than two. Examples of local indices include Common Neighbors, Salton index, Jaccard index, Sorensen index, hub promoted index, Leicht-Holme-Newman index, Preferential Attachment index, Adamic Adar index and Resource Allocation index. Local similarity indices are widely used in real applications as they lower resource utilization and computational complexity while maintaining best prediction performance. For example, local similarity indices get a greater prediction accuracy with AUC (Area Under Curve) scores up to 0.99, by combining topological structure with the computed similarity score (Xu and Yin, 2017a) or with additional community information (Sun et al., 2017). However, since local indices only consider neighbor nodes, some interesting and potential links may be missed. Table 1 shows researches in the year 2016 and 2017 that focused on improving the prediction accuracy in varying networks by improvising existing local similarity indices. Zeng (2016) proposed a combination of two local indices, Common Neighbors and Preferential Attachment that achieved even greater prediction performance than global indices. On the other hand, Hou and Liu (2017) defined the similarity intensity of varying networks to explain different behaviors of prediction in different networks. However, the local indices can be further improved by developing a parallel algorithm for the similarity calculation task.

b) Global indices

Global indices compute similarity score based on the global link structure of graph, where nodes have a path distance of more than two. In other words, global indices utilize the whole network of topological information to score each link. As compared to local index approaches, global indices identify all direct and indirect paths that are interesting to be included in the similarity score. Existing global indices include Katz index, the Leicht-Holme-Newman index and the Matrix Forest index. Table 2 shows existing researches utilized the global similarity approach for link prediction in social networks. Muniz et al. (2018) proposed a combination of global similarity indices and content-based measure in their work to improve the performance of link prediction. In the context of link prediction in large networks, global similarity indices are time-consuming and computationally complex due to high dimensionality of networks. Nevertheless, Coskun and Koyuturk (2015) extensively proposed dimensionality reduction to a global based approach and achieved better prediction accuracy of 0.8 AUC score. In the future, there remains a need for more scalable global approach to handle link prediction in a distributed environment.

c) Quasi-local indices

Quasi-local approaches use additional topological information as global indices do. They compute the score based on nodes with a path distance of no more than two, which is similar to local index approaches. Quasi-local indices, which include local path index, local random walk and superposed random walk, provide a tradeoff between the complexity of a model and prediction accuracy. They achieve higher prediction accuracy than local approaches as they consider additional topological information while acquiring lower computation complexity (Ozcan and Oguducu, 2016). Liu et al. (2016) proposed a method, Degree-related clustering ability path (DCP) outperformed four conventional common-neighbor based methods in terms of its accuracy and precision. The performance of quasi-local approaches is highly dependent on the dataset and the application. Therefore, the quasi-local approach can be further improved by developing an algorithm that is able to compute similarity index for the entire network with more accuracy and stability. Table 3 shows the current quasi-local similarity approaches for link

Table 1
Local similarity approaches for link prediction in social networks.

Year	Reference	Objectives	Application	Category	Approaches
2018	Wu et al., 2018	Propose and investigate the power of local asymmetric clustering information to improve prediction accuracy	Link prediction in Dolphins, Mouse, Macaque, Food, C.elegans, Yeast, Hamster, Political Blogs, USAir	Local	Asymmetric link clustering
2017	Hou and Liu, 2017	Define the similarity intensity to quantify the governance of similarity in complex networks	Link prediction in Facebook, Yelp, Gowalla, Flixster, The Trust, Pretty-Good-Privacy, Enron, Yeast and PDZBase	Local	Symmetrical similarity index
2017	Sun et al., 2017	Improve link prediction accuracy	Link prediction in C.elegans, Zachary Karate club, Lesmis, Netscience, Dolphins, Polbooks, Metabolic, Polblogs, Football, Hep	Local	Common neighbor and community structure
2017	Qian et al., 2017	Improve performance	Link prediction in C.elegans, NetScience, Jazz, USAir, Food Web, Facebook, Political Blogs, Router, Power, WCGScience, Yeast	Local	TPSR indices (Topological properties and Strong ties, clustering coefficient, degrees)
2017	Xu and Yin, 2017b	Improve prediction accuracy and propose a new similarity	Link prediction in USAir, NetScience, Political Blogs, Yeast, C.elegans, Power, Router, Email	Local	CRA index, common neighbors and end nodes
2016	Zeng, 2016	Improve prediction accuracy based on known interactions	Link prediction in PPI, NetScience, Grid, Political Blogs, USAir	Local	Common neighbors and preferential attachment
2016	Wu et al., 2016	Improve performance in predicting missing links	Link prediction in Food, Grassland, Jazz, Dolphins, Macaque, Mouse, Political Blogs, Email, Grid, INT	Local	Local link structure and clustering coefficient of common neighbors
2016	Yao et al., 2016	Consider link prediction in a dynamic network	Link prediction in Bibliographic database DBLP	Local	Weighted network, common neighbor and intimacy between common neighbors

prediction in social networks.

3.2. Probabilistic

Probabilistic approaches solve the problem of link prediction by building a statistical probability model that fits with the network structure. The model, specified with parameters, computes a mathematical statistic to produce a probability value for each pair of nodes. Then, the probability values are classified according to the hypothesis where the higher the probability value of the node pairs, the higher the possibility of link formation between the node pairs. Table 4 lists recent probabilistic models proposed by previous researchers. The existing link prediction probabilistic approaches, specifically in social networkdomain, reviewed in this paper are divided into four categories namely, Probability Tensor Factorization model, Probability Latent Variable model, Markov model and Link Label Modeling.

a) Probability Tensor Factorization

Gao et al. (2012) introduced Probability Tensor Factorization model which is the extension of Probability Matrix Factorization (Salakhutdinov and Mnih, 2008) to the tensor factorization version; as to consider multi-relational social network problem. Later, Ermis and Cemgil (2014) introduced the Probability Tensor Factorization model that considered Variational Bayes which was based on tensor factorization and led to an improved prediction performance in larger scale network.

b) Probability Latent Variable

Inspired by properties of block structure that show links in the same adjacent matrices of blocks of networks that often have similar properties; Yang et al. (2014) proposed Probability Latent Variable model that combined the ideas of low-rank approximations and block structure for matrices to give better prediction accuracy than the original method alone, that was low-rank approximations.

c) Markov model

Markov model, a stochastic model, is used to model systems that change randomly in the theory of probability. Markov model visualizes the dynamic network evolution as a process. Due to the integration of time scale with user interactions, Das and Das (2017) proposed a stochastic Markov model to demonstrate link prediction in time-varying social networks with a better prediction performance than the existing dynamic link prediction. However, the tradeoff for using probabilistic model is as the complexity of the model increases, the number of parameters set also increases.

d) Link label modeling

Finally, a link label modeling is a probabilistic approach that specifically focuses on solving link label prediction problem in signed social networks. In signed social networks, the nature of interactions or links exist among users can be of positive or negative. Positive links signify friendship or approval, whereas negative links indicate antagonism or disapproval. Recently, Javari et al. (2018) introduced Link Label

Table 2
Global similarity approaches for link prediction in social networks.

Year	Reference	Objectives	Application	Category	Approaches
2018	Muniz et al., 2018	Improve prediction performance of link prediction methods	Link prediction in (Astrophysics, Cond-mat, Gr-qc, Hep-th, Hep-ph)	Global	Weighted similarity index combining temporal, topological and contextual information
2015	Coskun and Koyuturk, 2015	Improve prediction performance by handling high dimensionality problem in large networks	Link prediction in co-authorship network DBLP	Global	Global Topological Similarity method with dimensionality reduction

Table 3
Quasi-local similarity approaches for link prediction in social networks.

Year	Reference	Objectives	Application	Category	Approaches
2017	Wang et al., 2017	Improve link prediction accuracy using community information	Link prediction in (C.elegans, Dolphins, Football, Karate, Metabolic, NetScience, Political Blogs, Power, USAir, and Yeast)	Quasi-local	Intra-community and Resource Allocation index, (ICRA)
2016	Ozcan and Oguducu, 2016	Link prediction in temporal network using time series method	Link prediction in DBLP coauthorship networks	Quasi-local	LP, LRW, SRW, CN, AA, PA
2016	Liu et al., 2016	Quantify the clustering ability of nodes	Link prediction in (Dolphin, SmaGri, USAir, Geom, Yeast, Metabolic, Political Blogs, Email, NetScience, C.elegans)	Quasi-local	DCP (Degree related and clustering ability of each path)
2016	Chen et al., 2016	Improve similarity based method for large networks and propose new method of global based similarity score	Link prediction in (PPI, NetScience, Grid, INT, Political Blogs, USAir)	Quasi-local	Node_LP (SRW and LRW)

Table 4
Probabilistic approaches for link prediction in social networks.

Year	Reference	Objectives	Approaches
2018	Javari et al., 2018	Introduce a novel probabilistic approach for link prediction to adapt the sparsity of data in social networks	Link label modeling based on the local and global structure
2017	Das and Das, 2017	Propose a novel Markov prediction model, with time-varying graph in dynamic social networks	Stochastic Markov model and Time-varying graph
2017	Zhao et al., 2017	Present a Bayesian probabilistic approach that incorporates various kinds of node attributes in directed and undirected relational networks	Node Attributes Relational Model
2014	Ermis and Cemgil, 2014	Propose a probabilistic approach based on Variational Bayes (VB) to provide better prediction performance in large scale network	Bayesian Tensor Factorization Model via Variational Bayes (VB)
2014	Yang et al., 2014	Find missing edges using a probabilistic latent variable model and convex nonnegative matrix factorization with block structures detection	Probabilistic Latent Variable Model and Low-Rank Approximation
2012	Gao et al., 2012	Focus on the task of link prediction in multiple relation types among object pairs in multi-relational networks	Probabilistic Latent Tensor Factorization with Hierarchical Bayesian

Modeling based on local and global network structures to adapt data sparsity problem in predicting link. Although probabilistic models demonstrate better prediction performance than similarity approaches, the application on large scale network often results in long computation times, depending on its model complexity.

