Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 11, No. 3, 2023, 487-504.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Link Prediction in Social Networks: A Bibliometric Analysis and Review of Literature (1987-2021)

Akram Pasandideh and Mohsen Jahanshahi*

Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

Article Info

Article History:
Received 10 February 2023
Revised 29 March 2023
Accepted 10 July 2023

DOI:10.22044/jadm.2023.12508.2425

Keywords:
Bibliometric, Social Networks, Social Network Analysis, Link Prediction, Bradford's Law.

**Corresponding author: mjahanshahi@iauctb.ac.ir (M. Jahanshahi).*

Abstract

Link prediction (LP) has become a hot topic in the data mining, machine learning, and deep learning community. This study aims to implement bibliometric analysis to find the current status of the LP studies and investigate it from different perspectives. The present study provides a Scopus-based bibliometric overview of the LP studies landscape since 1987 when LP studies were published for the first time. Various kinds of analysis including document, subject, and country distribution are applied. Moreover, author productivity, citation analysis, and keyword analysis are used, and Bradford's law is applied to discover the main journals in this field. Most documents were published by conferences in the field. The majority of LP documents have been published in the computer science and mathematics fields. So far, China has been at the forefront of publishing countries. In addition, the most active sources of LP publications are lecture notes in computer science including sub-series lecture notes in Artificial Intelligence (AI) and lecture notes in Bioinformatics, and IEEE Access. The keyword analysis demonstrates that while social networks had attracted attention in the early period, knowledge graphs have attracted more attention, recently. Since the LP problem has been approached recently using machine learning (ML), the current study may inform the researchers to concentrate on ML techniques. This is the first bibliometric study of "link prediction" literature, providing a broad field landscape.

1. Introduction

Link prediction (LP) is regarded as a sub-task of social network analysis (SNA) that provides us with the upcoming or implicit links of a network. The definition is composed of different components, which are presented in the following parts.

1.1 Social networks

A social network (SN) is defined as a social structure including some social actors (like persons or organizations), a series of dyadic relations, and other social communications among the actors. The SN view presents a set of approaches to analyze the entire social entities' structure and different theories describing the models seen in those structures [1]. SNs are collections of entities or groups that are interconnected such as collaboration and socialization. In SNs, nodes and edges are used to symbolize entities and their interactions, respectively [2]. Overusing has led to the problem of examining and analyzing structures and shared content that comprises SNs. SNs have turned out to be significant data sources for examining the users' features and rating personalities for different aims. Moreover, there are various research fields, like recommendation systems development for SN users [2], educational evaluations [3, 4], and sentiment [5].

1.2 Social network analysis (SNA)

The SNA notion was first suggested by Mitchell (1969). We could conduct a visual and mathematical examination of human relations via SNA that aids us to make sense of the SN, and the SNs' complicated structure analyzes the SNs' evolution and network dynamics, and notice complicated interaction patterns and common characteristics of the network. One of the major aims of SNA is to notice and recognize social relations patterns between actors and to categorize the effect (profits or limitations) of the social structure on the actors and network operations [6, 7, 12].

Three major approaches to SNA include graph representation, content mining, and semantic analysis [8]. Content mining takes into account comprehending the models and recognizing elements that lead to information distribution in online SNs. Some elements are regarded, like hashtags, URLs, sentiment examination, and sarcasm recognition. SNA's semantic analysis is conducted because of the lack of semantics support in graph representation and content analysis [8]. The use of new SNA methods has become progressively popular and has been used in various contexts, such as social mobility [9], citations study [10, 11, 13], contacts among members of abnormal groups, corporate power [14, 15], international trade exploitation [16], class structure [17, 18], expert search system, 19, 20], social recommendations [21], LP, and many others.

The SNA key tasks consist of various measures to rate nodes (or edges), recognizing SNs from social occasions, virtual marketing, community detection, the networks' incentives design, specifying hidden social hierarchy, the formation of network, and LP issue [22]. In the SNA, learning how a network evolves and how new links are formed over time is critical. This task is known as the LP problem [23].

1.3 Co-authorship networks

Co-authorship networks. Co-authorship networks are one of the most widely used academic social networks. In co-authorship networks, each node in the co-authorship network refers to an author. Edges in the co-authorship network refer to a co-authored relationship. Scholars study co-author networks from various perspectives. Scholars study collaboration behaviors according to co-author networks. Furthermore, collaboration teamwork has been found to be a new research pattern [84].

Co-authorship networks link researchers to writing joint research papers in co-authorship. They can be used for trend analysis and collaboration pattern mining. There are many real-world applications of collaboration networks to recommender systems for citing relevant papers, finding experts, etc. In particular, it is important to establish communications between researchers with similar research interests for universities and improve the quality of interdisciplinary research projects.

The network science method for predicting collaborations is based on the link prediction (LP) problem [89] applied for missing edge reconstruction. Basically, one needs to construct vector edge feature representation and then apply a standard machine learning framework for classification or regression. Link prediction models are applied in Web applications [86], dating recommendations [87], and paper recommendations for large digital libraries [88].

1.4 Co-citation networks

Co-citation is defined as two publications that are cited together in one article. Co-citation networks are constructed based on articles' citation relationships. Scholars generate co-citation networks from publications and study scholars' behaviors from co-citation networks.

Some academic social relationships may be not discovered through co-author networks but can be discovered by co-citation networks. Co-citation analysis is one of the most commonly used bibliometric analysis methods. When two publications are frequently co-cited by the other articles, the two references may have something in common. As an advanced bibliographic technique, co-citation analysis is commonly used to discover the clusters of co-citation pairs, which enables scholars to obtain new insights into research trends [84].

1.5 Link prediction

The issue of forecasting new relations within the networks is called LP [24]. LP identifies missing links (in static networks) or anticipates the future links probability (in dynamic networks) [25]. The task of LP can be classified as follows: Missing LP [26, 27], and Future LP [28, 29]. LP is a growing topic in the physical science and software engineering fields. There is a comprehensive scope of LP techniques, like similarity-oriented indices, probabilistic approaches, and dimensionality reduction methods. Learning-based techniques are divided into clustering-oriented and information-theoretic models. LP is done in various kinds of networks, for instance, directed, temporal, bipartite, and heterogeneous networks [30]. The LP concept is beneficial in a few regions of application such as automatic hyperlink creation, website hyperlink prediction on the internet, web of science domain, and friend recommendation on Facebook. LP is of crucial importance in the realm of recommender systems, too. Bioinformatics protein-protein interactions (PPI) have also been executed using LP. In insecurity concern areas, LP serves to differentiate hidden links between terrorists and their organizations [30].

Bibliometric is a useful tool for literature and research methods analysis regarding a special research area both quantitatively and qualitatively [31]. It aids in evaluating a research area's progress, recognizing the source of the most related and significant document, identifying main authors and institutions, and investigating potential research topics [32]. Many bibliometric studies have been conducted on computer science and data mining topics. These works contain research that explored specific research areas, like topic modeling [33], sentiment analysis [34], big data [35], machine learning [35], computational linguistics from an overall view [36], and deep learning research status analysis [37]. Through reviewing the literature, it is revealed no study has been conducted to systematically review LP research with large-scale bibliographic data [38, 32].

Also, Behrouzi *et al*. (2020) undertook a study on the keywords of journals and conferences to reach a complete explanation and comparison between them as the two main openings of the research. As a result, the structural features of journal and conference keyword networks are almost alike, both can be used as real resources for semantic bibliographic analysis in computer science. Zhang *et al*. (2021) proposed to use the impacts of novelty, bibliometric, and academic-network characteristics on the paper citation counts. They qualified the networking of neural models and implemented the linked neurons weighted product to assess the understanding degree for each characteristic about citations. The results confirmed that the effect of novelty, bibliometric, and academic-network-related factors on citation counts varies meaningfully among the four fields of study, including library and information science, nuclear science and technology, computer science and software engineering, and history. Also, it was revealed that the effect of various features in the novelty category on citation counts is higher than in the bibliometric and academic networks. In addition, individual characteristics in the novelty category are not always the most important factors. Given that no such research has been done on LP, this article aims to provide a bibliometric analysis of LP documents in more than the past three decades and paint a global and full picture of the patterns and dynamics of LP research from the following perspectives: distribution of subject areas, the trend of annual publication of LP articles, distribution of document types, most productive and impactful authors, most productive countries, most active institutions, most prolific sources, and most cited papers keyword analysis.

2. Data and Methods

At present, the researchers, institutions, and different organizations linked to science produce many publications such as journal articles, book chapters, and conference proceedings, among different document types, and, usually, this information is stored in bibliographical databases such as the famous Scopus, dimensions or Web of science [91]. The analytical study obtains the following main conclusions:

• First, the three databases are found to differ significantly in their journal coverage, with Web of Science having the most selective journal coverage, while Dimensions has the most exhaustive journal coverage. It is found that almost all journals indexed in Web of Science are also covered by Scopus and dimensions. Scopus indexes 66.07% more unique journals as compared to Web of Science.

