Theme Evolution of Researches on Knowledge Graphs Based on Visualization Analyses of Data

Kejun Chen, Boyang Xie and Sanhong Deng*
School of Information Management, Nanjing University, Nanjing, China

*Corresponding author email: sanhong@nju.edu.cn

Abstract. The knowledge graph, as a significant branch of Library and Information Science, has also become a vital part of Computer Science. In a bid to reveal the research trend and hot topics of the knowledge graph domain, this paper collected latest data about literature on knowledge graphs from SCI-Expanded and EI Compendex database. This paper took advantage of Citespace for visualization analyses of data, which discerned the research topics, hot issues and theme changes of the knowledge graph domain over time. Through analyses, we conclude that the knowledge graph has developed greatly in both Library and Information Science and Computer Science. In Library and Information Science domain, researches focus on the plotting of knowledge graphs and visualization of knowledge. In Computer Science domain, hot issues are the techniques of constructing knowledge graphs, which include entity recognition, knowledge integration, etc., and their applications in Semantic Web. Among all the research topics, the domain knowledge graph is one of the paramount research trends. Compared with previous researches, this paper keeps up with the times and can reveal the theme evolution of researches on knowledge graphs more precisely.

Keywords: Knowledge graph; Analyses of theme evolution; Citespace; Visualization.

1. Introduction
With the advent of the knowledge economy era, researchers’ attention is no longer limited to knowledge itself, but is increasingly paid to the connections between knowledge.[1] Knowledge graphs are emphasized by more and more researchers. This is reflected by the phenomena that an increasing number of researchers participate in the researches of knowledge graphs, achieving abundant results, and the application fields continue to be broadened.

Analyses of theme evolution of a domain can reveal the hot issues, frontier fields and evolution of themes over time. Several researchers have previously studied the development of knowledge graphs. For example, Ruibin Wei (2010), basing the research on the data from CNKI, WANFANG and CQVIP, as well as other tools such as UCINET, visually analysed domestic researches on knowledge graphs.[2] Siluo Yang (2012) collected data from CNKI, analysed in terms of publication years, journals, authors, and institutions of related papers, and drew the collaboration network in the field of knowledge graphs.[3] Shujin Cao (2015) explored the trend of knowledge graph research using data from SSCI and CSSCI.[4] Ni Zhang (2017) took advantage of Citespace to show the hot issues and trends of the knowledge graph domain based on data from Web of Science and CSSCI.[5] However, since recent years has witnessed the profound development of knowledge graphs, previous researches cannot reflect the current status of the knowledge graph domain precisely. Moreover, previous researches were almost limited to Library and Information Science domain but ignored the development and application of knowledge graphs in Computer Science domain. Consequently, this paper leverages latest data, conducts a comprehensive and up-to-date analysis of the development of
knowledge graphs, which is intended to help researchers grasp the latest development trend of the knowledge graphs. This paper will first introduce the concept of the knowledge graph and then use Citespace to visualize the research progress and theme evolution of knowledge graphs based on the data obtained from SCI-Expanded and EI Compendex database, and analyse its development trend.

2. The Origin and Concept of Knowledge Graphs

The concept of the knowledge graph can be traced back to the 1950s. In 1955, Eugene Garfield published the paper concerning citation indexing in *Science*, establishing foundations of the citation analysis.

In 1961, D Price proposed quantitative statistical models to research the scientific knowledge and found that the models presented exponential growth.

Such a kind of method to present the development of scientific knowledge with graphs effectively became the initial knowledge graph. In 1964, Garfield created the Science Citation Index, making it possible to conduct efficient and scientific citation analyses.

With the development of computer science, visualization technology and scientometrics, the knowledge graph has become one of the hottest issues in the field of Bibliometrics nowadays. The knowledge graph uses the method of information visualization, citation analysis and scientometrics to mine, analyse and present the knowledge and its connections.

