

Article

Topical Evolution and Thematic Progression of Research Frontiers: The Field of Knowledge Graphs

Elaheh Hosseini^{1,*}, Alireza Shahbazi², Alireza Dehbozorgi³
¹Department of Information Science, Faculty of Education & Psychology, Alzahra University, 19958-13151 Tehran, Iran

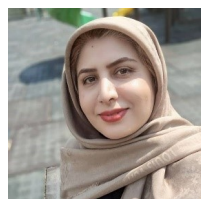
²Data Mining Research Laboratory, Iran University of Science and Technology (IUST), 16849-33177 Tehran, Iran

³Department of Linguistics, Faculty of Humanities, Tarbiat Modares University, 14115-141 Tehran, Iran

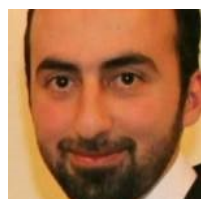
*Correspondence: E.hosseini@alzahra.ac.ir (Elaheh Hosseini)

Academic Editor: Natália Tognoli

Submitted: 6 July 2024 Revised: 12 January 2025 Accepted: 13 February 2025 Published: 3 April 2025



Dr. Elaheh Hosseini is an Assistant Professor of Information Science in the department of Information Science, housed within the faculty of Education & Psychology at Alzahra University in Tehran, Iran. She earned her Ph.D. from Alzahra University. Her research interests are diverse range of interests, focusing on knowledge graphs, linked data, semantic web, knowledge retrieval, scientometrics, social network analysis, trend analysis, and AI. Moreover, she is interested in the intersection of AI and knowledge representation and reasoning. In addition to her research, she is recognized for her role in advancing the field through her editorial work and active participation in professional associations.



Alireza Shahbazi is a knowledge representation and reasoning specialist. He works as senior researcher in data mining research laboratory at Iran University of Science and Technology (IUST) in Tehran, Iran. He is founder and CEO of Borhan company, which specializes in the development of semantic web technologies and advanced reasoning systems. His research expertise encompasses areas such as the semantic web, semantic reasoning, and knowledge retrieval, with a strong emphasis on integrating multidisciplinary knowledge—including philosophy, logic, programming, and AI—into both the services and products offered by his company as well as his academic pursuits.



Alireza Dehbozorgi is a dedicated researcher specializing in artificial intelligence, linguistics, and neural machine translation at LingoAI, with a profound curiosity for understanding the mathematical and computational foundations of languages. His work primarily centers on advancing the fields of natural language processing (NLP), natural language understanding (NLU), Good Old-Fashioned Artificial Intelligence (GOFAI), and Artificial General Intelligence (AGI). His research interests extend to knowledge graphs, knowledge representation, and computational linguistics.

Abstract

The paper aimed to analyze the thematic growth and topical evolution of the intellectual structure in Knowledge Graphs (kGs) during 2013–2023. This applied research used an analytical and descriptive method, co-word techniques, scientometric indicators, and social network analysis. A web-based interface of Bibliometrix, Python, Microsoft Excel, VOSviewer, UCINET, and SPSS was used for descriptive bibliometric study, data analysis, and network structure visualization. China was the most prolific country. Nine major topic clusters were identified based on the co-occurrence network. The most mature and mainstream thematic cluster was the ‘application of knowledge-based systems’. Six clusters were identified in the network structure: ‘knowledge graph’, ‘knowledge engineering’, and ‘knowledge discovery’ as niche themes, and ‘ontology’, ‘semantic web’, and ‘linked data’ as basic themes. Moreover, six main cluster evolutions during 2021–2023 were identified: ‘ontology’, ‘natural language processing’, ‘machine learning’, ‘protein’, and ‘knowledge engineering’, and ‘knowledge graph’.

Keywords: knowledge graphs; thematic clusters; words’ co-occurrence; research frontiers; thematic progression; topical evolution