• Second, the varied journal coverage of the three databases results in variation in research output volume, rank, and global share of different countries. Therefore, drawing data from different databases may produce different outcomes for any bibliometric evaluation exercises done at the level of countries.

• Third, the three databases also vary in their coverage of different subject areas. The Web of Science and Scopus have most of their coverage in life sciences, physical sciences, and technology. On the other hand, dimensions appear to have significantly better coverage of social sciences and arts and humanities [92, 93].

Scopus is a major indexing database and one of the key sources for bibliometric studies [39-43]. Therefore, we applied this database as our core data source. The data was collected from the Scopus core collection on November 1, 2021. The search statement is as follows:

TITLE-ABS-KEY ("link prediction"(: The raw bibliographic data of the papers were then saved. Moreover, some bibliometric software was applied, including Bibexcel and VOSviewer [44]. TITLE-ABS-KEY is a Scopus reserved key, and it searches in the title, abstract, and keywords of documents for the provided term. Bibexcel version

2008-08 [45] was applied to explore the most typical titles, subject categories, the most prolific authors and journals, and the most significant articles. Furthermore, to examine qualitatively, we applied SJR (Scimago Journal and Country Rank) of all sources from Scimago that are a free database to access journals and countries' bibliometric indices based on Scopus data. SJR is a journal bibliometric indicator like impact factor (IF) that influences citations according to the source journals' reputation [46, 47]. That is to say, more impactful journals' citations are not regarded as equal to those of less impactful ones. Keyword analysis, and identifying productive countries were done using VOSviewer version 1.6.7 [44]. Bradford's law was used to identify the core sources of information[1].

3. Results and Discussion

This section presents the bibliometric analysis of the results and discussion. As a result, a total of 3,955 research articles were obtained with the aforementioned retrieval strategy.

3.1 Trend of annual publication of LP articles

Figure 1 shows the annual trend of LP publications. From 1987 to 2021, the average growth rate of LP research was 33.08%. The rate of growth between 1987-2005, 2006-2010, and 2011-2021 were 31.66%, 55.13%, and 21.20%, respectively. The publications' number increased suddenly in 2019, accounting for a 43.59% growth rate (from 445 to 639 papers).

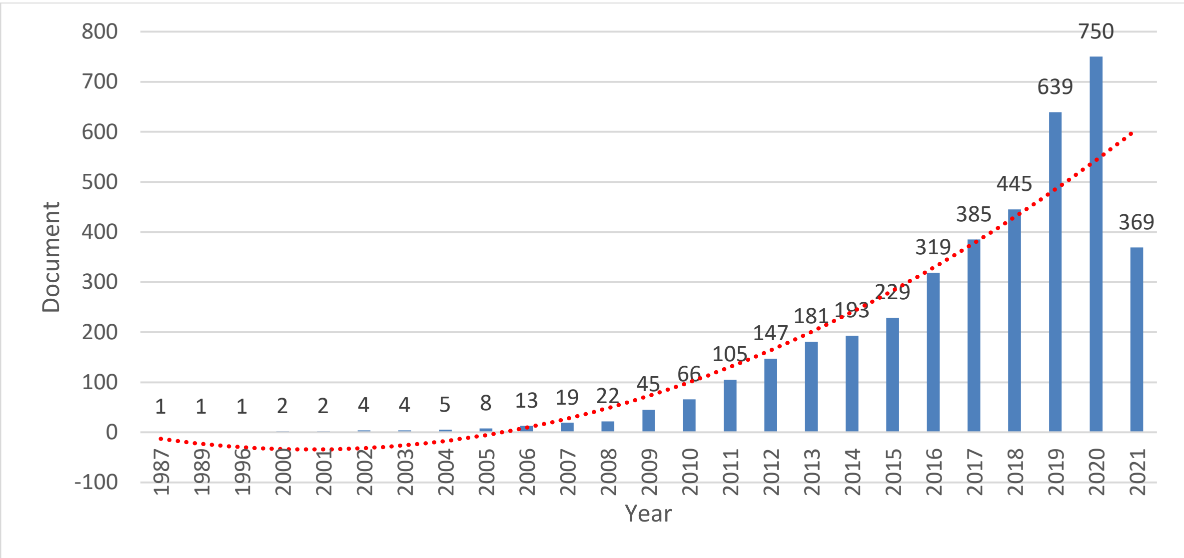


Figure 1. Annual publications of LP articles' trend. (The illustrated trend is a fitted polynomial regression).

It should also be noted that due to the increase in conference titles, there is a jump in the number of conference papers between 2018-2019, which is a growth of 445 to 639 papers.

3.2 Distribution of document types

Figure 2 illustrates the document type distribution. Many documents were published by technical conferences (2,120 documents, ~%53.60 of all documents), and it is followed by journals that published 1,572 documents (~%39.74 of all documents). Meantime, 'LP' is considered a 'computer science 'subfield, these data are in harmony with the source of the major publication in computer science, that is conference papers [33].

3.3 Distribution of subject areas

Figure 3 shows the subject areas' distribution. The majority (nearly half of the documents) of LP documents (%45.67) have been published in computer science. Meanwhile, LP documents gained popularity in mathematics (%14.99), engineering (%11.70), and decision sciences (%6.35). The results are significant because LP originally was a task of machine learning and data mining, which are their computer science subfields. On the other hand, the root of graph theory is originally from mathematics. LP is also important for social science because SN analysis is an important topic in this field [48]. Even though LP has recently emerged from computer science, it has been investigated in Physics and Astronomy [49,50], social sciences [40,51], business,

[1] rank.sid.ir/Bradford

management, and accounting [52, 53] because of its importance.

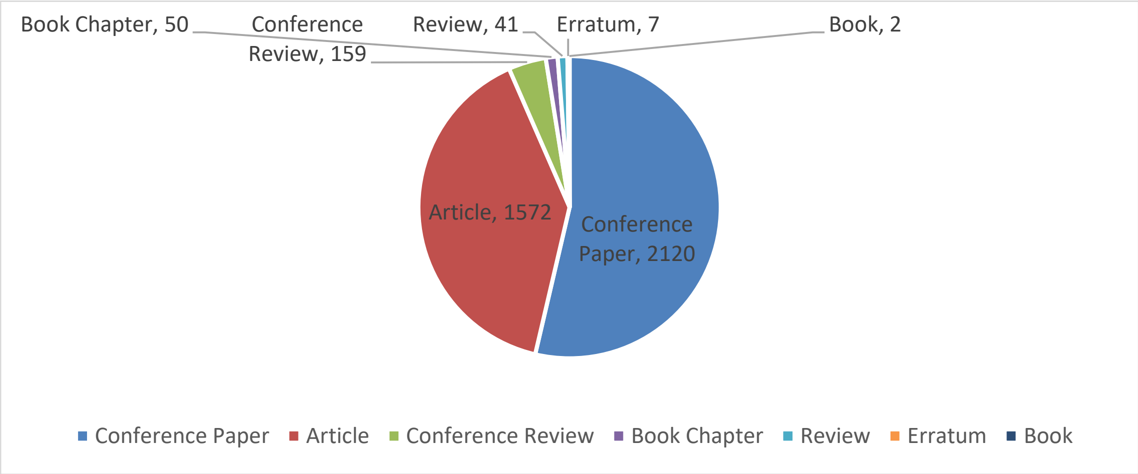


Figure 2. Document types distribution.