It is possible to reveal inner connection of the large-scale scientific knowledge via knowledge graphs and Citespace, the typical application of knowledge graphs which is used as the major analysing tool in this paper. Meanwhile, the knowledge graph is also introduced to the domain of computer science thanks to the development of ontology and semantic web. For example, Google redefined the concept of the knowledge graph in 2012 and has made great progress. In the field of Computer Science, the knowledge graph is considered as knowledge bases with directed graph structures and some technology such as knowledge fusion and entity recognition is used to construct knowledge graphs. Also, the knowledge graph is applied to smart search, construction of the knowledge base, knowledge engineering and so on.

3. Data Collection and Cleaning

Bibliometrics and thematic analysis are the main methods when analysing a specific domain and evaluating development trends.

Citespace, which is developed by Chaomei Chen, can perform visual analyses of the documents data from many dimensions.

We chose SCI-Expanded and EI Compendex as the data source of this passage because the knowledge graph per se is of practical use and its theories mainly involve Library and Information Science and Computer Science. Also, since the papers in the field of Computer Science prefer to be published in the top conferences first, we chose 10 top conferences involving knowledge graphs according to the research report *Research Report of Knowledge Graph* written by Aminer.

They are ACL, EMNLP, WWW, ISWC, IJCAI, AAAI, COLING, KR, KDD and CIKM. According to the definition of the knowledge graph in a narrow sense proposed by Siluo Yang, we selected “knowledge map, knowledge domain map, knowledge graph” as the search terms of knowledge graphs. Zeyuan Liu reckoned that though the knowledge map can be seen as one form of the knowledge graph, researches on traditional knowledge maps and their application are mostly limited to the aspect of management. However, compared with knowledge maps, knowledge graphs can exhibit mutual connections of knowledge and represent knowledge and semantics better. As a consequence, we manually exclude the documents concerning traditional knowledge maps when cleaning the data.

The time of retrieval was October 28, 2020. The search query in SCI-Expanded was as follows: (TI=knowledge map* OR knowledge domain map* OR knowledge graph*) AND (Document Types=Article), timespan=all years. After retrieval, we got 1104 records.

The search query in EI Compendex was as follows: Subject/Title/Abstract for TI= (knowledge map* OR knowledge domain map* OR knowledge graph*) AND Conference information for (ACL OR EMNLP OR WWW OR ISWC OR IJCAI OR AAAI OR COLING OR KR OR KDD OR CIKM).

After retrieval, we got 3231 records.

After data cleaning, deduplication and merging, we got 1211 relevant documents.
The earliest document is *Knowledge Map - An Approach to Knowledge Acquisition in Developing Engineering Expert Systems* written by LU, SCY in 1987. At this time, although the word “knowledge graph” had not been proposed formally, the research had contained mechanism and mapping skills of knowledge graphs. After that, papers in the field of knowledge graphs kept accumulating. Chronological distribution of papers is presented in figure 1. As it shows, the number of papers about knowledge graphs is still mounting.

**Figure 1.** Chronological Distribution of Papers.

The institution distribution of papers is more scattered. Most of the institutions published one or two papers and the institution which is leading in the number of publication is Metaphacts GmbH. Metaphacts GmbH published 4 papers, followed by University of Mannheim and Tsinghua University, both of which published 3 papers.

1211 papers are from 4011 authors. table 1 lists top 10 authors in terms of the quantity of papers.

| Author             | The Quantity of Papers | Institution                  | Author          | The Quantity of Papers | Institution               |
|--------------------|------------------------|-------------------------------|-----------------|------------------------|---------------------------|
| Lehmann, Jens      | 17                     | University of Bonn            | Stepanova, Daria| 8                      | Chuvash State University  |
| Zhiyuan Liu        | 15                     | Tsinghua University           | Wang J.         | 8                      | Beihang University        |
| Weikum, Gerhard    | 13                     | Max Planck Institute for Informatics | Juanzi Li     | 8                      | Tsinghua University       |
| Urbani, Jacopo     | 11                     | Vrije University              | Krotzsch, Markus| 8                      | TU Dresden                |
| Maosong Sun        | 8                      | Tsinghua University           | McGuinness, Deborah L. | 8                      | Rensselaer Polytechnic Institute |

4. Analyses of Themes

4.1. Keyword Co-occurrence and Keyword Clusters

Keywords are an abstract summary of literature and can precisely reveal the research issues. Co-occurrence analyses and cluster analyses of keywords in one particular domain assist in learning about the research topics and hot issues of the domain. This passage leverages Citespace to conduct keyword co-occurrence analyses and keyword cluster analyses, respectively shown in figure 2 and figure 3. In figure 2, every single node denotes a keyword and links between nodes represent the co-occurrence of
keywords. The bigger the node is, the more frequently the keyword which it denotes appears in the literature. In figure 3, different colour blocks stand for different research topics, beside which are the labels of the clusters, summarizing the corresponding research topics. Table 2 shows the top 20 keywords in terms of frequency.