1. Introduction

Knowledge graphs (KGs) have become crucial to contemporary information retrieval and knowledge-driven artificial intelligence (AI). KGs are structured representations of knowledge, including entities (nodes) and relationships (edges). A KG is a labeled graph wherein meanings are interlinked to these entities and links from a particular domain. The formal semantics empower computers to process these entities and links meaningfully and unambiguously (Chaudhri et al, 2022; Peng et al, 2023; Tian et al, 2022). The concept of KG can be traced back to the 1960s and 1970s when scholars started probing semantic networks (Simsek et al, 2023) to represent knowledge in a structured format, such as the semantic memory proposed by Quillian (Kumar et al, 2022). During the 1980s, researchers developed formalisms for representing knowledge, including Minsky's frame theory (1975) (Löebner, 2021). KL-ONE, a frame-based knowledge representation language, was the first attempt at creating a standard for providing and reasoning about knowledge (Brachman and Schmölze, 1985). The idea of the Semantic Web emerged in the late 1990s and early 2000s to make the World Wide Web more machine-readable, interrelated, and interconnected (Hall and Tiropanis, 2012). Key milestones include the Resource Description Framework (RDF) in 1999 (Lassila and Swick, 1998) and the Web Ontology Language (OWL) in 2004 (McGuinness and VanHarmelen, 2004). The initial large-scale KGs were created in the 2000s, significantly impacting AI and information retrieval. As an early, community-driven KG, DBpedia aimed to extract structured information from Wikipedia and embody it in RDF format in 2007 (Morsey et al, 2012).

Google Knowledge Graph was launched in 2012 by Google to improve search results by presenting more relevant and structured search hits. The Google search engine now relies heavily on it (Oramas et al, 2016).

This long and rich history of KGs necessitates a deep understanding and study of the scientific approaches, changes, trends, and developments in this field over time; in this way, the thematic growth and maturity of this domain can be evaluated in various thematic frontiers through multidimensional analysis. Understanding past trends, current developments, and topical evolution can lead to greater awareness for policymakers and funders to make data-driven and knowledge-based decisions and for semantic developers to adopt a more informed approach to the fast-paced and dynamic growth of technology to develop predictive semantic tools. Stakeholders and researchers can also prevent redundant research and push the boundaries of knowledge with new global trends and emerging themes. Creating such a conscious interaction and professional collaboration among these stakeholders leads to greater scientific convergence, more in-depth critical decisions, and more practical technologies; these outcomes enhance the quantitative and qualitative aspects of these trends and

strengthen the connection and development of knowledge and society. Therefore, this study aimed to analyze the thematic growth and topical evolution of research frontiers related to KGs from 2013 to 2023. The paper's originality and novelty lie in using different methods and mapping tools simultaneously, such as co-word analysis, scientific mapping, social network analysis (SNA), network structure visualization, hierarchical clustering, Biblioshiny, and VOSviewer (1.6.19, <https://www.vosviewer.com/>), to provide an appropriate, illustrative, systematic, and inclusive approach to KGs.

The following questions were addressed in this research:

Q1: What are the top scientific publications on KGs in terms of the main information related to the year, source, country, affiliation, and author?

Q2: How is the intellectual structure of KGs yielded by VOSviewer and analyzed in terms of network structure and overlay visualization based on the co-occurrence algorithm?

Q3: How is the KG cluster analysis in terms of the factorial analysis of the conceptual structure map method?

Q4: How are the KG clusters in terms of the thematic map?

Q5: How are the VOSviewer clusters of KGs in terms of maturity and development visualized and analyzed by the strategic diagram (SD)?

Q6: How are the KG clusters in terms of thematic evolution?

2. Literature Review

The following provides brief definitions of the concepts used in the query. KGs are structured representations of information that organize data into entities and relationships that are foundational in semantic web technologies. They enable data to be presented in a format that is human-readable and machine-interpretable (Verma et al, 2023). The construction of KGs is designed, curated, and optimized by knowledge engineers to ensure accuracy and usability of the system. They focus on knowledge representation, which involves creating formal structures, such as ontologies, to encode information in a machine-readable format that captures the meaning and interrelationships of concepts (Pan et al, 2024).

Ontologies, a core component of knowledge representation, define the hierarchical structure and semantics of concepts within a domain, enabling machines to reason and infer new knowledge. They provide a formal framework for representing entities, their attributes, and the relationships between them. By standardizing the vocabulary and structure of a domain, ontologies facilitate interoperability and support the integration of diverse data sources. This structured representation aligns closely with the goals of the semantic web, which seeks to make web data interconnected and machine-readable (Guizzardi and Nicola, 2024).

The most important part of these are technologies like RDF, which encode data as triples (subject, predicate, object) to establish relationships between entities. The functionality of the linked data approach is driven by the capability to connect and query structured data on a large scale. Together, these approaches provide the foundation for creating and utilizing knowledge graphs in intelligent systems (Lissandrini et al, 2025).