3.3. Algorithmic

Algorithmic approaches are another high performance approach that are widely explored among researchers in the link prediction literature. The algorithmic approaches allow researchers to perform calculation, data processing and automated prediction tasks in a process that follows certain rules and problem-solving operations. In comparison to the similarity and probabilistic approaches, the algorithmic approaches often employ extra information from the network and consider other factors that affect link formation For instance, algorithmic methods use community structure information (Mohan et al., 2017), users' behavior (Xiao et al., 2018), temporal information, weighted information, network structural information, latent features of the network and matrix factorization and subsequently enhance the performance of link prediction. Algorithmic approaches that are reviewed in this paper fall into three categories namely, Metaheuristic, Matrix Factorization and Machine Learning.

a) Metaheuristic

Metaheuristic is a framework with high-level problem-independent techniques that provides a set of guidelines or strategies on how to solve a set of problems by using heuristic optimization method in order to exploit its best capabilities and achieve better solutions. Several successful metaheuristic algorithms which were applied for link prediction in social networks are as listed in Table 5. Bliss et al. (2014) proposed the first kind of metaheuristic approaches which introduced an evolutionary algorithm in combination with sixteen similarity indices. They also proved that the evolved predictor outperformed other individual similarity indices. Next, Bastami et al. (2018) proposed a multi-level approach that showed high accuracy and speed with much less CPU time and memory space usage. The approach was inspired by graph structure and community features, making it beneficial for big social

Table 5
Metaheuristic for link prediction in social networks.

Year	Reference	Objectives	Approaches
2018	Srilatha and Manjula, 2018	Propose firefly based link prediction algorithm to improve performance of similarity approach	Similarity index with Swarm firefly algorithm
2018	Bastami et al., 2018	Propose multi-level approach consisting of method scalability, parallel prediction operations, prediction accuracy, speed improvement, computational complexity reduction (by subgraph optimization) and local prediction improvement to improve speed and accuracy of link prediction	gravitation based, multi-level learning and subgraph optimization
2014	Bliss et al., 2014	Propose an evolved predictor consisting of sixteen similarity indices and apply Covariance Matrix Adaptation Evolution Strategy (CMA-ES) for link prediction in large scale dynamic network	CMA-ES with sixteen similarity indices

networks. On the other hand, Srilatha and Manjula (2018) proposed a bio inspired algorithm, Swarm firefly algorithm to enhance the performance of conventional similarity approaches for link prediction. The proposed method outperformed Jaccard Indices in terms of precision. In future work, it will be interesting to extend the proposed metaheuristics approaches with more network information and features.

b) Matrix Factorization

Matrix Factorization algorithm is a class of Collaborative Filtering algorithms used in Recommender system. Matrix Factorization solves the link prediction problem as a matrix representation and employs latent feature factors exist for each object. Matrix Factorization makes predictions by taking the product of two lower dimensional matrices. (Zhifeng Wu and Chen, 2016) Shows that the matrix factorization technique has the potential to solve link prediction; yet it is unstable due to noise, bias, and variance. Therefore, they include bagging technique to increase the stability of their datasets and subsequently improve prediction performance (Ahmed et al., 2018). Recently proposed a new efficient solution that used non-negative Matrix Factorization technique for link prediction problem in dynamic networks. The technique expressed implicit features, in other words, the latent features that they learned from

Table 6
Matrix Factorization for link prediction in social networks.

Year	Reference	Objectives	Approaches
2018	Ahmed et al., 2018	Propose Non-negative Matrix Factorization methods in dynamic networks and demonstrate better performance of NMF weighted representation	Non-NMF algorithm with temporal information
2017	Wang et al., 2017	Propose a kernel framework based on the non-negative matrix factorization to solve structural link prediction problems in complex networks	Kernel framework based on Matrix Factorization
2017	Wu et al., 2016	Treat link prediction problem as a collaborative filtering and introduce matrix factorization with bagging technique to improve the algorithm performance	Matrix factorization with Bagging technique

the dynamic and topological graph structure. It also expressed network dynamics efficiently and made better predictions than other similarity-based methods. Table 6 summarizes previous link prediction researches that were based on Matrix Factorization.

c) Machine Learning

Machine Learning algorithm is one of the major used strategies in performing link prediction analysis. Machine learning strategy produces high prediction performance with less computational complexity. At present, researchers utilize different machine learning techniques which include supervised learning (Fu et al., 2018), unsupervised learning (Muniz et al., 2018) and deep learning (Li et al., 2018). In terms of performance, algorithmic methods often perform better than similarity approaches. However, in the case of large networks, the network formation mechanism is more complicated, leading to high computational complexity and time. According to Wang et al. (2017), high computational complexity could be significantly reduced through parallel computing. As social networks keep evolving with more and more advanced features, there is a need for a way to adapt an efficient algorithmic method for link prediction. Table 7 shows more researches that utilized Machine Learning approach.

3.4. Hybrid

Hybrid approaches are a combination of one or more methods and extra components, attached to a framework or model that subsequently serves a better performance result to predict future links. Hybrid approach is, for instance, when a similarity approach is combined with an algorithmic approach or vice versa. Previously, Bliss et al. (2014)

Table 7
Machine learning for link prediction in social networks.

Year	Reference	Objectives	Approaches
2018	Li et al., 2018	Propose deep learning method which significantly reduces the computational complexity and increases scalability to the algorithm to large networks	Deep learning Temporal Restricted Boltzmann and gradient Boosting Decision Tree
2018	Li et al., 2018	Propose a Deep Dynamic Network Embedding for link prediction task specifically to model evolving pattern of each node	Deep Dynamic Network Embedding
2018	Fu et al., 2018	Adopt a supervised learning methods for link prediction task and focus on weighted network	Supervised learning method, RF, GBDT and SVM
2017	Mohan et al., 2017	Develop a parallel algorithm using Bulk Synchronous Parallel programming model to predict missing links and rank the newly predicted links	Bulk Synchronous Parallel programming models
2017	Asil and Gurgen, 2017	Improve link prediction process with supervised algorithms and fuzzy rule in weighted networks	Supervised algorithm and fuzzy rule based
2016	Zhang et al., 2016	Improve link prediction task by introducing novel deep learning framework using deep neural networks (DNNs)	SPEAK topology features and Deep neural network
2016	Chen et al., 2016	Propose a new link prediction algorithm based on AUC to increase link prediction efficiency in high dimensional and complex networks	AUC_LP algorithm

Table 8
Hybrid approaches for link prediction in social networks.

Year	Reference	Objectives	Combination	Approaches
2018	Wang et al., 2018	Develop fusion models which fuse the adjacent matrix with symmetric and asymmetric topological metrics together in one unified probability matrix factorization framework	Similarity and Probabilistic	Symmetric and asymmetric topological metrics in Probability Matrix Factorization
2018	Srilatha and Manjula, 2018	Propose firefly based link prediction algorithm to improve performance of similarity approach	Similarity and Algorithmic	Similarity index with Swarm firefly algorithm
2018	Muniz et al., 2018	Improve the performance of link prediction method by combining contextual and temporal information with topological data in weight computation	Similarity and Algorithmic	Weighted similarity indices in unsupervised link prediction with temporal, contextual and topological attributes
2018	Aghabozorgi and Khayyambashi, 2018	Handle imbalance social network data for link prediction problem with supervised learning framework	Similarity and Algorithmic	Local similarity index and supervised learning framework (GBM and LDA)
2017	Qiu et al., 2017	Propose an innovative and efficient link prediction based on spectral analysis that use the properties of edges directly	Similarity and Algorithmic	Spectral analysis link prediction in Artificial Neural Network
2016	Ozcan and Oguducu, 2016	Propose a novel link prediction method based on neural network for evolving networks to handle complex non-linear temporal patterns	Similarity and Algorithmic	Quasi local similarity index with NARX neural network
2015	Coskun and Koyuturk, 2015	Propose two dimensionality reduction techniques in a similarity based approach to handle high dimensionality of large social networks	Similarity and Algorithmic	Global similarity index with dimensionality reduction algorithm
2015	Deylami and Asadpour, 2015	Propose an efficient model using similarity based link prediction while considering the community information using the hierarchical community detection algorithms	Similarity and Algorithmic	Hierarchical community detection algorithm, Infomap and a similarity index, Adamic Adar
2014	Bliss et al., 2014	Propose an evolved predictor consisting of sixteen similarity indices and apply Covariance Matrix Adaptation Evolution Strategy (CMA-ES) for link prediction in large scale dynamic network	Similarity and Algorithmic	CMA-ES with sixteen similarity indices

proposed a hybrid solution to improve the prediction performance in large dynamic social networks by combining multiple similarity indices with an operation that needed an evolutionary algorithm; a meta-heuristics algorithm. Similarity approaches may be computationally simple and produce high prediction performance. However, when it comes to applying to large and complex social networks, similarity approaches degrade in performance. Hence, these shortcomings call for hybrid solution due to the efficiency of the algorithmic approach itself in large and complex social networks. In accordance to that, Deylami and Asadpour (2015) proposed an algorithm; a combination of a graph community detection algorithm with a similarity index, Adamic Adar, which improved the precision of link prediction.

Probability Matrix Factorization is another type of hybrid solution that combines a similarity approach and a probabilistic approach. However, the solution still bears a problem that limits its applicability and prediction accuracy of link prediction. This is due to the fact that the matrix adjacent only observed present links in the network, where the network is rich with topological information such as local and global similarity measure. Wang et al. (2018) used a Probability Matrix Factorization framework to develop fusion models of symmetric and asymmetric topological measure.

Nevertheless, most of the hybrid approaches focus on improvising the combination of similarity approaches and algorithmic approaches due to the specialty and advancement that algorithmic approaches provide. For example, parallel tasks computation, automation process, and extra prediction measures. These combinations make prediction task efficient in most of complex social networks as proven by researchers listed in Table 8.

4. Link prediction applications

This section discusses prevailing link prediction applications focusing on social network domains. Majority of previous link prediction applications implemented for prediction purpose for example, predict future link connection, predict missing and spurious link in the social networks (Liben-Nowell and Kleinberg, 2003). However, there are several new applications that serve more significant purposes than only predicting link tasks. The recent three categories of proposed link prediction applications are detection, recommendation and influence analysis. Section 2.3 explains and discusses these applications against the taxonomy. As social networks emerged, link prediction researchers began to consider new contexts for the applications which are further explained in Section

Table 9
Link detection application in social networks.