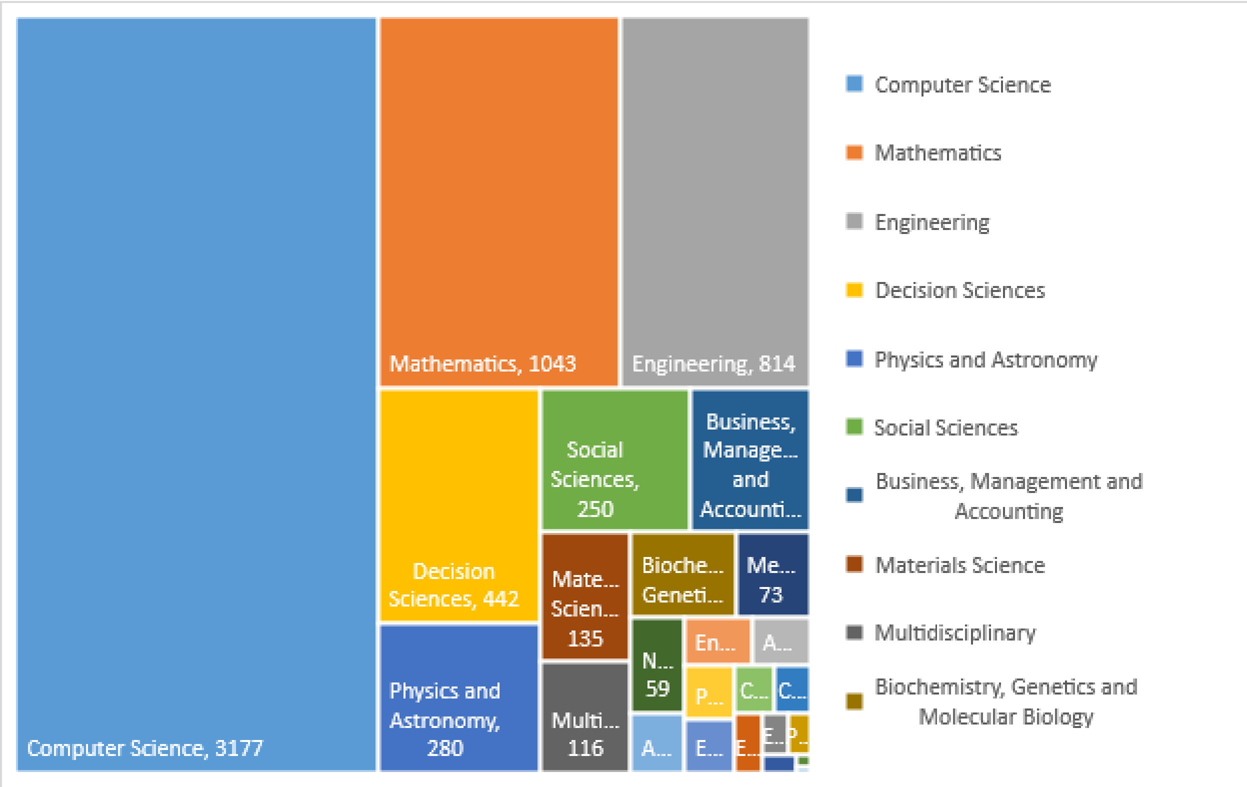


Figure 3. Subject categories with more than 100 LP documents.

3.4 Most productive countries

The most productive countries are listed in Table 1. These documents originated from 80 countries. The top 10 countries accumulatively have contributed 91.43% to all LP documents. China is the most productive country, which contributed 41.54% of documents followed by the United States, which contributed 79% to LP documents.
Other productive countries include India (6.44%), Australia (3.94%), the United Kingdom (3.41%), Germany (3.36%), Japan (2.95%), France (2.83%), Canada (2.65%), and Italy (2.52).
Figure 4 indicates the most prolific countries' collaboration map. The nodes' size in the graph illustrates the number of documents that were published by the respective countries.

Table 1. Most productive countries in LP research.

Rank	Countries	Number of documents	Percentage of 3955
1	China	1643	41.54
2	The United States	862	21.79
3	India	255	6.44
4	Australia	156	3.94
5	United Kingdom	135	3.41
6	Germany	133	3.36
7	Japan	117	2.95
8	France	112	2.83
9	Canada	105	2.65
10	Italy	100	2.52
sum	-	3618	91.43

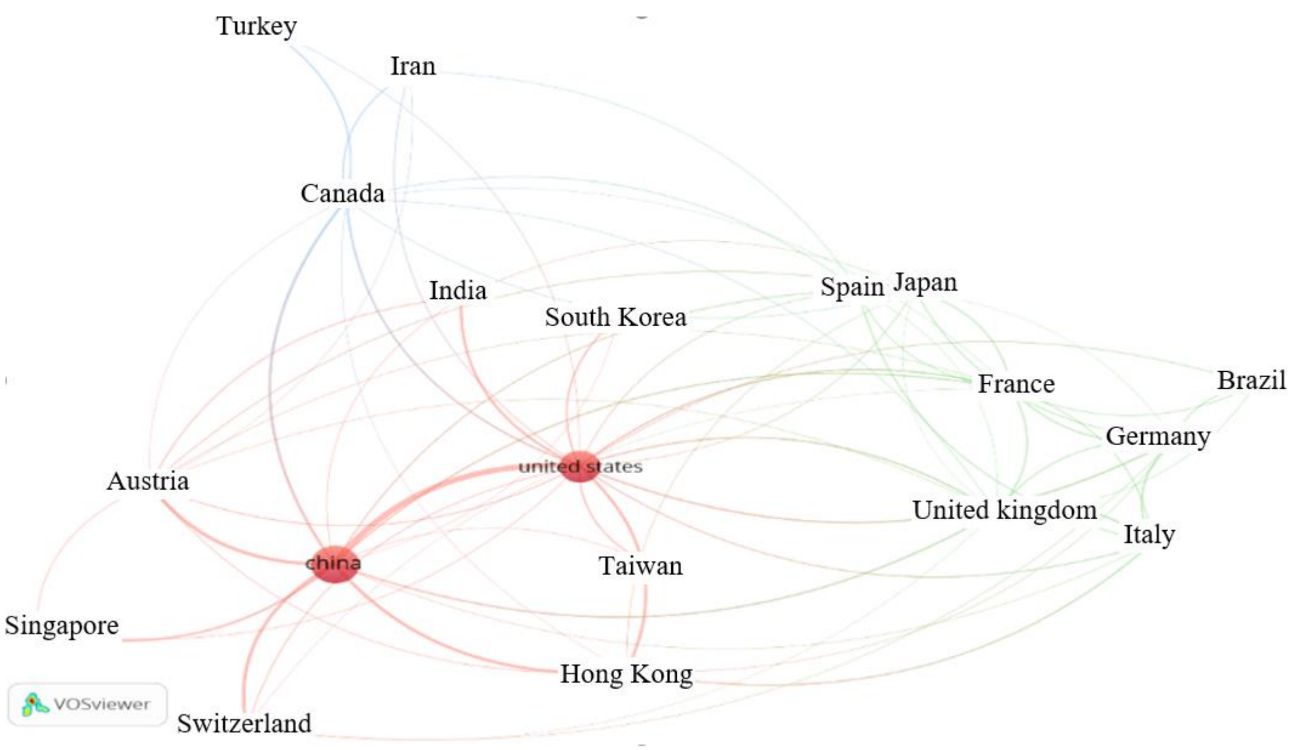


Figure 4. Top country's scientific collaboration.

Co-authorship was indicated by the graph edges, and the node clusters were shown by the node's colors. Clustering was performed by applying the VOS algorithm based on collaborations [54]. Three clusters could be described relatively through geographical distribution. The first Cluster consists of East Asian countries, like China, Hong Kong, India, Singapore, South Korea, Taiwan, and Australia. The second cluster consists of European countries like France, Germany, Italy, Spain, and the United Kingdom. The third cluster three includes two neighboring Asian countries, namely Iran and Turkey. In addition, the United States is regarded as a bridge between China and European countries.

3.5 Most productive and impactful authors

We provide the top products and impact based on the H-index[2] [21], the number of documents, Affiliation, and research field [55, 56], and these indices are computed with only the authors' LP documents than all articles of the authors. Based on

[2] The h-index is proposed by Hirsch (2007) that is defined as the H-index of a researcher is h if h of his/her papers have at least h citations each, and the other papers have at most h citations each.

Table 2, regarding the number of LP documents, the most prolific author is Philip S. Yu, a Computer Science professor at the University of Illinons at Chicago who works on *Data mining*, *Database, and Privacy* (with H-index; 112 and 46 of documents). His main research interests consist of *big data*, *data mining* (especially *Graph/Network mining*), *social networks*, *privacy-preserving data publishing*, *data stream*, *database systems*, *Internet applications, and technologies.* He is considered a top professor in the computer science department at UIC, and he gained the Wexler Chair in Information and Technology[3].

Table 2. Top 10 authors in LP documents.

H-index	Unit	All articles	Affiliation	Research field
112	Yu, P.S.	46	The University of Illinons at Chicago	Data mining, privacy-preserving publishing, and mining
68	Zhou, T.	27	The University of Electronic Science and Technology of China, Chengdu, China	Physics and Astronomy, Mathematics, Computer Science
49	Chawla, N.V.	27	University of Notre Dame, Notre Dame, United States	Computer Science, Mathematics, Engineering
39	Lehmann, J.	21	Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, Sankt Augustin, Germany	Computer Science, Mathematics, Social Sciences
26	Wu, B.	19	Beijing University of Posts and Telecommunications, Beijing, China	Computer Science, Mathematics, Engineering
23	Kaya, M.	19	Firat Üniversitesi, Elazig, Turkey	Computer Science, Medicine, Engineering
21	Chen, L.	18	Yangzhou University, Yangzhou, China	Computer Science, Engineering, Mathematics
18	Zhang, J.	19	The University of California, Davis, Davis, United States	Computer Science, Engineering, Decision Sciences
10	Zhu, X.	18	Beijing University of Posts and Telecommunications, Beijing, China	Physics and Astronomy, Computer Science, Mathematics
7	Liu, S.	19	Information Engineering University China, Zhengzhou, China	Computer Science, Engineering, Physics, and Astronomy

3.6 Most active institutions

The most productive institutions are listed in Table 3. Two important findings should be noted here. First, 9 of the Top 10 institutes, and academic centers involved in LP research are from China. The institutions of China published lots of documents. It can be due to China's population and its scientific productivity. Most of these productive institutions include universities, and research institutions like *the Chinese Academy of Sciences* and *Beijing University of Posts and Telecommunications*. Second, among the list of 10 active institutions and universities in this field, the only "University of Illinois at Chicago" with 45 documents is from the United States, which is well-known for data mining research.