A knowledge base is a centralized repository of structured information designed to store facts, rules, and relationships about a specific domain. To update a knowledge base, processes like knowledge extraction and knowledge discovery are essential (Fan et al, 2021). Knowledge extraction involves identifying and structuring information from unstructured or semi-structured sources. It often utilizes Natural Language Processing (NLP) and machine learning techniques to extract entities, relationships, and attributes. Knowledge discovery, on the other hand, focuses on uncovering hidden patterns and employs data mining, statistical analysis, and AI-driven methods to reveal new information that can enrich the knowledge base. Underlying these processes is the semantic data model, which provides a framework for representing and organizing data with a focus on meaning and relationships (Sirichanya and Kesorn, 2021).

A bibliometric approach can facilitate the analysis and visualization of knowledge development. There are numerous indications that this method can be employed effectively to explore the advancement of a field within a particular domain of knowledge (Börner et al, 2003). Examples include a recommendation system from the perspective of KG (Shao et al, 2021), the Internet of Things (Jaya et al, 2022), iMetrics (Khasseh et al, 2017), AI related to the tourism industry (Kong et al, 2023), humanoid robot technology (Kumari et al, 2019), data mining and machine learning techniques (DosSantos et al, 2019), and educational data mining (Baek and Doleck, 2022).

Some studies have also been conducted on KG-related subfields using bibliometric approaches or co-word analysis, such as semantic web (Ding, 2010), ontology (Wu and Ye, 2021; Zhu et al, 2015), and linked data (Hosseini et al, 2021, 2023; Kyaw and Wang, 2018; Liu X and Liu Z, 2019).

Chen et al (2021b) conducted topic analysis in KG research over 3 decades using bibliometrics and topic modeling. Their findings showed that KG embedding is the most researched topic, followed by a search based on KGs and KGs for intangible cultural heritage. The findings indicated that China is the most prolific country, and the Chinese Academy of Sciences is the most prolific institution. Significant research topics included *heterogeneous KG embedding*, *KG construction and completion*, *knowledge reasoning over KGs*, *KG representation*, *relation extraction for KG construction*, *application domains*, *entity profiling*, and *semantic similarity measures*.

The main difference between the cited paper and the present study lies in several factors, including the data, tools, and query for data extraction, period, and data analysis methods. Although both papers used data from the Web of Science (WoS), our paper focused on the core collection and some subject areas, while the cited paper only focused on the Science Citation Index (SCI), Social Science Citation Index (SSCI), and all subject areas. The query constructed by our study was more complex and detailed. Our paper covered the recent decade (2013–2023), while the cited paper covered a longer period (1991–2020). The cited study conducted its analysis using the Structural Topic Model (STM) and SNA, while our analysis concentrated on the co-occurrence of terms using tools such as VOSviewer and Biblioshiny. Therefore, these two studies have different perspectives on KGs.

A study (Chen et al, 2021a) concluded that KGs have faced significant developments in library and information sciences and computer sciences. Research primarily centers on KG visualization in library and information sciences. In computer sciences, constructing KGs is a popular study area involving techniques such as entity recognition and knowledge integration, along with their applications in the semantic web. Among all the research topics, constructing domain KGs stands out as one of the most prominent research trends.

Ji et al (2021) conducted a comprehensive survey of research on KGs, covering major topics such as KG representation learning, knowledge acquisition and completion, temporal KGs, and knowledge-aware applications. They provided an overview of existing achievements and proposed directions for future KG research. Their review also highlighted the significant efforts made to address challenges in knowledge representation. Their study identified several issues in the field, such as interpretability, complex reasoning, scalability, knowledge aggregation, and automatic construction and dynamics.

Turki et al (2021) analyzed scholarly research on open KGs over a decade (2013–2023) from Scopus. The findings showed that the number of publications has steadily increased each year. They identified three themes in this field: creation and enrichment of KGs, evaluation and reuse of these graphs, and integration of KGs into natural language processing systems. Additionally, the study highlighted specific tasks that had received significant attention, such as entity linking, KG embedding, and graph neural networks.

Salatino et al (2021) proposed a new approach to identifying, examining, and predicting research subjects using a comprehensive KG that categorizes research papers based on their topics in the Computer Science Ontology. The cited study outlined how this method was utilized to create bibliometric analysis and novel tools for studying and forecasting research trends. Topics like information retrieval, AI, data mining, computer vision, semantic web, and image qual-

ity were categorized based on the performance of random forest and ordered by F1 score.

Based on the literature review, it can be concluded that KG has been the focus of study for bibliometric experts due to its importance and interdisciplinary nature. However, the main difference between the present and previous studies is that this study performed a more comprehensive and multidimensional examination of this subject area. The goal was to represent thematic clusters simultaneously and, more precisely, using multiple bibliometric tools and techniques to measure the growth and maturity of this field over a decade. Besides, the capabilities of various tools in analyzing and visualizing thematic topics were demonstrated.