Year	Reference	Objectives	Category	Context	Approaches	Evaluation	Strength or weakness
2018	Kagan et al., 2018	Fake profiles detection in Academia, Arxiv, DBLP, Flixster, and Yelp	Anomaly Detection	Directed network, Weighted network	Unsupervised Anomaly detection algorithm, Random Forest classifier, Algorithmic link prediction	Confusion Matrix, Precision, ROC curve, AUC	It might be less effective on malicious users that had specific targets and strategies
2017	Teng et al., 2017	Anomaly pattern detection in dynamic and multi-attributed network, Twitter	Anomaly Detection	Dynamic network	Multi-view Time-series Hypersphere Learning (MTHL) algorithm	Confusion Matrix, Kappa statistic	The proposed model was parameter dependence
2005	Rattigan and Jensen, 2005	Anomalous link discovery in DBLP to improve accuracy rate of link prediction analysis	Anomaly Detection	Directed network	Anomalous Link Discovery (ALD) model, Katz measure, Similarity	ROC curve	The proposed work was evaluated on small network
2018	Cheng et al., 2018	Link community detection in artificial networks, GN and LFR networks	Community Detection	Complex network	Similarity measures and novel Community Detection Link Prediction (CDLP) algorithm	Normalized mutual information, NMI metric	The proposed work was evaluated on small network
2017	Mohan et al., 2017	Community detection in ca-GrQc (collaboration network), Facebook, Enron, Amazon, Youtube and LiveJournal	Community Detection	Complex network, Dynamic network	Parallel label propagation algorithm, similarity measure using Bulk Synchronous Parallel programming model	Accuracy, AUC measure	Link prediction in distributed system involved a lot of message generation which affected the performance of the system
2017	De Bacco et al., 2017	Community detection in Indian village social support network	Community Detection	Directed network, multilayer network	Multilayer community detection model	AUC measure	With the model ability to describe a wide variety of graph structures, it performed well on synthetic and real data
2018	Zhang and Lv, 2018	Event detection in Meetup.com datasets	Event Detection	Heterogeneous network	Personalized random walk with restart (RWR) algorithm	Accuracy, Precision, Recall and F1 score	The framework performed better on active users and groups than on inactive ones
2017	Hu et al., 2017	Event detection in communication network Vast and Enron mail network	Event Detection	Dynamic network	LinkEvent framework with SimC similarity computing algorithm and Event detection algorithm, EventD	ROC curve, AUC measure	The proposed model was parameter dependence

4.4.

4.1. Detection

Detection in social networks is a process of monitoring, identifying and discovering something hidden, underlying the network structure. Three categories of detection applications that use link prediction approaches in social networks are anomaly detection, community detection and event detection. Table 9 shows the latest trend of the application of link prediction for detection purposes.

a) Anomaly detection

Anomaly detection application works in domains that deal with security issues. For example, information that is provided by users when they register on a social network platform is not necessarily correct. This phenomenon gives opportunity to any user to create a fake account. These fake users vandalize the network and subsequently create harm to other users. To ensure the security of social networks, researchers proposed several methods to detect fake users in social networks. Among the methods are machine learning, graph-based or crowdsourcing (Adewole et al., 2017). Recently, Kagan et al. (2018) presented an efficient generic fake profiles detection algorithm that utilized link prediction measures in social networks. The algorithm proved its efficiency in revealing fake users as it produced lower false positive rate and higher AUC scores. Meanwhile, Teng et al. (2017) implemented a novel optimization algorithm for anomaly detection in dynamic and multi-attributed network such as Twitter. Anomaly detection was also applied along with link prediction analysis to improve the prediction performance by eliminating anomalies or outliers within the data and links which were different from the behavior of the typical links, as presented in Rattigan and Jensen (2005). However, the aforementioned proposed works are still ineffective when fake or anomalous users have specific strategies to evade from being detected. Also, anomaly link detection is still at a very preliminary stage of development. It requires further enhancement and examination to curb future security issues in link prediction applications.

b) Community detection

Community detection refers to the procedure of identifying groups in social networks according to its structural properties. Among the algorithms of community detection that were developed in the past are spectral optimization, graph partitioning and spectral clustering (Javed et al., 2018). To further improve community detection in social networks, Cheng et al. (2018) incorporated a similarity approach along with the community detection algorithm in order to identify missing links and remove fake links from the network. In addition, Mohan et al. (2017) proposed an advanced community detection system that utilized a parallel similarity measure algorithm to increase the scalability of the algorithms of link prediction for community detection. In general, link prediction should be explored to detect more accurate community structures as both link prediction and community detection are mutually beneficial.

c) Event detection

Event detection applications help social network users to discover the latest and popular events. The applications identify any important events that occur in current real-world situations, such as popular trend, political issues and economic crisis (Zhang and Lv, 2018). Zhang et al. (2016) introduced a personalized link prediction algorithm, Personalized Random Walk with Restart (RWR) for group-based event detection in Meetup.com network. Their work showed a high prediction performance for detecting event in three specific cities. Detection applications are needed for social network users to enjoy a safe, secure and beneficial social environment.

Table 10
Link recommendation application in social networks.

Year	Reference	Objective	Context	Approaches	Evaluation	Strength or weakness
2015	Song et al., 2015	Enhance link recommendation in signed network, Wikipedia, Slashdot and Epinions	Signed network	Linear time probabilistic models, Efficient Latent Link Recommendation (ELLR)	Generalized AUC, AUC, Recall and Mean Average Precision	The work only considered latent features and did not incorporate explicit features
2015	Garg et al., 2015	Product recommendation in Amazon based on product reviews	Dynamic network	Infomap algorithm, PageRank and Eigenvalue centrality measures	Accuracy	The proposed work was computationally complex for larger networks
2014	Barbieri et al., 2014	User recommendation in Twitter and Flickr	Directed network	Who to Follow and Why (WTFW), a stochastic topic model	Accuracy, ROC curve and stability	The model provided accurate link prediction and contextualize explanations to support the predictions
2012	Dong et al., 2012	User recommendation across different networks, Epinions, Slashdot, Wikivote, Twitter and Facebook	Heterogeneous network, Cross network	Ranking Factor Graph Model (RankFG) with network structure information	ROC curve, AUC measure	The proposed model was network dependent that needed a more generic model to suit other networks
2012	Papadimitriou et al., 2012	Friends recommendation in Epinions, Facebook, and Hi5 networks	Directed network, Signed network	FriendLink algorithm; exploits local and global similarity	Precision, Recall and AUC	The proposed algorithm outperformed the existing global-based friend recommendation algorithms in terms of time complexity
2011	Facebook and Leskovec, 2011	Link recommendation in Facebook and large collaboration networks	Directed network	Supervised Random Walks	ROC curve, AUC and Precision	The proposed approach was not limited to link recommendation only but could be applied to many other such as anomaly detection
2011	Scellato et al., 2011	Place recommendation in Gowalla network	Dynamic network, Heterogeneous network	Supervised learning prediction framework	ROC curve, AUC	The framework enabled the prediction of new social ties even for users who did not have any friendship connection yet, provided that they visited and checked-in at places

4.2. Recommendation

Recommendation is a common functionality in online social networks especially when it comes to recommend a new user to another user in the same network. A recommendation system gives users a more favorable social environment. The system helps users to find relevant information, social connection, and interest within a short time and it is not only limited to user recommendation. Recommendation systems also suggests a list of user, place, item and more, to users of what might interest them, in decreasing order of ranked list. The recommendation can be based on several factors such as user's preferences. As most social networks applied more context information such as location, time, interest, behavior and identity, researchers began to propose diverse recommendation applications. The main approaches used for recommendation system are Collaborative Filtering(CF), Memory-based, Model-based, Content-based, Graph-based and Context-Aware based (Campana and Delmastro, 2017). Then, integrating link prediction approaches were also introduced. For example, Papadimitriou et al. (2012) applied a friend recommendation system that exploited local and global links similarity computations in Epinions, Facebook and Hi5 networks. In comparison to the existing friend recommendation approaches, their proposed algorithm proved that utilizing link prediction approaches provided more accurate and faster recommendations. Dong et al. (2012) presented a similar approach to friend recommendation where it utilized a prediction function based on a probabilistic approach. While utilizing link prediction approaches, their approach implemented the recommendation application across different heterogeneous networks. Link prediction can also be applied in product recommendations. Evidently, Garg et al. (2015) fused a community detection on Amazon user and item graphs to improve product recommendation while utilizing link prediction algorithm. On the other hand, Song et al. (2015) presented a new probabilistic model for link recommendation in a signed network that considers latent features. The proposed model showed that emphasizing signed context of social networks helped to enhance recommendation performance. Table 10 lists existing recommendation applications that used link prediction technique.

4.3. Influence analysis

The emergence of social networks brings information like news, to the world at a glance, thus making social influence analysis a significant part of social networks. Influence analysis helps to understand people's social behaviors and provides a theoretical basis for decision making, public opinion and current economic stability. Various metrics are proposed to study how influence contributes to information spread (Peng et al., 2018). To understand the information influence in social networks, researchers applied link prediction concept to visualize in graph representation form. Then, they monitored the emerging users' behavior, information and news spread over social networks. Table 11 below displays the proposed applications of influence analysis in social networks using link prediction. Takahashi et al. (2011) came out with an application to detect emerging topics based on the mentioning behavior of Twitter users; where the topic was of something that they liked to discuss, comment or forward to their friends. Although the approach was conducted offline, it could detect the emergence of new topic efficiently using a probability model. Influence analysis is also beneficial in learning how the networks evolve. For example, Zhang et al. (2015) studied the phenomenon of “following” links in microblogging networks and proved that their method could effectively learn the diffusion strength in different patterns. In addition, Lu and Szymanski (2017) proposed another influence analysis application that used link prediction method proposed to predict viral events in a global scale where it produced prediction with high accuracy.

4.4. Application of link prediction in different social network contexts

Context is referred to a set of conditions or circumstances that surrounds a situation or event. In social networks, the context is distinguished based on the network graph representations, the requisites information and analytical purposes. The taxonomy illustrated in Fig. 7 shows that there exist several contexts of social networks namely, Directed network, Weighted network, Bipartite network, Cross network, Dynamic network, Heterogeneous network, and Signed network. Up to the present, researchers considered all these different contexts so as to serve different analytical purposes in link prediction. Table 12

Table 11
Link influence analysis application in Social networks.