3.7 Core sources

To identify the most important sources of information in the field, we use Bradford's law [56]. The law indicates that if the journals of a field are arranged acco
rding to the number of articles into three zones, the journals' number in each zone could be geometrically proportional to 1: n: n² [7]. Table 4 shows the number of published papers' number in each journal and the relative ranking according to Bradford's law zoning. 3,955 published LP papers were published by 1,090 sources. Among them, 650 sources only published one article, which constitutes the peripheral sources. The series formed by the zones is 19: 145: 926, which is approximately equal to 1: 7: 7^2. Table 5 provides the list of the core sources and their information.

[3] https://scholar.google.com/citations?hl=en&user=D0lL1r0AAAAJ

This relatively low number of sources provides about one-third of the publications in the field.

Table 3. Most productive institutions in LP research.

Institution	Country	Docs
Chinese Academy of Sciences	China	123
Beijing University of Posts and Telecommunications	China	104
Tsinghua University	China	89
Ministry of Education China	China	88
The University of Chinese Academy of Sciences	China	78
The University of Electronic Science and Technology of China	China	75
National University of Defense Technology	China	73
Nanjing University	China	51
The University of Illinois at Chicago	United States	45
Zhejiang University	China	43

Table 4. Bradford's core sources of LP research.

Zone	Number of journals	Number of articles	Multiplier
1	19	1313	-
2	145	1310	7.631578947
3	926	1309	6.386206897
Total	1090	3932	7.008892922

Table 5 shows the core sources of LP. Most sources' subject category is Computer Science (14/19) followed by Mathematics and Engineering (4/19). Journals consist of the majority of sources (9/19) followed by a conference source (8/19). *Lecture Notes in Computer Science* published the most number of documents (409/3955) with SJR 0.249.

3.8 Most cited papers

Citation count is regarded as an efficient index of research impact [50]; therefore, we apply this index to provide the most significant papers in the field. Table 6 shows the ten papers with the highest number of citations. Interestingly, seven out of ten papers include conference papers that show the respective significance of conferences to journals in the field. In the following, we provide a summary of these articles and their innovations.

The most effective paper is a conference paper published in *the 22nd ACM SIGKDD international conference on Knowledge discovery, and data mining proceedings*, which has been cited 3,160 times. Grover and Leskovec [57] proposed node2vec as an efficient scalable algorithm for feature learning in networks. The node2vec learns node representation in a low dimensional space of features that implement the probability of keeping network neighborhoods of nodes. Liben-Nowell and Kleinberg [49] applied different algorithms, like common neighbors, preferential attachment, Adamic/Adar, Jaccard, Sim Rank, Katz, Hitting time, and PageRank for the LP problem. Their results showed among others that the Katz measure, its variants, some simple measures such as common neighbors, and the Adamic/Adar measure perform well. Tang, Qu [58] suggested Large-Scale Information Network Embedding (LINE) that could simply expand to networks through many vertices and edges. It has meticulously developed objective functions, which keep the first-order and second-order proximities that complement each other. Bordes, Usunier [59] introduced Translating Embeddings (TransE) which is an energy-oriented model that makes knowledge-based embeddings. It models relations as translations performed on the entity embedding space. The tail entity embedding must be near the head entity embedding plus relationship embedding. Lü and Zhou [60] summarized LP algorithms, like the similarity-oriented algorithms (local similarity, global similarity, quasi-local indices), maximum likelihood approaches (hierarchical structure model and stochastic block model), probabilistic models (probabilistic relational models, probabilistic entity relationship models, stochastic relational models), underlining the recommendations from physical views and methods, like the random-walk-based approaches. and the maximum likelihood methods.

Table 5. List of core sources of LP research.

Row	Journals	#Docs	SJR (2020)	Subject area and category	Publisher
1	Lect. Notes Comput. Sci.	409	0.249	Computer Science, Mathematics	Springer Verlag
2	IEEE Access	97	0.587	Computer Science, Engineering, Materials Science	Institute of Electrical and Electronics Engineers Inc.
3	Phys A Stat Mech Appl	96	0.64	Physics and Astronomy, Mathematics	Elsevier
4	Int Conf Inf Knowledge Manage	77	0.519	Business, Management, Accounting, Decision Sciences	Association for Computing Machinery (ACM)
5	Commun. Comput. Info. Sci.	60	0.16	Computer Science, Mathematics	Springer Verlag
6	CEUR Workshop Proc.	57	0.177	Computer Science	---
7	Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.	55	1.004	Computer Science	---
8	ACM Int. Conf. Proc. Ser.	52	0.182	Computer Science	Association for Computing Machinery (ACM)
9	Web Conf. - Proc. World Wide Web Conf., WWW	46	0.545	Engineering	---
10	Proc. IEEE Int. Conf. Data Min. ICDM	43	1.36	Computer Science	IEEE Computer Society
11	IEEE Trans Knowl. Data Eng.	43	1.36	Computer Science	IEEE Computer Society
12	Knowl. Based Syst.	41	1.587	Business, Management and Accounting, Computer Science, Decision Sciences	Elsevier
13	Sci. Rep.	39	1.24	Multidisciplinary	Nature Publishing Group
14	Adv. neural inf. process. Syst.	36	1.4	Computer Science	---
15	Proc. - IEEE Int. Conf. Big Data, Big Data	34	0.231	Computer Science, Mathematics	---
16	AAAI Conf. Artif. Intell., AAAI	34	0.63	Computer Science	American Association for Artificial Intelligence (AAAI) Press
17	Soc. Netw. Analysis Min.	33	0.457	Computer Science, Engineering, Social Science	Springer-Verlag Wien
18	PLoS ONE	33	0.99	Multidisciplinary	Public Library of Science
19	Adv. Intell. Sys. Comput.	31	0.184	Computer Science, Engineering	Springer Verlag

Table 6. Top 10 most cited papers.

Author	Year	#Citations	Type	Source
[18]	2016	3160	Conference Paper	Proceedings of the 22nd ACM SIGKDD international conference on Knowledge Discovery and data mining
[19]	2007	2425	Article	Journal of the American Society for Information Science and Technology
[20]	2015	2184	Conference Paper	Proceedings of the 24th International Conference on the World Wide Web
[21]	2013	1961	Conference Paper	Advances in Neural Information Processing Systems
[22]	2011	1501	Article	Physica A: Statistical Mechanics and its Applications
[23]	2016	1112	Conference Paper	Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
[24]	2015	954	Conference Paper	Proceedings of the National Conference on Artificial Intelligence
[25]	2009	922	Article	European Physical Journal B
[26]	2014	886	Conference Paper	Proceedings of the AAAI Conference on Artificial Intelligence
[27]	2011	676	Conference Paper	Proceedings of the fourth ACM international conference on Web Search and Data Mining

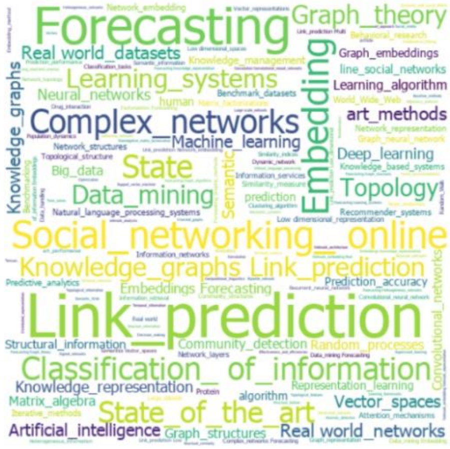
In addition, they offered three common applications, including network reconstruction, network evaluation evolving mechanisms, and partially labeled network classification. Wang and Cui [61] proposed the Structural Deep Network Embedding (SDNE), which is a semi-supervised deep model with multiple non-linear layers. The model jointly utilizes first and second-order proximity. Zhou, Lü [62] empirically explored some LP algorithms for relation estimation based on node similarities. Numerical findings regarding the nine local similarity indices specified that the simplest indices, common neighbors, have the best application, and the Adamic-Adar index is the next. The results are to some extent in agreement with Liben-Nowell and Kleinberg [49]. Furthermore, they suggested a new measure, resource allocation (RA), inspired by the resource allocation process, which is equal to the one-step random walk starting from the common neighbors.