3. Methodology

This applied study was conducted using a descriptive and analytical approach, scientometric techniques, co-word analysis, and SNA. The research population included all the keywords extracted from all the documents about KGs indexed in the WoS during 2013–2023.

Two domain experts initially compiled a list of seed keywords related to KGs, including terms like ‘knowledge graph’, ‘ontology’, and ‘semantic web’. This set of keywords was used to query and retrieve papers featuring these terms in their titles, abstracts, or author-defined keywords. Next, we extracted all author-defined keywords from the highly cited papers identified. These keywords were reviewed by the domain experts, who removed any irrelevant terms. The remaining relevant keywords were then incorporated into the final keyword query for KGs. The following researcher-made query was used in an advanced search in the WoS Core Collection from Clarivate Analytics on April 5, 2023. As a result, 13,662 documents were retrieved. The keyword column was then unmixed, refined, and cleaned for preprocessing.

TS = (“knowledge graph **” OR “knowledge engineer**” OR “knowledge represent” OR “knowledge base” OR “knowledge extraction” OR “knowledge discovery” OR “knowledge embedding” OR “ontolog**” OR “semantic web” OR “linked data” OR “RDF**” OR “semantic data model**”) and 2023 or 2022 or 2021 or 2020 or 2019 or 2018 or 2017 or 2016 or 2015 or 2014 or 2013 (Publication Years) and Article (Document Types) and English (Languages) and Science Citation Index Expanded (SCI-EXPANDED) or Social Sciences Citation Index (SSCI) or Arts & Humanities Citation Index (A&HCI) (Web of Science Index) and Computer Science or Information Science Library Science or Medical Informatics (Research Areas)

The authors legally accessed the WoS database through their institution. While there was no definitive scientometric data source, the WoS was selected due to its extensive coverage of global scientific content from 1900 onwards, providing a wealth of valuable information. The study of science, technology, and knowledge has greatly benefited from its interdisciplinary coverage, making the

WoS a useful resource for appropriate literature. The WoS Core Collection (All Indexes) offers access to the best scholarly publications across various subjects, including humanities, social sciences, arts, and sciences (Clarivate, 2023).

Bibliometric analysis is a valid approach to classify and quantitatively evaluate the bibliographic content within a specific scientific field (Lazarides et al, 2023). Co-word analysis is a helpful method for analyzing the relationships between words in a document and can provide valuable insights into the intellectual structure of a particular research area. It is a quantitative method to analyze the semantic relationships between words in a text. It is often used in scientometrics to identify co-occurrence patterns of terms in scientific research papers (Chen et al, 2019). This method can be used to map out the intellectual structure of a particular research area and identify the most important topics and themes (Cobo et al, 2011). Co-word analysis is implemented within a co-word matrix by using mathematical algorithms to identify thematic clusters in a field. The value assigned to a cell representing two words is determined based on the number of times these two words appear together in a single document. A higher frequency of co-occurrence indicates a closer relationship between the two words (Cho, 2014). Co-word analysis represents emerging and developed thematic clusters, which can provide insight into the potential direction of future research (Mokhtarpour and Khasseh, 2021).

The R program (4.3.0, <https://www.r-project.org/>) and Biblioshiny, a web-based interface of Bibliometrix (4.1.2, <https://www.bibliometrix.org/>), were used to conduct a descriptive bibliometric study (Aria and Cuccurullo, 2017).

VOSviewer is software for generating visual maps from network data and facilitating the identification of links between concepts within clusters. Although originally developed for bibliometric network analysis, it can be utilized with any form of network data. The software allows for the construction of these maps and their visual exploration (Van Eck and Waltman, 2014).

SNA is a method for studying social structures and relationships between nodes (Kumari et al, 2019). Several measures and metrics can be used in SNA, such as centrality and density (Isfandyari et al, 2023). In the SD, two axes show density and centrality: the x-axis indicates centrality, and the y-axis represents density. Consequently, the diagram is divided into four quadrants, displaying different degrees of centrality and density (Hu et al, 2013). A high level of centrality indicates that a cluster holds a more significant position within the field. As demonstrated in Fig. 1, the first quadrant consists of mature clusters situated at the center of the field due to their high levels of centrality and density. This means they exhibit strong internal correlations and maturity and have an extensive and robust relationship with other clusters. The second quadrant consists of well-developed but isolated clusters with low centrality