Year	Reference	Objectives	Context	Approaches	Evaluation	Strength or weakness
2017	Lu and Szymanski, 2017	Predict viral events through influence analysis in GDELT dataset	Complex network	Stochastic propagation model, Parallel inference algorithm	Accuracy and F1 score	Small datasets used for evaluation
2015	Zhang et al., 2015	Study diffusion phenomenon of “following” links in microblogging networks, Twitter and Weibo	Dynamic network	Following Link Cascade model	Precision, Recall, F score, and AUC	Demonstrated that distinguishing the diffusion effects in different triadic structures could effectively activate followers based on the diffusion process
2011	Takahashi et al., 2011	Discover topic emergence in Twitter, Youtube, and NASA	Dynamic network	Probability models, with change-point analysis, Sequentially Discounting Normalized Maximum Likelihood Coding (SDNML)	Not available	The analysis was conducted offline

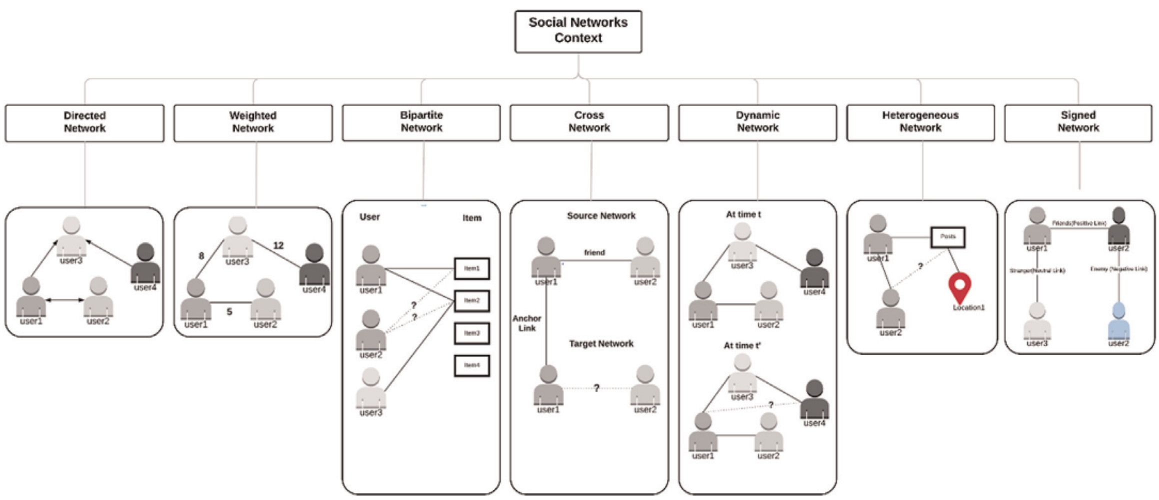

Fig. 7. Context for link prediction in social networks.

emphasizes the implementation of link prediction applications that considered each network context and discussed future work to be done on the latest issues.

a) Directed network

A social network context can be directed or undirected. In a directed network context, all edges between two nodes have directions associated with them as illustrated in Fig. 7. For instance, a ‘follower’ network, given two nodes x and y, a connection established between the two can be mutually connected such that x following y or y following x. Schall (2014) presented a triadic graph pattern into link prediction similarity approaches to predict missing links in directed social networks. The work proved that a pattern-based approach performed best in a directed network as compared to a local approach. In directed networks, modeling links could be time consuming especially in social networks with high sparsity. Bolun Chen et al. (2016) introduced a novel efficient algorithm based on Area Under Curve, AUC and treated link prediction into an optimization problem. As compared to other approaches, the algorithm achieved accurate results in a short period of time in consideration of network direction.

b) Weighted network

A social network context can also be defined as weighted or un-weighted, in a sense where researchers take the weights of edges or links into consideration (Zhu and Xia, 2016). In a weighted network context, the link between the nodes, defined with a sum of weights, denotes its degree. Liu et al. (2018) studied and analyzed the importance of considering weights in link prediction; distinguishing strong links from weak links. The heavier the weight of the link, the stronger the links or ties between nodes; strong ties increased the performance accuracy of prediction. However, not all weights of networks were effective enough to be considered in link prediction as it also depended on the network background. Bütün et al. (2018) introduced new methods to improve link prediction accuracy where the methods considered various contexts including weighted, link direction and time information. The work proved that links having heavier weight contributed more to the score of the link prediction measures.

c) Bipartite network

Social network contexts are divided into unipartite or bipartite. A social network with unipartite context has one type of node and link in the graph representation. On the other hand, bipartite network context has one type of connections and the nodes are partitioned into two different categories such that no two nodes of the same category are adjacent. All nodes in a bipartite network are connected to a node from the opposite category. Fig. 7 illustrates examples of bipartite network namely, author-paper collaborator network, group member-activities network and user-item network. A bipartite network with x users and y items can be represented as an adjacency matrix A. Benchettara et al. (2010) showed that bipartite nature of graph enhanced the prediction performance substantially by introducing a dyadic topological approach to measure the similarity of the links in a product recommendation network, Mondomix and academic collaboration recommendation network, DBLP. Li and Chen (2013) implemented link prediction in a user-item bipartite network using a graph kernel based approach to infer whether or not a user has a link with an item. Meanwhile, Gao et al. (2017) proposed a novel Potential Link Prediction, PLP algorithm to

Table 12
Link prediction applications in different social network contexts..

Year	Reference	Objectives	Context	Future work
2017	Shang et al., 2017	Propose new directional prediction method based on randomized algorithm for link prediction in directed social networks	Directed network	Extend the proposed work with deeper understanding of the role of link direction
2016	Chen et al., 2016	Propose an efficient algorithm to handle high dimensionality and sparsity of complex network for link prediction in directed social networks, USAir, Political Blogs	Directed network	Improve the proposed work by selecting appropriate network characteristics
2014	Schall, 2014	Propose a link prediction approach using triad graph patterns in directed social networks, Github, GooglePlus, and Twitter	Directed network	Extend the proposed work in a weighted network context
2018	Bütün et al., 2018	Introduce a neighbor-based link prediction methods that considers weighted, link direction and temporal information of social networks to improve link prediction accuracy	Weighted network, Directed network	Extend the work with content-based and topological features in a supervised learning algorithm
2018	Liu et al., 2018	Propose null models that consider weight of the links to analyze the effect of various weight measures for link prediction in networks	Weighted network	Use the proposed work to improve the performance of other applications in weighted complex networks
2017	Gao et al., 2017	Reduce computation time and superior quality of link prediction in Southern Women, Divorce and Scotland bipartite network	Bipartite network	Extend the proposed work in a larger dataset
2016	Zhao et al., 2016	Improve weight-based music recommendation model by considering the influence of users' similarity and users' preference	Bipartite network, Weighted network	Collect more users' preferences data to overcome data sparsity problem
2014	Fu et al., 2014	Predict users' future action by identifying and computing a better proximity measure between two nodes in a social user-item network, DBLP	Bipartite network	Utilize the proposed method for community detection in a social user-item network
2014	Zhang et al., 2014	Improve accuracy of missing and spurious link detection using bi-directional intra-similarity measures in bipartite networks for recommendation in music rating website, RYM, and author-paper network, Econophys	Bipartite network	Extend the bi-directional methods in directed networks
2013	Li and Chen, 2013	Consider recommendation task as a link prediction problem in user-item interaction graphs and present a graph kernel based approach to infer whether a user may have a link with an item	Bipartite network	Identify a systematic approach that combines graph kernels with appropriate node information to fully exploit the ability of graph kernels in recommendation
2012	Xia et al., 2012	Improve accuracy of link prediction by proposing two novel measures of structural holes in bipartite networks, IMDb director-actor network	Bipartite network	Conduct further research on weighted and directed bipartite networks and investigate structural holes methods on general graphs
2010	Benchettara et al., 2010	Show the capability of bipartite nature of graph to enhance the prediction performance substantially by introducing a dyadic topological approach to measure the links likelihood in a product recommendation network, Mondomix and Academic collaboration recommendation network, DBLP	Bipartite network	Further proposed an approach to handle weighted temporal networks
2018	Liu et al., 2018	Propose an efficient method to detect similar groups across online social networks, Twitter, LiveJournal, Flickr, Last.fm and MySpace using random walk	Cross network	Explore more advanced models and expand the feature sources to obtain better detection results
2017	Zhang et al., 2017	Propose a sparse low-rank Matrix estimation based prediction and use heterogeneous information to find similar users tend to be link across two networks, Foursquare and Twitter	Cross network	Extend the proposed work in a larger dataset
2016	Lee and Lim, 2016	Investigate how users maintain friendships' across multiple social networks through link prediction using supervised and unsupervised methods across Twitter and Instagram	Cross network	Expand the work to include larger and diverse social networks with overlapping user communities
2016	Sajadmanesh et al., 2016	Propose a meta-path based approach to solve anchor link prediction problem and extract a feature vector to predict the formation of anchor links across Twitter and Foursquare networks	Cross network	Perform dimensionality reduction to reduce features, to lower the complexity and better performance
2013	Zhang et al., 2013	Propose a supervised method, SCAN-PS to improve link prediction results for new users in the target network in cross aligned networks, Twitter and Foursquare	Cross network	Model the link prediction method in a temporal model
2019	Xu et al., 2019	Propose a distributed link prediction algorithm based on label propagation in dynamic network, Enron- e-mail and collaboration network	Dynamic network	Extend with more temporal information to improve efficiency to predict links
2018	Ahmed et al., 2018	Demonstrate better performance of similarity indices based on NMF weighted representation of link prediction in dynamic networks, Irvine, Enron, and SocioPatterns	Dynamic network	Not available
2017	Das and Das, 2017	Propose efficient method based on stochastic Markov model with time varying graph for link prediction in dynamic networks, Twitter, Facebook, and DBLP	Dynamic network	Prove the scalability of the proposed model theoretically from the order of the Markov model
2017	Ma et al., 2017	Improve accuracy of link prediction in temporal networks using NMF-based frameworks in DBLP network, cellphone and breast-cancer network	Dynamic network	Extend the application of the temporal link prediction
2016	Ahmed and Chen, 2016	Extend method for solving link prediction in temporal uncertain networks by integrating time and global topological information to obtain more accurate results in high school dynamic contact networks, SocioPatterns, Gulf, and Balkan	Dynamic network	Find an efficient approach for link prediction in uncertain dynamic networks where the occurrence probability of each edge in the network changes over time
2016	Ahmed et al., 2016	Propose a fast similarity-based algorithm to predict future links by choosing a proper number of sampled paths while	Dynamic network	Find an efficient way to obtain the optimal parameters values for a given dataset