Lin, Liu [24] proposed Translating Relations (TransR), which is a knowledge graph embedding model, that lodges entities and relations in different entity and relation spaces and learns embeddings through translation between projected entities. Their experiments indicated that TransR attains major developments compared to TransE and TransH. Wang, Zhang [63] introduced Translating on Hyperplanes (TransH) for knowledge graph embedding. TransH overcomes the errors of TransE concerning the automatic/one-to-many/many-to-one/many-to-many relations while taking over its efficiency. TransH models a relation as a hyperplane with a translation operation. The results show that TransH performs significantly better than TransE. Backstrom and Leskovec [64] suggested Supervised Random Walks (SRW) for LP and link recommendation that combines the network structure with node and edge-level features.

3.9 Keyword analysis

We examined the most frequently applied key terms in two time periods from 2012 to 2021. Figure 5 shows the wordclouds of these periods (2012-2016 and 2017-2021), which are applied to summarize texts, visually [65]. The size of the key terms in the wordclouds indicates their frequency in that period. Table 7 shows the most applied key terms in the periods. The right column of Table 7 reveals the key terms' growth in the second period relative to the first one. The first column's red cells indicate key terms, which do not exist in the second Period's top 20 lists and green cells include the key terms that appeared in the second period's top 20 lists. That is to say, the red cells illustrate the key terms that lost their significance, and the green ones are key terms that have become more important in recent years. Common key terms in the first and second periods are social networking (online), complex networks, data mining, artificial intelligence, learning systems, topology, graph theory, semantics, and community detection.

2012-2016



2017-2021

Figure 5. wordclouds of keywords in two time periods.

The most frequently used key terms in the first period (2012-2016) are social networking (online), SNs, algorithms, complex networks, data mining, artificial intelligence, learning systems, topology, SNA, and graph theory. The most frequently used keywords in the second period (2017-2021) include embeddings, knowledge graphs, social networking (online), complex networks, data mining, learning systems, network embedding, graph theory, semantics, and knowledge representation. As it is obvious from this figure, LP and forecasting (which are rough synonyms) are the most common key terms in the first and second periods, respectively. Some key terms in the first period disappeared from the list in the second period, such as SNs, algorithms, and SNA. Moreover, other key terms appeared in the second period, such as embeddings, knowledge graphs, and network embedding. The presence of key terms, such as embeddings, network embedding, machine learning, deep learning, and learning systems show that the approach of the LP community has shifted to the use of machine learning methods, data mining, and especially deep learning.

Table 7. Most used keywords in two different periods (2012-2016) and (2017-2021).

Rank	Keyword	2012-2016	Keyword	2017-2021
1	Social networking (online)	408	Embeddings	584
2	Social networks	196	Knowledge graphs	331
3	Algorithms	144	Social networking (online)	329
4	Complex networks	143	Complex networks	303
5	Data mining	122	Data mining	287
6	Artificial intelligence	97	Learning systems	274
7	Learning systems	82	Network embedding	273
8	Topology	68	Graph theory	243
9	Social network analysis	62	Semantics	230
10	Graph theory	58	Knowledge representation	227
11	Semantics	52	Classification (of information)	209
12	Prediction	48	Vector spaces	178
13	Community detection	44	Deep learning	166
14	Prediction accuracy	44	Graphic methods	153
15	Information services	39	Machine learning	152
16	World wide Web	39	Topology	152
17	Websites	36	Artificial intelligence	148
18	Heterogeneous networks	35	Graph structures	143
19	Big data	34	Community detection	122
20	Factorization	33	Network structures	121

In the following, we will briefly explain the second period's most frequent key terms. LP is the first keyword and the main topic of the current paper, and we have explained it in the previous sections.

Forecasting is nearly anticipating the future as correctly as possible, regarding all of the information accessible, involving historic data and any future events' knowledge that may affect the predictions [65]. The embedding concept originated from natural language processing (NLP) [66]. Word embedding is a term used for the word representation usually using a real-valued vector that encodes the word's meaning [44]. Recently, the concept of embedding has appeared in the context of processing graph data using deep learning. These vectors encode the meaning of the network [67].

Social Networking (online) or Social Networking Service (SNS) is regarded as an online platform that individuals apply to form SNs or social links with other persons who have the same private or business affairs, actions, experiences, or real-life links [68]. Social networking (online) has three specific features; the user creates a "profile", the user can associate himself with other users, and related users can link publicly, and privately, and interchange text, images, video, and audio files.

Knowledge Graphs (KGs) offer an organized image of knowledge about the world and contain only a small subclass of what is true in the world [69]. KGs have several applications, such as semantic search based on objects, and relationships, natural language disambiguation, deep reasoning (e.g. IBM Watson), machine reading (e.g. text summarization), entity consolidation for big data, text analytics, engineering, and scientific applications [70]. Moreover, a knowledge graph is defined as a knowledge base that implements a graph-oriented data topology to link data. KGs are frequently stored interlinked descriptions of entities, objects, processes, states, or unique ideas with freestyle semantics [71]. KGs are widely applied in the AI field, like in information retrieval, processing natural language, and recommendation systems. Yet, KGs' open essence frequently shows that they are imperfect, containing self-defects. This generates the requirement to make a fuller knowledge graph for improving the KG's practical implementation. LP is regarded as a crucial task in knowledge graph development that applies existing ties to infer new ties to form a more complete knowledge graph [59].

Complex Networks (CN) are graphs (networks) that contain non-trivial topological features-features, which do not occur in simple networks, like frameworks or random graphs, but often pursue networks that are real systems representative. CNs study is a novel and dynamic field of scientific study [72, 21, 58]. CNs define a broad spectrum of systems in nature and society, much-quoted instances and the cell, a chemicals-linked network by chemical reactions or the Internet, a routers' network, and computers linked by physical links [72].

Data Mining (DM) is the mining approach, which determines patterns in big data sets and approaches at the joining of machine learning, statistics, and database systems. Disregarding the phase of raw analysis, it also includes a database, data management characteristics, data pre-processing model, inference reflections, interestingness metrics, complexity reflections, exposed structures post-processing, visualization, and online updating [60]. DM is considered an interdisciplinary field resulting from joining computer science and statistics with a general aim to extract information (with intelligent approaches) from a data set, and convert the information into a transparent format for further utilization [60, 61, 73]. DM is the analysis step of the Knowledge Discovery in Databases (KDD) approach [74].

Network Embedding (NE) assigns nodes to low-dimensional demonstrations in a network and effectively reserves the structure of a network. NE, as an efficient method of network depiction, can support upcoming network handing out and analysis activities, like node categorization, node clustering, network visualization, and LP [63].

Learning Systems (LS) is necessarily an artifacts collection, which is 'brought together', in a proper method, to make the situation that will ease different kinds of learning processes. LS could take various formats, including a book, a computer, an online forum, and a university. The majority of learning systems could present different kinds of learning resources, and depiction of methods for implementing these to attain specific learning results. They will also employ different strategies for evaluating the levels and quality of their users' gains [64].

Graph Theory (GT) is a branch of discrete mathematics focused on the study of graphs and has been used widely to model many kinds of ties and processes in physical, biological, and social information systems [75-77].

Graph neural networks (GNNs) are neural models that capture the dependence of graphs by message passing between the nodes of graphs. Over the past few years, graph neural networks have become influential and useful tools for machine learning tasks in the graph field. This growth owes to progress in expressive power, model flexibility, and training algorithms. Modeling physics systems, learning molecular fingerprints, predicting protein interface, and classifying diseases claim a model to learn from graph inputs. In other domains such as learning from non-structural data like texts and images, reasoning on extracted structures is a key research topic that also needs graph reasoning models.

In recent years, variants of GNNs such as graph convolutional network (GCN), graph attention network (GAT), and graph recurrent network (GRN) have confirmed ground-breaking shows on many deep learning tasks [79].

GNNs are a powerful computational tool to jointly learn graph structure and node/edge features, so that accomplished an unmatched accuracy in the link prediction problem, i.e. the task of predicting if two nodes are probable to be tied by an edge shortly. Graph Conversion Capsule Link (GCCL) technique, uses Capsule Networks (Caps Net), to resolve the link prediction problem. they have planned a conversion block to convert the link prediction problem into a binary classification task. Based on node features extracted by GNN, the feature map of each node pair (edge feature map) is removed by conversion and a Caps Nets architecture to extract features in the form of the vector [80].