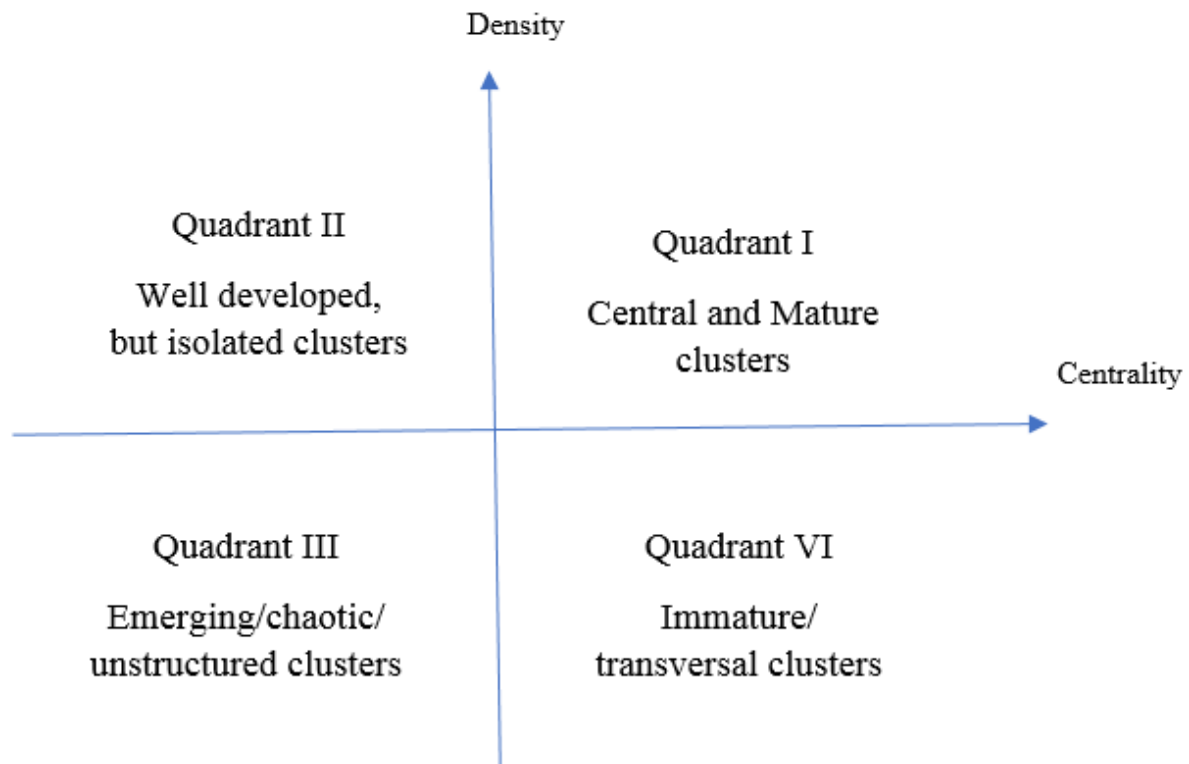


Fig. 1. Quadrants in a strategic diagram (Hu et al, 2013).

Table 1. Main information.

Description	Results
Sources (journals, books, etc.)	556
Documents	13662
Annual growth rate %	-18.51
Document average age	4.92
Average citations per doc	13.47
References	1
Keywords plus (ID)	8106
Author's keywords (DE)	29403
Authors	31813
Authors of single-authored docs	723
Single-authored docs	823
Co-authors per doc	3.97
International co-authorships %	31.09
Article	13038
Article; book chapter	11
Article; data paper	5
Article; early access	239
Article; proceedings paper	364
Article; retracted publication	5

and high density. It contains clusters that are not central but have progressed well. This suggests that these clusters are not pivotal but evolving. Clusters in the third quadrant have low centrality and low density, indicating emerging or declining themes with little attention. They display a relatively fragmented structure and are underdeveloped. Fi-

nally, the fourth quadrant comprises central clusters that are not yet developed. They have high centrality but low density and are immature due to their central position within the field (Hu et al, 2013; Khasseh et al, 2017).

Adjacency matrices, as co-occurrence square matrices, were exported by a code in NumPy and Panda libraries in Python (3.10.9, <https://www.python.org/>) (Python Software Foundation, 2023). Then, correlation matrices were generated as square matrices by SPSS (22.0, IBM SPSS Statistics, Armonk, NY, USA) (IBM Corp, 2013). Next, UCInet software (6, <https://sites.google.com/site/ucinetsoftware/home>) was utilized to measure the centrality and density of each cluster (Borgatti et al, 2002). Finally, an SD was created using Microsoft Excel (2016, Microsoft, Albuquerque, NM, USA) based on these metrics.