![Figure 2. Keyword Co-occurrence.](image)

**Table 2.** Top 20 Keywords in Descending Order of Frequency.

| Count | Keyword                              | Count | Keyword                        |
|-------|--------------------------------------|-------|--------------------------------|
| 145   | knowledge graph                      | 13    | network                        |
| 27    | ontology                             | 13    | semantics                      |
| 26    | web                                  | 13    | knowledge management           |
| 25    | knowledge graph embedding            | 13    | knowledge mapping              |
| 23    | system                               | 12    | neural network                 |
| 21    | management                           | 11    | database                       |
| 21    | knowledge map                        | 11    | link prediction                |
| 18    | semantic web                         | 11    | recommender system             |
| 15    | framework                            | 11    | Citespace                      |
| 13    | visualization                        | 11    | knowledge graph completion     |
| 13    | knowledge representation             |       |                                |

![Figure 3. Keyword Cluster.](image)
As is shown in figure 2, hot keywords in knowledge graph domain encompass knowledge graph, ontology, web, knowledge graph embedding, etc. The keyword cluster analysis divides all the keywords into 13 categories, covering research fields such as ontology, knowledge engineering, knowledge base, knowledge management, research trends, information visualization and natural language processing.

Based on the visualization results and extensive reading of relevant literature, we conclude that the hot topics of knowledge graph domain mainly have the following three aspects.

1. Research on the theories and the drawing methods of knowledge graphs in the field of scientometrics. In the scientometrics domain, researchers maintain that knowledge graphs visualize the knowledge resources and their carriers human beings possess over time, analyse and reveal the knowledge together with its inner relationships, and create an atmosphere where knowledge is shared aiming to contribute to scientific research and collaboration.[8] For example, a host of researches focus on topics such as knowledge map, knowledge mapping domain, information visualization and bibliometric analysis.

2. Research on the technology of constructing knowledge graphs and their applications in the semantic web in the field of Computer Science. In Computer Science domain, the knowledge graph, expressing semantics, is defined as a semantic network which is capable of showing relationships among entities.[14] Many researchers take advantage of computer technology such as knowledge representation, natural language processing, to construct knowledge graphs. What is more, some researchers successfully applied knowledge graphs to computer science in various aspects, including recommender systems and knowledge engineering. For instance, Bounhas(2020) built a morpho-semantic knowledge graph serving the purpose of Arabic information retrieval.[15] Bhatt(2020), intending to improve AI, gave input data semantics with the assistance of knowledge graphs.[16]

3. Domain knowledge graphs and interdisciplinary application of knowledge graphs. Due to the ability to show development process and knowledge structures of disciplines, the knowledge graph has been put into widespread use in many domains. In figure 3, #cluster 8 is diseases and #cluster 9 is geo-ontology, which show the application of knowledge graphs in geography and medicine. Feicheng Ma(2014), basing the research on PubMed database, built knowledge maps of global translational medicine research and revealed the publication trends.[17] Jonas Jetschmi(2017) constructed enterprise knowledge graphs, which were expected to increase the ability of organization management.[18]

4.2. Keyword Bursts and Timeline Analyses

Dynamic changes of keywords of a field reflect the changes of research topics over time. Keyword bursts analyses and keyword timeline analyses can show the changes in research topics in a field. We drew the pictures of keyword bursts and the keyword timeline in knowledge graph domain using Citespace.