(*continued on next page*)

Table 12 (*continued*)

Year	Reference	Objectives	Context	Future work
		reducing the computation time in temporal network, Reality mining, Haggle, DBLP, and IEE repository		
2015	Yuan et al., 2015	Propose a distributed link prediction algorithm based on clustering in dynamic networks, USAir	Dynamic network	Extend the application with directed and weighted contexts
2017	Dai et al., 2017	Propose LPMR algorithm for link prediction in multi-relational networks and use the influence between different types of relations in YouTube, Disease-Gene and Climate network	Heterogeneous network	Extend the proposed application with weighted information
2017	Gupta et al., 2017	Propose a framework HeteClass, explore the network schema to generate a set of meta-paths for classification and incorporate the knowledge of domain expert in heterogeneous social networks DBLP and Flick Fashion	Heterogeneous network	Extend proposed framework for multi-label classification problem to improve the accuracy
2017	Shakibian and Moghadam Charkari, 2017	Improve meta-path based similarity indices to predict future links in heterogeneous networks, DBLP by introducing meta-path based link entropy	Heterogeneous network	Validate the efficiency of the proposed work in a high sparsity and noisy networks
2016	Shakibian et al., 2016	Propose a multilayered approach in a multilayered heterogeneous network, DBLP by exploring the network layers with different set of meta-paths	Heterogeneous network	Extend the framework in weighted heterogeneous networks
2011	Rossetti et al., 2011	Propose scalable predictors with structural measures to solve multidimensional link prediction problem in DBLP and IMDb	Heterogeneous network, Dynamic network	Not available
2018	Li et al., 2018	Propose novel Framework of Integrating both Latent features and Explicit features (FILE), to improve link prediction performance in Epinions, Slashdot, Wikipedia and Bitcoins networks and deal with no-relation status	Signed network	Explore more explicit features for the proposed work to enhance performance and further test the effectiveness using field experiments
2018	Gu et al., 2018	Propose novel algorithm based on latent space mapping in Epinions, Slashdot and Wikipedia networks to achieve higher quality results of link prediction	Signed network	Not available
2015	Song et al., 2015	Propose Linear time probabilistic models, Efficient Latent Link Recommendation (ELLR) for link recommendation in signed network, Wikipedia, Slashdot and Epinions	Signed network	Extend the work with investigate side information of users/items
2014	Papaoikonomou et al., 2014	Propose an investigating ways using frequent subgraph between two related users to estimate link signs with high accuracy for friends recommendation in Epinions, Slashdot and Wikipedia	Signed network	Further the work into current social networks that have not support signed links like Facebook or Twitter
2010	Leskovec et al., 2010	Improve performance of edge sign prediction problem using machine learning framework in Epinions, Slashdot and Wikipedia	Signed network	Further strengthen the connections between local structure and global structure for signed links

perform link prediction of superior quality and reduce computation time in a bipartite network. However, there is a need to consider this bipartite social network context in a larger dataset in order to implement the work in real social networks.

d) Cross network

Based on the reviews presented in section 3 and 4, the majority of the link prediction researches considered single social network context only, which is rather unrealistic. As a matter of fact, people tend to use multiple social networks respectively so as to benefit from each of the social network services. For instance, people use Facebook to socialize with friends, Twitter to write a post or read the latest news and Instagram to share photos of themselves. Hence, researchers eventually came out with cross social network context, where link prediction applications in cross network have a source and target network, with nodes represent a similar user from each network while a link that connects between the two-node sets is an anchor link. Different social networks serve different purposes to their users. Therefore, link prediction researchers can gain more information on and preferences of a user who signs up to multiple social networks. For example, Zhang et al. (2013) proposed an algorithmic approach of link prediction to handle the cold start problem, where there was a lack of information on a new user of a particular social network due to trivial interactions made with other users. Meanwhile, Lee and Lim (2016) applied link prediction across two different social networks, Twitter and Instagram, to study how their users maintained their friendships across multiple social networks. On the other hand, Liu et al. (2018) proposed a method to detect similar groups that existed across multiple social networks such as Twitter, LiveJournal, Flickr, Last. fm and MySpace. The proposed works can be extended by exploring various feature sources to give even more benefits to all communities and organizations worldwide.

e) Dynamic network

A dynamic network context of social networks is associated with time information. Initially, researchers proposed link prediction in a static context where link connections between nodes did not change over time. However, social network interactions are often dynamic, thus network topology changes where nodes and link formation may appear and disappear over time. Researchers now have to consider the recurring-link situation in social networks as a time-varying graph analysis. In time-varying graph, network states are represented over a time interval. Link prediction in a dynamic context of social network includes a well-defined time interval into the applications. On the other hand, exploiting the network temporal information into a link prediction approach shows higher quality prediction results, in comparison to that of a network without temporal information. For instance, Ahmed et al. (2016) presented a fast similarity-based algorithm to predict future links, where their work combined snapshots of temporal network information into a weighted graph. The temporal information gave greater importance and higher quality prediction results with optimal time complexity. Later, Ahmed and Chen (2016) proposed an extended algorithm to improve potential link prediction in uncertain dynamic social networks. The algorithm performed random walk in order to compute similarity scores between the nodes. By casting a probabilistic approach, stochastic Markov Model with temporal information, Das and Das (2017) proved that the model also achieved great prediction accuracy in dynamic social networks namely Twitter, Facebook and DBLP. They also suggested further researches to prove the scalability of the Markov model. Ahmed et al. (2018) demonstrated that similarity indices based on non Matrix-Factorization for link prediction performed better in dynamic social networks rather than in static representation networks. Finally, the latest research by Xu et al. (2019) attempted to cater not only the dynamic properties of social networks but also the scalability of the link prediction algorithm in large networks. They implemented algorithm based on label propagation in a distributed environment, Spark. However, there remains a need to extend their work such that by adding more temporal information.

f) Heterogeneous network

A social network is considered heterogeneous when it involves different kinds of nodes and links; where each link is semantically different as illustrated in Fig. 7. The types of nodes include users, posts and locations while the links among the nodes include social interaction links, location tags and photo tags. Several meta-path based similarity indices namely, PathSim (Sun et al., 2011), HeteSim (Shi et al., 2014) and random walk are used to solve link prediction in heterogeneous social networks. The concepts of meta-paths guide network mining and help in understanding the semantic meaning of the objects and relations in the network. Shakibian et al. (2016) explored heterogeneous network layers to have a different set of meta-paths for efficient link prediction in multilayered heterogeneous networks, DBLP. Shakibian and Moghadam Charkari (2017) later improved the accuracy of link prediction in heterogeneous networks by employing link path entropy. Then, Gupta et al. (2017) proposed HeteClass framework; a meta-path based approach to exploit multiple path dependencies of link prediction in heterogeneous networks, DBLP and Flick Fashion. Meanwhile, Dai et al. (2017) used influence vectors between different relations and a non-negative matrix factorization method to predict links and achieved higher-quality prediction results than other similar algorithms.

g) Signed network

Social networks in the context of signed network define the relationship between two nodes labeled either as positive or negative link. Traditional link prediction in unsigned networks considers all connections as positive links. Be that as it may, social interaction involves both positive and negative relationships such that people form links to distinguish friends from foes and vice versa. Positive link indicates a relationship as trustworthy while negative link indicates a relationship as untrustworthy. Leskovec et al. (2010) casted link prediction problem in signed social networks to determine the signs of links. They proved the employment of both positive and negative links as useful as the result showed a significant improvement in overall prediction performance of link existence. Link prediction in signed social networks is crucial in many real-world applications, especially recommendation systems. For example, Papaoikonomou et al. (2014) focused on investigating ways of using frequent subgraph between two related users to estimate the link type with high accuracy and eventually proved their work to be beneficial in friend recommendation system. Similarly, Gu et al. (2018) proposed a novel algorithm based on latent space mapping which significantly reduces the time complexity for link prediction in signed social networks. However, in reality, besides positive and negative relationship, there exists another kind of social status that is, no-relation which refers to strangers or frenemies. In consideration of that, Li et al. (2018) proposed a new framework that integrated both latent and explicit features named FILE, to handle the no-relation situation and improve link prediction performance in signed social networks.

5. Open issues and future trends

Numerous studies addressed the strength and weaknesses of link prediction applications in social networks. However, as social networks continue to advance and become more complex, the link prediction tasks become more challenging. Several existing issues regarding link prediction in social networks are yet to be adequately discussed. This section presents the open issues of link prediction in social networks and future trends that are recommended by previous researchers.

5.1. Dataset

Social networks involve millions of active users that generate large social network data (Kemp, 2018; Razak et al., 2020). However, the veracity of data in social networks leads to issues in the data preparation step for example, incomplete or missing data (Aminzadeh et al., 2015; Liaqat et al., 2017). This phenomenon resulted in noise and unreliable dataset that subsequently affect prediction link performance in social networks (Liu et al., 2013). To address this issue, it is recommended to include a proper filtering and preprocessing step to obtain a high quality social network data to predict link.

The next issue is the imbalance of datasets which often occurs in link prediction that uses machine learning classifiers. The classification in machine learning model involves the procedure to divide the datasets into a training set and a test set. However, high sparsity of social networks may lead to an imbalance of training datasets provided for classification and eventually mislead the prediction task (Moradabadi and Meybodi, 2018). Hence, we recommend future researches to concentrate on modifying both training and test sets to fit the standard machine learning classifier.

The final issue regarding dataset for link prediction is insufficient dataset provided for evaluation of the proposed link prediction application. Evidently, several works overlooked the size of dataset (Zhao et al., 2016). When a proposed application is tested on a small dataset instead of real social network applications, it leads to a false evaluation issue. Therefore, to overcome this issue, we call upon link prediction researchers to document their data collection process more thoroughly, so it can be easily replicated by other researchers. Moreover, the majority of the researches reviewed in this paper only used links of a text data type when in fact, social networks have other types of data such as images, videos and sounds. Thus, we suggest to consider these various types of data in link prediction applications, in the future.