Factorial analysis is a statistical method used to analyze relationships between variables. Multiple correspondence analysis (MCA) is a type of factorial analysis; it is a multivariate statistical technique used to analyze relationships between categorical variables to identify patterns or trends in the data (Greenacre and Blasius, 2006).

One way to create thematic maps or thematic evaluations is by using clustering algorithms, such as the Walktrap clustering algorithm, a hierarchical clustering algorithm used to identify communities or clusters within a network. It is based on the idea that connected nodes are more likely to belong to the same cluster than unconnected ones. The algorithm works by iteratively merging connected nodes until a desired level of clustering is achieved.

Thematic maps can provide valuable insights into the spatial patterns and trends within a particular theme or topic (Brusco et al, 2022). Fig. 2 indicates the schematic methodology, including steps, methods, and tools.

4. Results

Table 1 presents the main information collected from the WoS database for 13,662 publications published during a decade (2013–2023). There were 13,038 articles. Growth was expressed annually at 18.51% as a negative percentage rate. Based on Fig. 3, in the last 10 years, the number of publications slightly decreased from 2013 to 2015 and has increased dramatically since 2015. Moreover, the average number of citations received per paper (times cited) was benchmarked at 13.47. Each paper was written by four authors on average (4.92).

Table 2. The most relevant sources.

Sources	Articles
IEEE Access	784
Expert System Representations	398
Knowledge-Based System	385
Bioinformatics	383
Semantic Web	349
Journal of Biomedical Informatics	257
Journal of Web Semantic	223
Information Sciences	183
Multimedia Tools and Applications	181
Future Generation Computer Systems-	160
The International Journal of Escience	

Table 2 presents the top 10 journals as the most relevant sources about KGs. The *IEEE ACCESS* journal published 784 articles and ranked first.

Between 2013 and 2023, an increase was observed in publications, particularly by IEEE Access, Expert System with Applications, Knowledge-Based System, Bioinformatics, and Semantic Web journals. Therefore, the journals have published a depth of knowledge and expertise on KGs and related subtopics. Fig. 4 depicts the sources' publication over time.

Table 3 and Fig. 5 show the top countries' publications. China ranked first (8313), followed by the USA and UK, respectively. Iran, a developing country, ranked 16th.

Many citations were reported from China, the USA, and Spain, ranking first to third, respectively. All countries retained their leading roles as the top 10 countries in the citations. Table 4 presents the citations per country.

Based on Table 5, China played a significant role in international research networks. China made a network collaboration with the USA as the first rank of scholarly communication related to KGs globally. Fig. 6 depicts the path of global professional collaboration.

Table 3. Most prolific countries.

Rank	Country	Freq.
1	China	8313
2	USA	4964
3	UK	1723
4	Spain	1629
5	Italy	1483
6	France	1376
7	Germany	1336
8	India	1154
9	Australia	1029
10	South Korea	838

Table 4. Countries, TC, and AAC.

Rank	Country	Total Citations (TC)	Average Article Citations (AAC)
1	China	40767	11.72
2	USA	28229	17.84
3	Spain	11104	14.35
4	UK	10426	16.90
5	Italy	9105	12.77
6	Germany	8856	18.11
7	Australia	8763	23.62
8	France	7636	15.09
9	India	5157	9.21
10	Korea	5049	12.53

Table 5. World collaboration frequency.

From	To	Frequency
China	USA	451
China	Australia	178
China	UK	159
USA	UK	139
UK	Italy	103
China	Canada	102
UK	Germany	102
China	Singapore	88
USA	Canada	85
USA	Germany	85

Fig. 7 displays the publications of affiliations over time. All the top affiliations had an increasing trend during the last decade. Tsinghua Univ, located in China, published 1154 articles and ranked first. This was followed by other Chinese affiliations, namely PEKING UNIV (974), TSINGHUA UNIV (1428), Univ Chinese Acad Sci (1180), and Zhejiang Univ (994). It is concluded that China produced extensive and profound knowledge related to KGs.