Graph neural networks (GNNs) have received notable success in link prediction (GNNLP) tasks. Existing efforts first predefine the subgraph for the whole dataset and then apply GNNs to convert edge representations by leveraging the neighborhood structure persuaded by the fixed subgraph. Since node connectivity in real-world graphs is complex, one shared subgraph is limited for all edges. So, the selections of subgraphs should be personalized to different edges. To bridge the gap, we introduce a Personalized Subgraph Selector (PS2) as a plug-and-play framework to routinely, personally, and inductively identify optimal subgraphs for different edges when executing GNNLP. PS2 is instantiated as a bi-level optimization problem that can be efficiently solved differently. Coupling GNNLP models with PS2, they suggest a brand-new angle toward GNNLP training: by first identifying the best subgraphs for edges; and then focusing on training the implication model by using the tested subgraphs [81]. Many Graph Neural Networks (GNNs) achieve poorly compared to simple heuristics on Link Prediction (LP) tasks. This is due to limitations in sensitive power such as the weakness to count triangles and because they cannot differentiate automorphic nodes (those having equal structural roles). Both expressiveness issues can be alleviated by learning link (rather than node) representations and incorporating structural features such as triangle counts. Since obvious link representations are often prohibitively expensive, recent works resorted to subgraph-based methods, which have achieved state-of-the-art performance for LP, but suffer from poor efficiency due to high levels of redundancy between subgraphs.

They analyzed the workings of subgraph GNN (SGNN) methods for link prediction. Based on their analysis, planned a novel full-graph GNN called ELPH (Efficient Link Prediction with Hashing) that permits subgraph sketches as messages to estimate the key components of SGNNs without obvious subgraph structure. ELPH is probably more expressive than Message Passing GNNs (MPNNs). It outperforms existing SGNN models on many standard LP benchmarks while being orders of magnitude more rapidly [82].

Link prediction, as one of the key problems for network-structured data, aims to predict whether there exists a link between two nodes. The traditional approaches are based on the clear similarity computation between the compact node representation by embedding each node into a low-dimensional space. To powerfully handle the intensive similarity computation in link prediction, the hashing technique has been effectively used to produce the node representation in the Hamming space. Currently, the Graph Neural Network (GNN) framework has been widely applied to graph-related tasks in an end-to-end manner.

They proposed a simple and effective model called #GNN, which balances the trade-off between accuracy and efficiency. #GNN can efficiently acquire node representation in the Hamming space for link prediction by exploiting the randomized hashing technique to implement message passing and capture high-order proximity in the GNN framework. The extensive experimental results demonstrate that the proposed #GNN algorithm achieves accuracy comparable to the learning-based algorithms and outperforms the randomized algorithm while running significantly faster than the learning-based algorithms. Also, the proposed algorithm shows excellent scalability on a large-scale network with limited resources [83].

4. Conclusion

In this paper, we presented a view on the global trends in LP in social networks research. We performed a bibliometric analysis, including the trend of annual publications, document types, subject areas, most productive countries, most productive, and impactful authors, most active institutions, core sources, most cited papers, and keyword analysis. The LP in Social Networks has got increasing attention in the related literature. Based on the analysis of this study, several findings are reported as follows. First, the productivity of literature on the LP is still developing. Second, the major institutions which have written the most LP papers are located in China. Third, the annual trend of LP publications shows that from 1987 to 2021 the average growth rate of research was 33.08%. Fourth, LP is chiefly applied to some fields of study, like computer science, mathematics, engineering, and decision sciences.

On the other hand, according to the analysis of the results of different years, several interesting conclusions could be found. First, computer science and mathematics are the fields that have the most contribution to LP research, and the number of articles, especially on engineering (%11.70) and decision sciences (%6.35) gradually increased. Second, in the topic of analysis, LP was gradually used in the two contexts of SN and SNA for discussion. LP is also important for social science because SN analysis is an important topic in this field. Although LP originated from computer science, in recent years, it has been investigated in physics, astronomy, social sciences, business, management, and accounting because of its importance. The results are significant because LP originally is a task of machine learning and data mining, which are their own computer science subfields. On the other hand, the root of GT is originally from mathematics.

Finally, using Bradford's law to identify the most important sources of information in the field of LP proved that the series formed by the zones is 19: 145: 926, which is approximately equal to 1: 7: 7^2 which means the data were consistent with Bradford's law. In addition, the 19 core journals in LP issues were identified. Among these core journals, the number of articles published in *Lect. Notes Comput. Sci.* is the largest. The main reason for this is that *Lect. Notes Comput. Sci.* is a journal mostly concentrated on computer science and mathematics. However, the more interesting point is that *IEEE Access* is a journal publishing basic studies on the various fields of computer science, engineering, materials science, and *Phys a Stat Mech Appl* is a journal mainly publishing studies on physics, astronomy, and mathematics in which both occupied a large number of published articles in LP. In addition to the above three core journals, the others include *Commun. Comput. Info. Sci., IEEE Trans Knowl Data Eng, Knowl Based Syst., Sci. Rep., Adv. neural inf. process. Syst., Soc. Netw. Analysis Min., PLoS ONE, and Adv. Intell. Sys. Comput.* among the core journals with the subject category of computer science followed by mathematics and engineering.

This study identified useful limitations that offer direct avenues for additional research. First, although we are confident that a single database, Scopus, is large enough to offer a wide variety of publications vital for our analyses, future studies can apply other databases, such as Web of Science to retrieve more in-depth issues. Second, our search strategy may not include all of the related documents. This study can be implemented for future further analysis, like keyword content analysis to reach a deeper comprehension of this topic in which it could present a taxonomy of tasks that are tackled by applying LP and their results.

Acknowledgments

This article was written without financial support. The supervisor "Mr. Dr. Mohsen Jahanshahi" (Tehran Islamic Azad University, Center Branch, Technical and Engineering Faculty), and also "Mr. Dr. Abdul Samad Karamatfar" (Scientific Information Center of Academic Jihad), who played a valuable role in promoting this work with his valuable comments. We sincerely thank you. No other government or private organization or agencies supported or funded this research work.

Reference

[1] J. Tang, *et al.*, "Line: Large-scale information network embedding", in *Proceedings of the 24th international conference on the world wide web,* 2015.

[2] J. Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques" (3rd Ed), Morgan Kauffman, 2011.

[3] U., G. Fayyad, Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases". *AI magazine*, vol. 17, no. 3: p. 37-37, 1996.

[4] P. Cui, *et al*., "A survey on network embedding". *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 5, pp. 833-852, 2018.

[5] M.N. Hallquist, and F.G. Hillary, "Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world". *Network Neuroscience*,. vol. 3, no. 1, pp. 1-26, 2018.

[6] P.Shah, *et al*., "Characterizing the role of the structural connectome in seizure dynamics". *Brain*, vol. 142, no. 7, pp. 1955-1972, 2019..

[7] B. Yoon and C.L. Magee, "Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction", *Technological Forecasting and Social Change*, vol. 132, Part C, pp.105–117, 2018.

[8] L. Bornmann, and H.-D. Daniel, "Does the h-index for ranking of scientists really work?" *Scientometrics*,. vol. 65, no. 3, pp. 391-392, 2005.

[9] JA. Morente-Molinera, G. Kou, K. Samuylov, R. Ureña, and E. Herrera-Viedma. "Carrying out consensual group decision-making processes under social networks using sentiment analysis over

comparative expressions". *Knowledge-Based Systems*; 165: 335-45, 2019.

[10] JC. Mitchell, "Social networks in urban situations: analyses of personal relationships in Central African towns". Manchester University Press; ISBN-13: 978-0719010354, 1969.

[11] Y-S. Su, C-L Lin, S-Y Chen, and C-F Lai. "Bibliometric study of social network analysis literature". *Library Hi Tech*.; vol. 38, no. 2, 420-33, 2020.

[12] J. Ding, *et al*., "Member structure and sharing behavior: social network analysis of CALIS online cataloging data in China". *The Journal of Academic Librarianship*, vol. 46, no. 2, p. 102115, 2020.

[13] J. Horkoff, F. B. Aydemir, E. Cardoso, T. Li, Maté, A., Paja, E. and P. Giorgini, "Goal-oriented requirements engineering: an extended systematic mapping study". *Requirements engineering*, vol. 24, pp. 133-160, 2019.

[14] Z. Bu, Y. Wang, H-J. Li, J. Jiang, Z. Wu, and J. Cao." Link prediction in temporal networks: Integrating survival analysis and game theory". *Information Sciences*.; 498: 41-61, 2019.

[15] S. Rafiee, C. Salavati, A. Abdollahpouri "CNDP: Link prediction based on common neighbor's degree penalization". *Physica A: Statistical Mechanics and its Applications.* 539: 122950, 2020.

[16] M. F. Jamaludin, and F. Hashim, 2017. "Corporate governance, institutional characteristics, and director networks in Malaysia. Jamaludin, MF and Hashim, F. Corporate governance, institutional characteristics, and director network in Malaysia". Asian Academy of Management *Journal of Accounting and Finance*, vol. 13, no. 2, pp. 135-154, 2017.

[17] I. Kuznetcova, M. Glassman and T. J. Lin, "Multi-user virtual environments as a pathway to distributed social networks in the classroom". *Computers & Education*, vol. 130, pp. 26-39, 2019.

[18] K. Lewis, and J. Kaufman, "The conversion of cultural tastes into social network ties". *American Journal of Sociology*, vol. 123, no. 6, pp. 1684-1742, 2018.