5.2. Network

Researchers began to consider the application of predicting links across different social networks. For example, Dong et al. (2012) described a link prediction application across 12 pairs of networks where each pair had a source and a target network. However, one of the open issues with this application was that the transformation of information between networks depended highly on a specific source and a target network. Therefore, there is a need in the future to further generalize the link prediction application in order for it to be universally accepted in multiple social networks Liu et al. (2018). Contexts of social networks such as directed, weighted and temporal contexts constitute the base and the nature of social networks. It is recommended for researchers to consider these contexts to ensure the efficiency of a link prediction application in real social networks.

5.3. Model parameter sensitivity

A probabilistic model approach for link prediction has specific parameters that the model user should specify with value each time the model runs. However, the open issue in using model approaches is how to obtain the optimal model parameters to result in an efficient link prediction. Zhang et al. (2017) proposed link prediction across two social networks, built based on a sparse and low-rank matrix model with a social structure, and incorporated the attribute information from a target network to a source network. However, assigning too large of weights of parameters of the attribute information made the model to be over fitted and later affected its performance. Hence, there remains a need to find a further efficient method of obtaining optimal parameter for the model.

5.4. Computation

Link prediction approaches in social networks, ranging from similarity to hybrid approaches, have different computation complexity which also depend on the size of the networks itself. Recently, Mohan et al. (2017) proposed an efficient and scalable method to handle the growing size of the networks. However, using a Synchronous Parallel programming model involved a lot of message generation and transfer which led to high computational complexity of the proposed method. Although the work proved the efficiency of implementing link prediction in a scalable model, the growth and sparsity of social network data posed a discrete challenge. Therefore, there is an urgency for researchers to consider steps in reducing the complexity, either time, cost or computation wise.

Real world social networks evolve over time and the link occurrence in such networks is not fixed and does not change accordingly. The dynamic nature of social network data makes its collection and preparation complicated and time consuming. For instance, the process of predicting future links requires the history of the links to be taken into account. This operation causes a massive amount of computation time and also memory space especially for a large network (Ahmed and Chen, 2016). It is recommended that such time computation issues to be tackled by utilizing distributed platforms like Hadoop or Apache Spark.

6. Conclusion

This paper presents a comprehensive and up-to-date literature review on link prediction analysis in social networks. To provide a general understanding of link prediction, the paper presents a systematic link prediction process execution in social networks beginning from data collection, preprocessing and prediction until evaluation. The paper also introduces various state-of-the-art link prediction approaches proposed in earlier researches. Unlike previous reviews on link prediction, this paper analyses and addresses how researchers combine link prediction as a base method to perform other social network analysis such as recommender systems, community detection, anomaly detection and influence analysis. Furthermore, the paper highlights the open issues and future research directions of link prediction in social networks. The findings reveal that link prediction approaches and applications in social networks to be quite challenging. This is due to the fact that social networks today have up to billions of active users which causes conventional link prediction techniques to be underperformed. However, there are still rooms for improvement on the existing link prediction analysis in social networks for example, introducing a more scalable approach that is capable to handle large social network data efficiently. A link prediction analysis approach should also consider the topology of social networks that is ever-evolving. Highly efficient methods are needed to infer time-varying information from social network data. Finally, there is a need for better approaches to perform link prediction across multiple social networks.

Conflict of interest

The author(s) declare(s) that there is no conflict of interest.

Acknowledgement

We gratefully thank the Ministry of Education, Malaysia for the support provided under Fundamental Research Grant scheme, FRGS/1/2017/ICT04/UM/02/2.

References

Adewole, K.S., Anuar, N.B., Kamsin, A., Varathan, K.D., Razak, S.A., 2017. Malicious accounts: dark of the social networks. J. Netw. Comput. Appl. 79, 41–67. https://doi.org/10.1016/J.JNCA.2016.11.030.

Aghabozorgi, F., Khayyambashi, M.R., 2018. A new similarity measure for link prediction based on local structures in social networks. Phys. Stat. Mech. Appl. 501, 12–23. https://doi.org/10.1016/J.PHYSA.2018.02.010.

Ahmed, N.M., Chen, L., 2016. An efficient algorithm for link prediction in temporal uncertain social networks. Inf. Sci. 331, 120–136. https://doi.org/10.1016/J.INS.2015.10.036.

Ahmed, N.M., Chen, L., Wang, Y., Li, B., Li, Y., Liu, W., 2016. Sampling-based algorithm for link prediction in temporal networks. Inf. Sci. 374, 1–14. https://doi.org/10.1016/J.INS.2016.09.029.

Ahmed, N.M., Chen, L., Wang, Y., Li, B., Li, Y., Liu, W., 2018. DeepEye: link prediction in dynamic networks based on non-negative matrix factorization. Big Data Mining and Analytics 1 (1), 19–33. https://doi.org/10.26599/BDMA.2017.9020002.

Aminzadeh, N., Sanaei, Z., Ab Hamid, S.H., 2015. Mobile storage augmentation in mobile cloud computing: taxonomy, approaches, and open issues. In: Simulation Modelling Practice and Theory. https://doi.org/10.1016/j.simpat.2014.05.009.

Asil, A., Gurgen, F., 2017. Supervised and fuzzy rule based link prediction in weighted co-authorship networks. In: 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, pp. 407–411. https://doi.org/10.1109/UBMK.2017.8093427.

Barbieri, N., Bonchi, F., Manco, G., 2014. Who to follow and why. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '14. ACM Press, New York, New York, USA, pp. 1266–1275. https://doi.org/10.1145/2623330.2623733.

Barnett, George A., 2011. Encyclopedia of social networks. Retrieved from. https://books.google.com.my/books/about/Encyclopedia_of_Social_Networks.html?id=HK5tNbVYs0UC&printsec=frontcover&source=kp_read_button&redir_esc=y#v=onepage&q&f=false.

Bastami, E., Mahabadi, A., Taghizadeh, E., 2018. A gravitation-based link prediction approach in social networks. In: Swarm and Evolutionary Computation. https://doi.org/10.1016/J.SWEVO.2018.03.001.

Benchettara, N., Kanawati, R., Rouveirol, C., 2010. Supervised machine learning applied to link prediction in bipartite social networks. In: 2010 International Conference on Advances in Social Networks Analysis and Mining. IEEE, pp. 326–330. https://doi.org/10.1109/ASONAM.2010.87.

Bliss, C.A., Frank, M.R., Danforth, C.M., Dodds, P.S., 2014. An evolutionary algorithm approach to link prediction in dynamic social networks. J. Comput. Sci. 5 (5), 750–764. https://doi.org/10.1016/J.JOCS.2014.01.003.

Bütün, E., Kaya, M., Alhajj, R., 2018. Extension of neighbor-based link prediction methods for directed, weighted and temporal social networks. Inf. Sci. 463–464, 152–165. https://doi.org/10.1016/J.INS.2018.06.051.

Campana, M.G., Delmastro, F., 2017. Recommender systems for online and mobile social networks: a survey. Online Soc. Netw. Media 3–4, 75–97. https://doi.org/10.1016/J.OSNEM.2017.10.005.

Chen, Bo-lun, Hua, Y., Yuan, Y., Jin, Y., Hua Yong Hua, Y., 2016. Link prediction on directed networks based on AUC optimization. IEEE Access 4. https://doi.org/10.1109/ACCESS, 2017.

Chen, Bolun, Chen, L., Li, B., 2016. A fast algorithm for predicting links to nodes of interest. Inf. Sci. 329, 552–567. https://doi.org/10.1016/J.INS.2015.09.047.

Cheng, H.-M., Ning, Y.-Z., Yin, Z., Yan, C., Liu, X., Zhang, Z.-Y., 2018. Community detection in complex networks using link prediction. Mod. Phys. Lett. B 32, 1850004. https://doi.org/10.1142/S0217984918500045, 01.

Coskun, M., Koyuturk, M., 2015. Link prediction in large networks by comparing the global view of nodes in the network. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW). IEEE, pp. 485–492. https://doi.org/10.1109/ICDMW.2015.195.

Dai, C., Chen, L., Li, B., Li, Y., 2017. Link prediction in multi-relational networks based on relational similarity. Inf. Sci. 394 (395), 198–216. https://doi.org/10.1016/J.INS.2017.02.003.

Das, S., Das, S.K., 2017. A probabilistic link prediction model in time-varying social networks. In: 2017 IEEE International Conference on Communications (ICC). IEEE, pp. 1–6. https://doi.org/10.1109/ICC.2017.7996909.

De Bacco, C., Power, E.A., Larremore, D.B., Moore, C., 2017. Community detection, link prediction, and layer interdependence in multilayer networks. Phys. Rev. 95 (4), 042317 https://doi.org/10.1103/PhysRevE.95.042317.

Deylami, H.A., Asadpour, M., 2015. Link prediction in social networks using hierarchical community detection. In: 2015 7th Conference on Information and Knowledge Technology (IKT). IEEE, pp. 1–5. https://doi.org/10.1109/IKT.2015.7288742.

Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N.V., Rao, J., Cao, H., 2012. Link prediction and recommendation across heterogeneous social networks. In: 2012 IEEE 12th International Conference on Data Mining. IEEE, pp. 181–190. https://doi.org/10.1109/ICDM.2012.140.

Ermis, B., Cemgil, A.T., 2014. A bayesian tensor factorization model via variational inference for link prediction. Retrieved from. http://arxiv.org/abs/1409.8276.

Facebook, L.B., Leskovec, J., 2011. Supervised random walks: predicting and recommending links in social networks. Retrieved from. https://www-cs.stanford.edu/~jure/pubs/linkpred-wsdm11.pdf.

Fu, C.-H., Chang, C.-S., Lee, D.-S., 2014. A proximity measure for link prediction in social user-item networks. In: Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014). IEEE, pp. 710–717. https://doi.org/10.1109/IRI.2014.7051959.

Fu, C., Zhao, M., Fan, L., Chen, X., Chen, J., Wu, Z., et al., 2018. Link weight prediction using supervised learning methods and its application to yelp layered network. IEEE Trans. Knowl. Data Eng. https://doi.org/10.1109/TKDE.2018.2801854, 1–1.

Gao, S., Denoyer, L., Gallinari, P., 2012. Probabilistic latent tensor factorization model for link pattern prediction in multi-relational networks. Retrieved from. http://arxiv.org/abs/1204.2588.

Gao, M., Chen, L., Li, B., Li, Y., Liu, W., Xu, Y., 2017. Projection-based link prediction in a bipartite network. Inf. Sci. 376, 158–171. https://doi.org/10.1016/J.INS.2016.10.015.