The top and most relevant authors and the number of their articles are reported in Table 6. Fractional authorship quantifies an individual author's contributions to a published set of papers (Aria and Cuccurullo, 2017). Zhang Y,

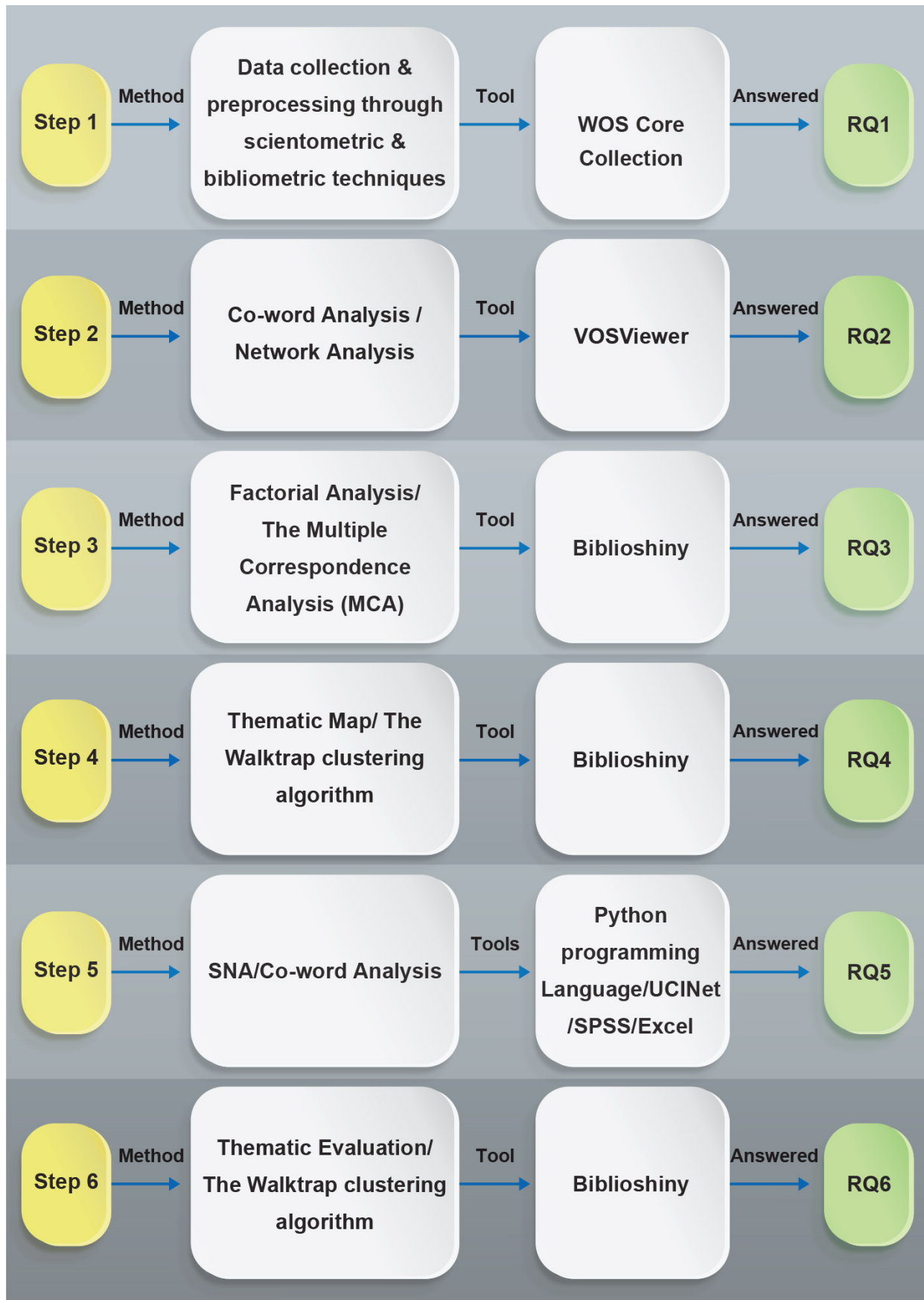


Fig. 2. Schematic methodology.