![Keywords with the Strongest Citation Bursts](image-url)
As is shown in the figures, in early times, researches on knowledge graphs focused on constructing knowledge graphs and applying them to knowledge management, information systems, etc. In 1998, Neches introduced the concept of ontology into computer science domain. Meanwhile Berners Lee put forward the concept of semantic web and expected to incorporate semantic elements into web pages. Moreover, some scholars attempted to combine knowledge graphs with ontology and semantic web for the purpose of constructing formal and standardized knowledge graphs that can express semantics between entities. A vital turning point in the field of knowledge graphs was that in 2012, Google proposed a semantic-based knowledge graph aimed at improving search engine capabilities. After that, knowledge graphs play an increasingly significant role in various fields such as information retrieval, machine translation and natural language processing. Also, more scholars concentrate on the construction techniques of knowledge graphs such as link prediction, linked data, entity identification and knowledge fusion.

5. Discussion and Prospect

5.1. The Knowledge Graph Domain will be Deeply Integrated with the Computer Science Domain
From the previous analyses, we can learn that the concept of semantic web and ontology has changed the research objects in the field of traditional knowledge graphs drastically. According to figure 3 and figure 4, currently, an array of researches are equipped with assorted computer science technology, including linked data, natural language processing, knowledge reasoning, computational modeling and relation extraction. Besides, knowledge graphs have excellent applications in computer science and artificial intelligence domain such as machine translation, intelligent retrieval, constructing dbpedia and deep learning. For example, Cudre-Mauroux(2020) applied knowledge graphs to big data integration. Hu X(2020) used knowledge graphs to optimize scalable aggregate keyword query. With the development of computer technology and the development of the application requirements of knowledge graphs, the construction of knowledge graphs in the future will rely on advanced computer technology, and there will also be more applications in the computer field.

5.2. Knowledge Graphs will have Broader Applications in Different Disciplines
The knowledge graph can vividly convey information such as subject structures, research frontiers, and research hot issues. At present, the knowledge graph has been applied in many fields such as geography, medicine, education, and sociology. In the future, with the advance in technology such as the semantic web and knowledge fusion, the knowledge graph can express semantics more accurately and adapt to the needs of specific fields better, thus playing its role in more fields.
5.3. Knowledge Graphs will be Confronted with Challenges such as Data Heterogeneity and Immature Construction Technology

Ni Zhang (2017) reckoned that data heterogeneity and complex knowledge structures may hinder the development of knowledge graphs since constructing knowledge graphs needed to collect data from multiple databases.[5] Guilin Qi (2017) contended that data quality in big data times was poor and the lack of relevant tools such as dictionaries would multiply the difficulty of knowledge graph construction.[9] Zhonggui Ma (2020) pointed out the challenges in terms of construction technology.[22] For instance, knowledge fusion technology is lack of dynamic evaluation and natural language is hard to process. Knowledge reasoning technology can hardly handle multiple relationships. As a consequence, there is a long way to go before the knowledge graph reaches its maturity.

6. Conclusion

Based on the bibliometric method and topic analysis method, this paper collects the latest data and analyses the evolution of research topics in the field of knowledge graphs more comprehensively. This paper reveals the research hot issues, topic evolution and development trends in the field of knowledge graphs. Recent years has witnessed the upsurge of literature on the knowledge graph and its remarkable development in various domains. In the field of Library and Information, the hot topics of knowledge graph are knowledge graph drawing, knowledge visualization, bibliometric analyses and so on. In the field of Computer Science, the researches on knowledge graph focus on the technology of knowledge graph construction, such as entity recognition, knowledge fusion, linked data, etc., and its application in the semantic web. At the same time, the construction of domain knowledge graphs has become a fundamental direction in the current knowledge graph field. The limitation of this article is that it only analyses the keywords of the data collected, but does not analyse other aspects such as the author cooperation network and the organization cooperation network. Future research will conduct a more comprehensive analysis, so as to understand the development status of the knowledge graph field more thoroughly.

Acknowledgments

This paper is supported by the key project Cultivation of Innovative Abilities to Discipline Competition under the Big Data Environment (SY201919) of Nanjing University’s “The 13th Five-Year Plan” experimental teaching reform research.