Garg, A., Viswanathan, S., Desai, S., 2015. CS224W Project Final Report: using community detection and link prediction to improve Amazon recommendations. Retrieved from. https://pdfs.semanticscholar.org/b77f/127e480f0b95bb5078cb9e4839f0099896f0.pdf?_ga=2.6285235.1228165773.1529993431-667948160.1522466478.

Gu, S., Chen, L., Li, B., Liu, W., Chen, B., 2018. Link Prediction on Signed Social Networks Based on Latent Space Mapping. Applied Intelligence. https://doi.org/10.1007/s10489-018-1284-1.

Gupta, M., Kumar, P., Bhasker, B., 2017. HeteClass: a Meta-path based framework for transductive classification of objects in heterogeneous information networks. Expert Syst. Appl. 68, 106–122. https://doi.org/10.1016/J.ESWA.2016.10.013.

Hou, L., Liu, K., 2017. Common neighbour structure and similarity intensity in complex networks. Phys. Lett. 381 (39), 3377–3383. https://doi.org/10.1016/J.PHYSLETA.2017.08.050.

Hu, W., Wang, H., Peng, C., Liang, H., Du, B., 2017. An event detection method for social networks based on link prediction. Inf. Syst. 71, 16–26. https://doi.org/10.1016/j.is.2017.06.003.

Javari, A., Qiu, H., Barzegaran, E., Jalili, M., Chang, K.C.-C., 2018. Statistical Link Label Modeling for Sign Prediction: Smoothing Sparsity by Joining Local and Global Information. Retrieved from. http://arxiv.org/abs/1802.06265.

Javed, M.A., Younis, M.S., Latif, S., Qadir, J., Baig, A., 2018. Community detection in networks: a multidisciplinary review. J. Netw. Comput. Appl. 108 (September 2017), 87–111. https://doi.org/10.1016/j.jnca.2018.02.011.

Kagan, D., Elovichi, Y., Fire, M., 2018. Generic anomalous vertices detection utilizing a link prediction algorithm. Soc. Netw. Anal. Min. 8 (1), 27. https://doi.org/10.1007/s13278-018-0503-4.

Kemp, S., 2018. Digital IN 2018: world'S internet users pass the 4 billion mark. Retrieved September 19, 2018, from. https://wearesocial.com/uk/blog/2018/01/global-digital-report-2018.

Kushwah, A.K.S., Manjhvar, A.K., 2016. A review on link prediction in social network. Int. J. Grid Distrib. Comp. 9 (2), 43–50. https://doi.org/10.14257/ijgdc.2016.9.2.05.

Lee, K.-W.R., Lim, E.-P., 2016. Friendship maintenance and prediction in multiple social networks. In: Proceedings of the 27th ACM Conference on Hypertext and Social Media - HT '16. ACM Press, New York, New York, USA, pp. 83–92. https://doi.org/10.1145/2914586.2914593.

Leskovec, J., Huttenlocher, D., Kleinberg, J., 2010. Predicting positive and negative links in online social networks. In: WWW '10 Proceedings of the 19th International Conference on World Wide Web, pp. 641–650. Retrieved from. http://snap.stanford.edu.

Li, Xin, Chen, H., 2013. Recommendation as link prediction in bipartite graphs: a graph kernel-based machine learning approach. Decis. Support Syst. 54 (2), 880–890. https://doi.org/10.1016/J.DSS.2012.09.019.

Li, T., Wang, B., Jiang, Y., Zhang, Y., Yan, Y., 2018. Restricted Boltzmann machine-based approaches for link prediction in dynamic networks. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2840054, 1–1.

Li, T., Zhang, J., Yu, P.S., Zhang, Y., Yan, Y., 2018. Deep dynamic network embedding for link prediction. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2839770, 1–1.

Li, Xiaoming, Fang, H., Zhang, J., 2018. FILE: a novel framework for predicting social status in signed networks. In: AAAI Conference on Artificial Intelligence. Retrieved from. www.aaai.org.

Liaqat, M., Chang, V., Gani, A., Hamid, S.H.A., Toseef, M., Shoaib, U., Ali, R.L., 2017. Federated cloud resource management: review and discussion. J. Netw. Comput. Appl. https://doi.org/10.1016/j.jnca.2016.10.008.

Liben-Nowell, D., Kleinberg, J., 2003. The link prediction problem for social networks. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management - CIKM '03. ACM Press, New York, New York, USA, p. 556. https://doi.org/10.1145/956863.956972.

Liu, Z., He, J.-L., Kapoor, K., Srivastava, J., 2013. Correlations between community structure and link formation in complex networks. PloS One 8 (9), e72908. https://doi.org/10.1371/journal.pone.0072908.

Liu, Y., Zhao, C., Wang, X., Huang, Q., Zhang, X., Yi, D., 2016. The degree-related clustering coefficient and its application to link prediction. Phys. Stat. Mech. Appl. 454, 24–33. https://doi.org/10.1016/J.PHYSA.2016.02.014.

Liu, B., Xu, S., Li, T., Xiao, J., Xu, X.-K., 2018. Quantifying the effects of topology and weight for link prediction in weighted complex networks. Entropy 20 (5), 363. https://doi.org/10.3390/e20050363.

Liu, X., Shen, C., Guan, X., Zhou, Y., 2018. We know who you are: discovering similar groups across multiple social networks. IEEE Trans. Syst. Man Cybern.: Systems 1–12. https://doi.org/10.1109/TSMC.2018.2826555.

Lu, X., Szymanski, B., 2017. Predicting viral news events in online media. In: 2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). IEEE, pp. 1447–1456. https://doi.org/10.1109/IPDPSW.2017.82.

Lü, L., Zhou, T., 2011. Link prediction in complex networks: a survey. Phys. Stat. Mech. Appl. 390 (6), 1150–1170. https://doi.org/10.1016/J.PHYSA.2010.11.027.

Ma, X., Sun, P., Qin, G., 2017. Nonnegative matrix factorization algorithms for link prediction in temporal networks using graph communicability. Pattern Recogn. 71, 361–374. https://doi.org/10.1016/J.PATCOG.2017.06.025.

Martínez, V., Berzal, F., Cubero, J.-C., 2016. A survey of link prediction in complex networks. ACM Comput. Surv. 49 (4), 1–33. https://doi.org/10.1145/3012704.

Mohan, A., Venkatesan, R., Pramod, K.V., 2017. A scalable method for link prediction in large real world networks. J. Parallel Distr. Comput. 109, 89–101. https://doi.org/10.1016/J.JPDC.2017.05.009.

Moradabadi, B., Meybodi, M.R., 2018. Link prediction in weighted social networks using learning automata. Eng. Appl. Artif. Intell. 70, 16–24. https://doi.org/10.1016/J.ENGAPPAI.2017.12.006.

Muniz, C.P., Goldschmidt, R., Choren, R., 2018. Combining contextual, temporal and topological information for unsupervised link prediction in social networks. Knowl. Base Syst. https://doi.org/10.1016/J.KNOSYS.2018.05.027.

Ozcan, A., Oguducu, S.G., 2016. Temporal link prediction using time series of quasi-local node similarity measures. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, pp. 381–386. https://doi.org/10.1109/ICMLA.2016.0068.

Papadimitriou, A., Symeonidis, P., Manolopoulos, Y., 2012. Fast and accurate link prediction in social networking systems. J. Syst. Software 85 (9), 2119–2132. https://doi.org/10.1016/J.JSS.2012.04.019.

Papaoikonomou, A., Kardara, M., Tserpes, K., Varvarigou, T.A., 2014. Predicting edge signs in social networks using frequent subgraph discovery. IEEE Intern. Comput. 18 (5), 36–43. https://doi.org/10.1109/MIC.2014.82.

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., Jia, W., 2018. Influence analysis in social networks: a survey. J. Netw. Comput. Appl. 106, 17–32. https://doi.org/10.1016/J.JNCA.2018.01.005.

Qian, F., Gao, Y., Zhao, S., Tang, J., Zhang, Y., 2017. Combining topological properties and strong ties for link prediction. Tsinghua Sci. Technol. 22 (6), 595–608. https://doi.org/10.23919/TST.2017.8195343.

Qiu, Y., Hu, R., Yan, R., Yao, Y., Jiao, M., 2017. The new link prediction methods based on spectral analysis. In: 2017 13th International Conference on Semantics, Knowledge and Grids (SKG). IEEE, pp. 106–112. https://doi.org/10.1109/SKG.2017.00025.

Rattigan, M.J., Jensen, D., 2005. The case for anomalous link discovery. ACM SIGKDD Explor. Newslett. 7 (2), 41–47. https://doi.org/10.1145/1117454.1117460.

Razak, C.S.A., Zulkarnain, M.A., Hamid, S.H.A., Anuar, N.B., Jali, M.Z., Meon, H., 2020. Tweep: a system development to detect depression in twitter posts. In: Lecture Notes in Electrical Engineering. https://doi.org/10.1007/978-981-15-0058-9_52.

Rossetti, G., Berlingerio, M., Giannotti, F., 2011. Scalable link prediction on multidimensional networks. 2011 IEEE 11th International Conference on Data Mining Workshops. IEEE, pp. 979–986. https://doi.org/10.1109/ICDMW.2011.150.

Sajadmanesh, S., Rabiee, H.R., Khodadadi, A., 2016. Predicting anchor links between heterogeneous social networks. https://doi.org/10.1109/ASONAM.2016.7752228.

Salakhutdinov, R., Mnih, A., 2008. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In: Proceedings of the 25th International Conference on Machine Learning - ICML '08. ACM Press, New York, New York, USA, pp. 880–887. https://doi.org/10.1145/1390156.1390267.

Scellato, S., Noulas, A., Mascolo, C., 2011. Exploiting place features in link prediction on location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '11. ACM Press, New York, New York, USA, p. 1046. https://doi.org/10.1145/2020408.2020575.

Schall, D., 2014. Link prediction in directed social networks. Soc. Netw. Anal. Min. 4 (1), 157. https://doi.org/10.1007/s13278-014-0157-9.

Shakibian, H., Moghadam Charkari, N., 2017. Mutual information model for link prediction in heterogeneous complex networks. Sci. Rep. 7, 44981. https://doi.org/10.1038/srep44981.