[19] L. Kirichenko, T. Radivilova, and A. Carlsson, "Detecting cyber threats through social network analysis: short survey". arXiv preprint arXiv:1805.06680, (2018).

[20] S. Nahhas, O. Bamasag, M. Khemakhem and N. Bajnaid, "Added values of linked data in education: A survey and roadmap". Computers, vol. 7, no. 3, 45, (2018).

[21] N.J. Van Eck, and L. Waltman, "VOS: A new method for visualizing similarities between objects, in Advances in data analysis", Springer. p. 299-306, 2007.

[22] J. Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques" (3rd Ed), Morgan Kauffman, 2011.

[23] L. Bornmann, and H.-D. Daniel, "Does the h-index for ranking of scientists really work?" *Scientometrics*, vol. 65, no. 3, pp. 391-392, 2005.

[24] G. Sondossi, A. Saebi, SA. Hashemi. "A Neighbor-based Link Prediction Method for Bipartite Networks". *Journal of Information and Communication Technology.* 13(47, 48), pp. 178-186, 2021.

[25] B. Bilecen, M. Gamper, and MJ. Lubbers. "The missing link: Social network analysis in migration and transnationalism". *Social Networks.* 53: 1-3, 2018.

[26] A. Kumari, RK. Behera, KS. Sahoo, A Nayyar, A Kumar Luhach, S Prakash Sahoo. "Supervised link prediction using structured-based feature extraction in social network. Concurrency and Computation", *Practice and Experience.* vol. 34, no. 13: e5839, 2022.

[27] I. Ahmad, MU. Akhtar, S. Noor, and A. Shahnaz. "Missing link prediction using common neighbor and centrality based parameterized algorithm". *Scientific reports*. 10: 364, 2020.

[28] X. Ma, P. Sun, and Y. Wang. "Graph regularized nonnegative matrix factorization for temporal link prediction in dynamic networks". *Physica A: Statistical mechanics and its applications.* 496: pp. 121-36, 2018.

[29] PM. Chuan, LH. Son, M. Ali, TD. Khang, LTH. uong, and N. Dey. "Link prediction in co-authorship networks based on hybrid content similarity metric". *Applied Intelligence*. vol. 48, no. 8: 2470-86, 2018.

[30] B. Bilecen, M. Gamper, MJ Lubbers."The missing link: Social network analysis in migration and transnationalism". *Social Networks.* 53: 1-3, 2018.

[31] M. E. Falagas, A.I. Karavasiou, and I.A. Bliziotis, "A bibliometric analysis of global trends of research productivity in tropical medicine". *Acta tropica*, 99(2-3), pp. 155-159, (2006).

[32] C. Chen and Song," Visualizing a field of research: A methodology of systematic scientometric reviews". *PLOS ONE*. 14(10): e0223994, 2019.

[33] B. González-Pereira, V.P. Guerrero-Bote, and F. Moya-Anegón, "A new approach to the metric of journals' scientific prestige: The SJR indicator". *Journal of Informetrics*, vol. 4, no.3, pp. 379-391, 2010.

[34] A. Keramatfar and H. Amirkhani," Bibliometrics of sentiment analysis literature". *Journal of Information Science.* vol. 45, no. 1, pp. 3-15, 2019.

[35] D. R. Raban and A. Gordon, " The evolution of data science and big data research: A bibliometric analysis". *Scientometrics,* vol. 122, no. 3, pp. 1563-1581, (2020).

[36] J. Li, Peng, Liu, S., X. Ji, X. Li, and X. Hu, "Link prediction in directed networks utilizing the role of reciprocal links". *IEEE Access*, vol. 8, pp. 28668-28680, (2020).

[37] S. Behrouzi, Z. S. Sarmoor, K. Hajsadeghi, and K Kavousi, "predicting scientific research trends based on link prediction in keyword networks". *Journal of Informetrics*, vol. 14, no. 4, pp. 101079, (2020).

[38] J.F. Burnham, Scopus database: a review. Biomedical Digital Libraries, vol. 3, no. 1: p. 1, 2006.

[39] H. Bingjie, N. Ling, and S. J. Abbas, "Research on Drug-Target Interactions Prediction: Network similarity-based approaches". In *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)* (pp. 168-173). IEEE, (2020, July).

[40] G.P. Khiste and R.R. Paithankar, "Analysis of Bibliometric term in Scopus". *International Journal of Library Science and Information Management (IJLSIM)*, vol. 3, no. 3, pp. 81-88, 2017.

[41] G.P. Khiste, "Publication productivity of 'consortia'by scopus during 1989-2016". *International Journal of Current Innovation Research*, vol. 3, no. 11, pp. 879-882, 2017.

[42] G.P. Khiste, D.B. Maskeand R.K. Deshmukh, "Knowledge management output in Scopus during 2007 to 2016". *Asian Journal of Research in Social Sciences and Humanities*, vol. 8, no. 1, pp. 10-19, 2018.

[43] S.A. Ebrahim, *et al.*, "Quantitative and Qualitative Analysis of Time-Series Classification Using Deep Learning". *IEEE Access*, 8: pp. 90202-90215, 2020.

[44] E. K. Pedersen, "Embeddings of topological manifolds". *Illinois Journal of Mathematics*, vol. 19, no. 3, pp. 440-447, (1975).

[45] N.J. Van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping". *scientometrics*, vol. 84, no. 2, pp. 523-538, 2010.

[46] O. Persson, R. Danell and J.W. Schneider,"How to use Bibexcel for various types of bibliometric analysis". *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday*, vol. 5, pp. 9-24, 2009.

[47] E-SM. El-Alfy and SA. Mohammed, "A review of machine learning for big data analytics: bibliometric approach".*Technology Analysis & Strategic Management.* vol. 32, no. 8, pp. 984-1005, 2020.

[48] S.C. Bradford," Sources of information on specific subjects". *Engineering*, vol. 137, pp. 85-86, 1934.

[49] D.Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks". *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019-1031, 2007.

[50] M. Mao, Z. Li, Z. Zhao and L. Zeng, "Bibliometric analysis of the deep learning research status with the data from Web of Science". In *Data Mining and Big Data: Third International Conference, DMBD 2018, Shanghai, China*, June 17–22, 2018, Proceedings 3 (pp. 585-595). Springer International Publishing, (2018).

[51] Y. Zhang, M. Wu, GY. Tian, G. Zhang, and J. Lu, "Ethics and privacy of artificial intelligence: Understandings from bibliometrics". *Knowledge-Based Systems.* 222: 106994, 2021.

[52] K. Berahmand, E. Nasiri, M. Rostami, and S. Forouzandeh, "A modified DeepWalk method for link prediction in attributed social network". *Computing*. vol. 103, no. 10, pp.
2227-49, 2021.

[53] A. Keramatfar and H. Amirkhani, "Bibliometrics of sentiment analysis literature". *Journal of Information Science*, vol. 45, no. 1, pp. 3-15, 2019.

[54] R. Cohen and S. Havlin, "Complex networks: structure, robustness, and function". *Cambridge University Press*, (2010).

[55] J.E. Hirsch, "Does the h index have predictive power?" *Proceedings of the National Academy of Sciences*, vol. 104, no. 49, pp. 19193-19198, 2007.

[56] T. Pradhan and S. Pal, "A hybrid personalized scholarly venue recommender system integrating social network analysis and contextual similarity". *Future Generation Computer Systems*, vol. 110, pp. 1139-1166, (2020).

[57] E. Roldan-Valadez, *et al.*," Current concepts on bibliometrics: a brief review about impact factor, Eigenfactor score, CiteScore, SCImago journal rank, source-normalized impact per paper, H-index, and alternative metrics". *Irish Journal of Medical Science* (1971-), vol. 188, no. 3, pp. 939-951, 2019.

[58] Tang J, Qu M, Wang M, Zhang M, Yan J, and Mei Q, editors. "Line: Large-scale information network embedding". *Proceedings of the 24th International Conference on World Wide Web*, 2015.

[59] Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O, editors. "Translating embeddings for modeling multi-relational data". *Neural Information Processing Systems (NIPS)*, 2013.

[60] L. Lü, and T. Zhou, "Link prediction in complex networks: A survey". *Physica A: statistical mechanics and its applications*, vol. 390, no. 6, pp. 1150-1170, (2011).

[61] D. Wang, P. Cui, and W. Zhu, " Structural deep network embedding". *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016.

[62] L. Lü and T. Zhou," Link prediction", (2013).

[63] Wang, Z., Zhang, J., Feng, J., and Chen, Z. "Knowledge graph embedding by translating on hyperplanes". In *Proceedings of the AAAI conference on artificial intelligence* (vol. 28, no. 1), (2014, June).

[64] L. Backstrom and J. Leskovec. "Supervised random walks: predicting and recommending links in social networks". In *Proceedings of the fourth ACM International Conference on Web Search and data mining*, 2011.