References

[1] Lixin Xia, Kaili Wang and Xiufeng Cheng 2019 A Visual Analysis of Evolutional Characteristics of Knowledge Mapping in China Information Science vol 37, No.3 pp 9-16
[2] Ruibin Wei 2011 Visualization Analysis of the Achievements in Mapping Knowledge Domain in China Library and Information Service vol 55, No.8 pp 126-130
[3] Siluo Yang and Ruizhen Han 2012 A Visual Analysis of the Status Quo and Trend of Knowledge Mapping Research Information and Documentation Services No.4 pp 22-28
[4] Shujin Cao, Yubing Wu, Jingzhu Wei and Cuichang Ma 2015 History, Schools and Trend in Knowledge Map: Investigation and Visualization Based on SSCI and CSSCI Journal of Library Science in China vol 41, No.219 pp 16-34
[5] Ni Zhang and Jingyuan Wang 2017 Knowledge Mapping Based on Citespace: Research Hotspots Analysis at Home and Abroad and Situation Prospects Information and Documentation Services No.3 pp 33-41
[6] Yue Chen, Zeyuan Liu, Jin Chen and Jianhua Hou 2008 History and theory of mapping knowledge domains Studies in Science of Science vol 26, No.3 pp 449-460
[7] D Price 1961 Science Since Babylon (New Haven: Yale University Press)
[8] Changjiang Qin and Hanqing Hou 2009 Mapping Knowledge Domain-A New Field of Information Management and Knowledge Management Journal of Academic Libraries vol 21, No.1 pp 30-37,96
[9] Guilin Qi, Heng Gao and Tianxing Wu 2017 The Research Advances of Knowledge Graph Technology Intelligence Engineering vol 3, No.1 pp 4-25
[10] Liang Zhu and Xianxue Meng 2013 The Comparative Study on Bibliometric Method and Content Analysis Method Library Work and Study Serial 208 pp 64-66

[11] Chen CM 2006 CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature Journal of the American Society for Information Science and Technology vol 57, No.3 pp 359-377

[12] Research Report of Knowledge graph (Beijing: AMiner) 2019 No.1

[13] Yue Chen and Zeyuan Liu 2005 The rise of mapping knowledge domain Studies in Science of Science vol 23, No.2 pp 149-154

[14] Zenglin Xu, Yongpan Sheng, Lirong He and Yafang Wang 2016 Review on Knowledge Graph Techniques Journal of University of Electronic Science and Technology of China vol 45 pp 589-606

[15] Ibrahim Bounhas, Nadia Soudani, and Yahya Slimani 2020 Building a morpho-semantic knowledge graph for Arabic information retrieval Information Processing & Management vol 57, No.6 pp 102-124

[16] Shreyansh Bhatt, Amit Sheth, Valerie Shalin and Jinjin Zhao 2020 Knowledge Graph Semantic Enhancement of Input Data for Improving AI IEEE Internet Computing vol 24, No.2 pp 66-72

[17] Feicheng Ma, Penghui Lyu, Qiang Yao, Lan Yao and Shijing Zhang 2014 Publication trends and knowledge maps of global translational medicine research Scientometrics: An International Journal for All Quantitative Aspects of the Science of Science Policy vol 98, No.1 pp 221-246

[18] Jonas Jetschni, Vera G. Meister 2017 Schema engineering for enterprise knowledge graphs: A reflecting survey and case study 8th International Conference on Intelligent Computing and Information Systems, ICICIS 2017, pp 271-277

[19] Xiao Tan and Zhiqiang Zhang 2020 Research Progress and Subject Analysis of Knowledge Graph Library and Information No.2 pp 50-63

[20] Cudre-Mauroux, P (Cudre-Mauroux, Philippe) 2020 Leveraging Knowledge Graphs for Big Data Integration: the XI Pipeline Semantic Web vol 11 pp 13-17

[21] Hu, X (Xin Hu), Duan, JL (Jiangli Duan), Dang, DP (Depeng Dang) 2020 Scalable aggregate keyword query over knowledge graph Future Generation Computer Systems vol 107 pp 588-600

[22] Zhonggui Ma, Runyu Ni and Kaihang Yu 2020 Recent advances, key techniques and future challenges of knowledge graph Chinese Journal of Engineering vol 42, No.10 pp 1254-1266