Shakibian, H., Charkari, N.M., Jalili, S., 2016. A multilayered approach for link prediction in heterogeneous complex networks. J. Comput. Sci. 17, 73–82. https://doi.org/10.1016/J.JOCS.2016.10.001.

Shang, K., Small, M., Yan, W., 2017. Link direction for link prediction. Phys. Stat. Mech. Appl. 469, 767–776. https://doi.org/10.1016/J.PHYSA.2016.11.129.

Shi, C., Kong, X., Huang, Y., Yu, P., Wu, B., 2014. HeteSim: a general framework for relevance measure in heterogeneous networks. *Knowledge and data engineering.* IEEE Trans. 26 (10) https://doi.org/10.1038/nature09973.

Song, D., Meyer, D.A., Tao, D., 2015. Efficient latent link recommendation in signed networks. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15. ACM Press, New York, New York, USA, pp. 1105–1114. https://doi.org/10.1145/2783258.2783358.

Srilatha, P., Manjula, R., 2018. Structural similarity based link prediction in social networks using firefly algorithm. In: Proceedings of the 2017 International Conference on Smart Technology for Smart Nation, SmartTechCon 2017. https://doi.org/10.1109/SmartTechCon.2017.8358434.

Sun, Y., Han, J., Yan, X., Yu, P.S., Wu, T., 2011. PathSim: meta path-based top-K similarity search in heterogeneous information networks. Proc. VLDB Endowm. 4 (11), 992–1003. https://doi.org/10.2200/S00433ED1V01Y201207DMK005.

Sun, Q., Hu, R., Yang, Z., Yao, Y., Yang, F., 2017. An improved link prediction algorithm based on degrees and similarities of nodes. In: 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). IEEE, pp. 13–18. https://doi.org/10.1109/ICIS.2017.7959962.

Takahashi, T., Tomioka, R., Yamanishi, K., 2011. Discovering emerging topics in social streams via link anomaly detection. Retrieved from. http://arxiv.org/abs/1110.2899.

Teng, X., Lin, Y.-R., Wen, X., 2017. Anomaly detection in dynamic networks using multi-view time-series hypersphere learning. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management - CIKM '17. ACM Press, New York, New York, USA, pp. 827–836. https://doi.org/10.1145/3132847.3132964.

Wang, P., Xu, B., Wu, Y., Zhou, X., 2015. Link prediction in social networks: the state-of-the-art. Sci. China Inf. Sci. 58 (1), 1–38. https://doi.org/10.1007/s11432-014-5237-y.
Wang, J., Ma, Y., Liu, M., Yuan, H., Shen, W., Li, L., 2017. A vertex similarity index using community information to improve link prediction accuracy. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, pp. 158–163. https://doi.org/10.1109/SMC.2017.8122595.
Wang, W., Feng, Y., Jiao, P., Yu, W., 2017. Kernel framework based on non-negative matrix factorization for networks reconstruction and link prediction. Knowl. Base Syst. 137, 104–114. https://doi.org/10.1016/J.KNOSYS.2017.09.020.
Wang, Z., Liang, J., Li, R., 2018. A fusion probability matrix factorization framework for link prediction. Knowl. Base Syst. 159, 72–85. https://doi.org/10.1016/j.knosys.2018.06.005.
Wu, Zhifeng, Chen, Y., 2016. Link Prediction Using Matrix Factorization with Bagging*. IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS).
Wu, S., Sun, J., Tang, J., 2013. Patent partner recommendation in enterprise social networks. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining - WSDM '13. ACM Press, New York, New York, USA, p. 43. https://doi.org/10.1145/2433396.2433404.
Wu, Zhihao, Lin, Y., Wang, J., Gregory, S., 2016. Link prediction with node clustering coefficient. Phys. Stat. Mech. Appl. 452, 1–8. https://doi.org/10.1016/J.PHYSA.2016.01.038.
Wu, Zhihao, Lin, Y., Zhao, Y., Yan, H., 2018. Improving local clustering based top-L link prediction methods via asymmetric link clustering information. Phys. Stat. Mech. Appl. 492, 1859–1874. https://doi.org/10.1016/J.PHYSA.2017.11.103.
Xia, Shuang, Dai, BingTian, Lim, Ee-Peng, Zhang, Yong, Xing, Chunxiao, 2012. Link prediction for bipartite social networks: the role of structural holes. In: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE, pp. 153–157. https://doi.org/10.1109/ASONAM.2012.35.
Xiao, Y., Li, X., Wang, H., Xu, M., Liu, Y., 2018. 3-HBP: a three-level hidden bayesian link prediction model in social networks. IEEE Trans. Comput. Soc. Syst. 1–14. https://doi.org/10.1109/TCSS.2018.2812721.
Xu, M., Yin, Y., 2017a. A similarity index algorithm for link prediction. In: 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE, pp. 1–6. https://doi.org/10.1109/ISKE.2017.8258724.
Xu, M., Yin, Y., 2017b. A similarity index algorithm for link prediction. In: 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE). IEEE, pp. 1–6. https://doi.org/10.1109/ISKE.2017.8258724.
Xu, X., Hu, N., Li, T., Trovati, M., Palmieri, F., Kontonatsios, G., Castiglione, A., 2019. Distributed temporal link prediction algorithm based on label propagation. Future Generat. Comput. Syst. 93, 627–636. https://doi.org/10.1016/J.FUTURE.2018.10.056.
Yang, Q., Dong, E., Xie, Z., 2014. Link prediction via nonnegative matrix factorization enhanced by blocks information. In: 2014 10th International Conference on Natural Computation (ICNC). IEEE, pp. 823–827. https://doi.org/10.1109/ICNC.2014.6975944.
Yao, L., Wang, L., Pan, L., Yao, K., 2016. Link prediction based on common-neighbors for dynamic social network. Procedia Comput. Sci. 83, 82–89. https://doi.org/10.1016/J.PROCS.2016.04.102.
Yuan, H., Ma, Y., Zhang, F., Liu, M., Shen, W., 2015. A distributed link prediction algorithm based on clustering in dynamic social networks. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, pp. 1341–1345. https://doi.org/10.1109/SMC.2015.238.
Zeng, S., 2016. Link prediction based on local information considering preferential attachment. Phys. Stat. Mech. Appl. 443, 537–542. https://doi.org/10.1016/J.PHYSA.2015.10.016.
Zhang, S., Lv, Q., 2018. Hybrid EGU-based group event participation prediction in event-based social networks. Knowl. Base Syst. 143, 19–29. https://doi.org/10.1016/J.KNOSYS.2017.12.002.
Zhang, Jiawei, Kong, X., Yu, P.S., 2013. Predicting social links for new users across aligned heterogeneous social networks. Retrieved from. http://arxiv.org/abs/1310.3492.
Zhang, P., Zeng, A., Fan, Y., 2014. Identifying missing and spurious connections via the bi-directional diffusion on bipartite networks. https://doi.org/10.1016/j.physleta.2014.06.011.
Zhang, Jing, Fang, Z., Chen, W., Tang, J., 2015. Diffusion of "following" links in microblogging networks. IEEE Trans. Knowl. Data Eng. 27 (8), 2093–2106. https://doi.org/10.1109/TKDE.2015.2407351.
Zhang, C., Zhang, H., Yuan, D., Zhang, M., 2016. Deep learning based link prediction with social pattern and external attribute knowledge in bibliographic networks. In: 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, pp. 815–821. https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2016.170.
Zhang, Jiawei, Chen, J., Zhi, S., Chang, Y., Yu, P.S., Han, J., 2017. Link prediction across aligned networks with sparse and low rank matrix estimation. In: 2017 IEEE 33rd International Conference on Data Engineering (ICDE). IEEE, pp. 971–982. https://doi.org/10.1109/ICDE.2017.144.
Zhao, D., Zhang, L., Zhao, W., 2016. Genre-based link prediction in bipartite graph for music recommendation. Procedia Comput. Sci. 91, 959–965. https://doi.org/10.1016/J.PROCS.2016.07.121.
Zhao, H., Du, L., Buntine, W., 2017. Leveraging node attributes for incomplete relational data. Retrieved from. http://arxiv.org/abs/1706.04289.
Zhu, B., Xia, Y., 2016. Link prediction in weighted networks: a weighted mutual information model. PloS One 11 (2), e0148265. https://doi.org/10.1371/journal.pone.0148265.

Nur Nasuha Daud received Bachelor degree in Computer Science and Information Technology (Software Engineering) from University of Malaya, Malaysia. Nasuha is currently pursuing a PhD program from the same university in the Department of Software Engineering, Faculty of Computer Science and Information Technology. Her Ph.D research is in Social network analysis with specific focus on Link Prediction. Her research interests include Social network analysis, Machine learning and Big Data. She is also a member of IEEE student and IEEE Women in Engineering.

Siti Hafizah Ab Hamid received BS (Hons) in Computer Science from University of Technology, Malaysia, MS in Computer System Design from Manchester University, UK., and the PhD in Computer Science from University of Malaya, Malaysia. She is currently an Associate Professor with the Department of Software Engineering, Faculty of Computer Science & Information Technology, and University of Malaya, Malaysia. She has authored over 80 research articles in different fields, including mobile cloud computing, big data, software testing, software engineering, machine learning and IoT.

Nor Badrul Anuar received his Master of Computer Science from University of Malaya in 2003 and a PhD at the Centre for Information Security & Network Research, University of Plymouth, UK in 2012. He is currently an Associate Professor with the Faculty of Computer Science and Information Technology, University of Malaya. He has authored over 128 research articles and a number of conference papers locally and internationally. His research interests include information security, intrusion detection systems, data sciences, high-speed networks, artificial intelligence, and library information systems.

Muntadher Saadoon received BS (Hons) in Software Engineering from Al-Rafidain University College, Iraq, and MS in Computer Science from University Putra Malaysia, Malaysia. He is currently a Ph.D. student in Software Engineering, University of Malaya, Malaysia. His research interests focus on software reliability and fault tolerance in distributed computing. He has strong technical knowledge in different technologies, including virtualization, cloud computing, big data, web services, and programming languages.

Firdaus Sahran received his Bachelor's Degree in Computer Science from University of Malaya, Kuala Lumpur, Malaysia, in 2018 majoring in Computer System and Networking. He is currently a researcher and PhD candidate in Network Analytics Lab, University of Malaya, Kuala Lumpur, Malaysia. His research interests include Software-Defined Networking (SDN), network security, and fault management.