[65] F. Heimerl, S. Lohmann, S. Lange, and T. Ertl, "Word cloud explorer: Text analytics based on word clouds". *47th Hawaii international conference on system sciences; IEEE*, 2014.

[66] RJ. Hyndman and G. Athanasopoulos," Forecasting: principles and practice". 2nd edition. OTexts, 2018.

[67] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing". *Expert Systems with Applications*. 166: 114060, 2021;

[68] A. Stolcke, Ries K, Coccaro N, Shriberg E, Bates R, Jurafsky D *Et al.* "Dialogue act modeling for automatic tagging and recognition of conversational speech". *Computational linguistics.* vol. 26, no. 3, pp. 339-73, 2000.

[69] Obar JA and Wildman S. "Social media definition and the governance challenge-an introduction to the special issue". *Telecommunications policy*. vol. 39, no. 9, pp. 745-50, 2015.

[70] Kazemi SM and Poole D. "Simple embedding for link prediction in knowledge graphs". *Advances in neural information processing systems.* vol. 31, 2018.

[71] J. P. McCusker, S. M. Rashid, N. Agu, K. P. Bennettand, D. L.McGuinness, "Developing Scientific Knowledge Graphs Using Why is*". In SemSci@ ISWC* (pp. 52-58), (2018, October).

[72] Z, editors. "Knowledge graph embedding by translating on hyperplanes". *Proceedings of the AAAI conference on artificial intelligence;* DOI: https://DOI.org/10.1609/aaai.v28i1.8870, 2014.

[73] Y. Lin, *et al*. "Learning entity and relation embeddings for knowledge graph completion". In *Proceedings of the AAAI Conference on Artificial Intelligence*. 2015.

[74] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information". *The European Physical Journal B*, vol. 71, no. 4, pp. 623-630, 2009.

[75] S. Wasserman, and F.K. aust, Social network analysis: Methods and applications (1994).

[76] S. Chakrabarti, *et al*., "Data mining curriculum: A proposal (Version 1.0)". *Intensive Working Group of ACM SIGKDD Curriculum Committee*, vol. 140: pp. 1-10. 2006.

[77] C. Clifton, "Encyclopædia Britannica: definition of data mining". Retrieved on, 2010. vol. 9, no. 12, p. 2010.

[78] M. Nilufar, and A. Abhari, "Incremental Text Clustering Algorithm for Cloud-based Data Management in Scientific Research Papers". In *2022 Annual Modeling and Simulation Conference (ANNSIM)* (pp. 778-789). IEEE. (2022, July).

[79] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, and M. Sun, "Graph neural networks: A review of methods and applications". *AI open*, vol. 1, pp. 57-81. (2020).

[80] X. Liu, X. Li, G. Fiumara, & P. De Meo," Link prediction approach combined graph neural network with capsule network". *Expert Systems with Applications*, *212*, 118737. (2023).

[81] Q. Tan, X. Zhang, N. Liu, D. Zha, L. Li, R. Chen, and X. Hu, "Bring your own view: Graph neural networks for link prediction with personalized subgraph selection". In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining* (pp. 625-633). (2023, February).

[82] B. P. Chamberlain, S. Shirobokov, E. Rossi, F. Frasca, T. Markovich, N. Hammerla, and M. Hansmire, "Graph Neural Networks for Link Prediction with Subgraph Sketching". *arXiv preprint arXiv:2209.15486*. (2022).

[83] B.W. Wu, C. Li, Luo, and W. Nejdl, "Hashing-accelerated graph neural networks for link prediction". In *Proceedings of the Web Conference 2021* (pp. 2910-2920). Skarding, J., Hellmich, M., Gabrys, B., & Musial, K. A Robust Comparative Analysis of Graph Neural Networks on Dynamic Link Prediction. *IEEE Access*, *10*, 64146-64160. (2022).

[84] X. Kong, Y. Shi, S. Yu, J. Liu, and F. Xia, "Academic social networks: Modeling, analysis, mining, and applications". *Journal of Network and Computer Applications*, vol. 132, pp. 86-103. (2019).

[85] I. Makarov, and O. Gerasimova, "Predicting collaborations in co-authorship network". In *2019 14th international workshop on semantic and social media adaptation and personalization (SMAP)* (pp. 1-6), IEEE.

[86] Sisay Fissaha Adafre and Maarten de Rijke. "Discovering missing links in wikipedia". In *Proceedings of the 3rd International Workshop on Link Discovery, LinkKDD '0*5, pages 90–97, New York, NY, USA, ACM, 2005.

[87] Lars Backstrom and Jure Leskovec. "Supervised random walks: Predicting and recommending links in social networks". In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11*, pages 635–644, New York, NY, USA, 2011.

[88] Qi He, Jian Pei, Daniel Kifer, Prasenjit Mitra, and Lee Giles. "Context aware citation recommendation". I*n Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pages 421–430, New York, NY, USA, 2010.

[89] David Liben-Nowell and Jon Kleinberg. "The link-prediction problem for social networks". *Journal of the Association for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[90] V. Smojver, Štorga, M., and G. Zovak, "Exploring knowledge flow within a technology domain by conducting a dynamic analysis of a patent co-citation network". *Journal of Knowledge Management*, vol. *25*, no. 2, pp. 433-453, (2021).

[91] A. Velez-Estevez, I. J. Perez, P. García-Sánchez, J. A. Moral-Munoz & M. J. Cobo, "New trends in bibliometric apis: A comparative analysis". *Information Processing & Management*, vol. *60*, no. 4, 103385, (2023).

[92] V. K. Singh, P. Singh, M. Karmakar, J. Leta, and P. Mayr, *The journal coverage of Web of Science*, "Scopus and Dimensions: A comparative analysis". *Scientometrics*, vol. 126, pp. 5113-5142, (2021).

[93] M. Khazaei, & N. Ashrafi-Payaman, "An Unsupervised Anomaly Detection Model for Weighted Heterogeneous Graph". *Journal of AI and Data Mining*, (2023).

پیش‌بینی پیوند در شبکه‌های اجتماعی: تحلیل کتاب‌سنجی و بررسی ادبیات (۱۹۸۷-۲۰۲۱)

اکرم پسندیده* و محسن جهانشاهی

گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی واحد تهران مرکزی، ایران.

ارسال ۲۰۲۳/۰۲/۱۰؛ بازنگری ۲۰۲۳/۰۳/۲۹؛ پذیرش ۲۰۲۳/۰۷/۱۰

چکیده:

پیش‌بینی پیوند به یک موضوع داغ در جامعه داده‌کاوی، یادگیری ماشین و یادگیری عمیق تبدیل شده است. هدف این مطالعه پیاده‌سازی تحلیل کتاب‌سنجی برای یافتن وضعیت فعلی مطالعات پیش‌بینی پیوند و بررسی آن از دیدگاه‌های مختلف است. مطالعه حاضر یک مرور کتاب‌سنجی مبتنی بر Scopus از چشم‌انداز مطالعات پیش‌بینی پیوند از سال ۱۹۸۷ ارائه می‌کند، زمانی که مطالعات پیش‌بینی پیوند برای اولین بار منتشر شد. انواع مختلفی از تجزیه و تحلیل از جمله سند، موضوع، و توزیع کشور استفاده می‌شود. علاوه بر این، از بهره‌وری نویسنده، تحلیل استنادی و تحلیل کلمات کلیدی استفاده می‌شود و قانون برادفورد برای کشف مجلات اصلی در این زمینه استفاده می‌شود. بیشتر اسناد توسط کنفرانس‌های این حوزه منتشر شده است. اکثر اسناد پیش‌بینی پیوند در زمینه‌های علوم کامپیوتر و ریاضیات منتشر شده است. چین تاکنون پیشتاز کشورهای منتشر کننده بوده است. علاوه بر این، فعال‌ترین منابع انتشارات پیش‌بینی پیوند، یادداشت‌های سخنرانی در علوم کامپیوتر از جمله یادداشت‌های سخنرانی زیر مجموعه در هوش مصنوعی و یادداشت‌های سخنرانی در بیوانفورماتیک و دسترسی IEEE است. تجزیه و تحلیل کلمات کلیدی نشان می‌دهد که در حالی که شبکه های اجتماعی در دوره اولیه توجه را به خود جلب کرده بودند، نمودارهای دانش اخیرا توجه بیشتری را به خود جلب کرده‌اند. از آنجایی که مشکل پیش‌بینی پیوند اخیراً با استفاده از یادگیری ماشین مورد بررسی قرار گرفته است، مطالعه فعلی ممکن است به محققان کمک کند تا روی تکنیک‌های یادگیری ماشین تمرکز کنند.

کلمات کلیدی: کتاب سنجی، شبکه‌های اجتماعی، تحلیل شبکه‌های اجتماعی، پیش‌بینی پیوند، قانون برادفورد.