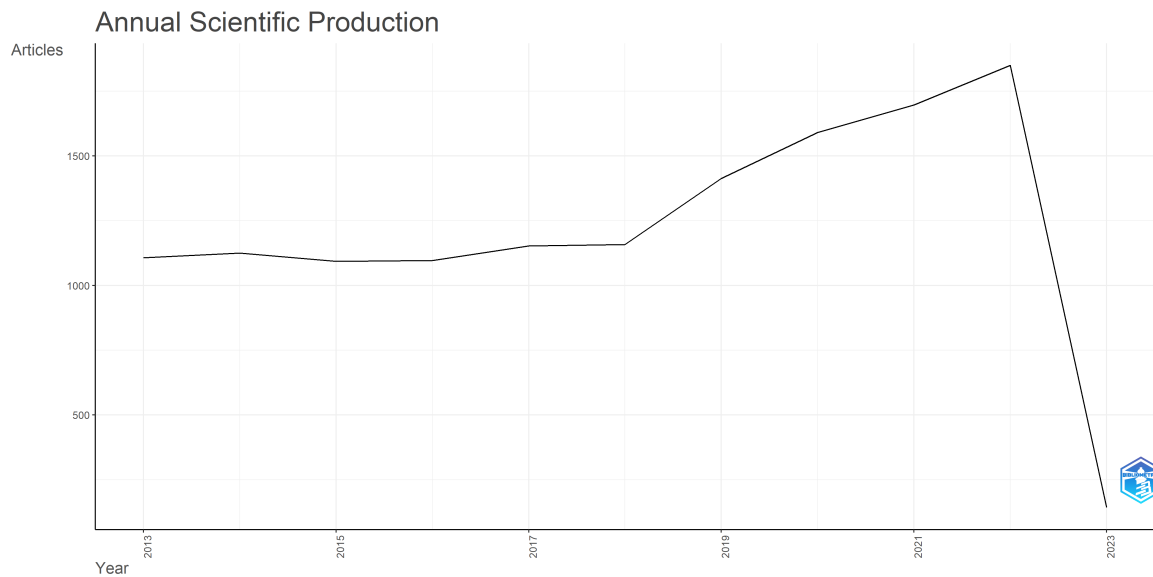


Fig. 3. Annual scientific production.

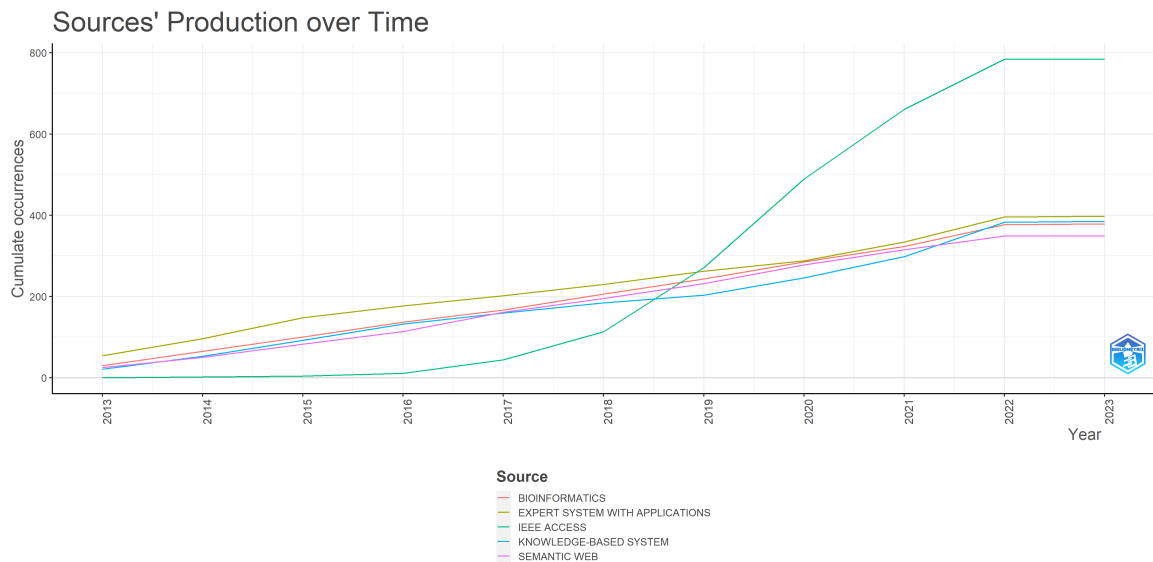


Fig. 4. Source growth.

Table 6. Prolific authors.

Rank	Authors	Articles	Articles Fractionalized
1	ZHANG Y	69	14.94
2	LIU Y	61	15.91
3	WANG Y	60	14.44
4	WANG J	55	12.80
5	LIU J	54	12.42
6	WANG H	52	12.49
7	CHEN L	49	10.19
8	LI J	48	10.71
9	ZHANG J	48	11.42
10	LEE S	46	10.27

with 69 articles and a contribution of almost 14 (Articles Fractionalized = 14.94), was the top writer in the field.

Table 7 lists the characteristics of the most cited documents related to KGs. The articles were cited exclusively in the last 10 years from various fields, such as bioinformatics and computer sciences. A total of 2004 articles among 13,662 had no citation (15%). This does not necessarily mean the articles have not been cited elsewhere. The data pertains solely to citations within the WoS, not other citation sources.

The software's threshold value, determined by trial and error, was ≥ 10 co-occurrences, which formed eight main clusters of the 29,411 keywords. Of these, 707 meet the threshold, including 707 keywords. Fig. 8 depicts

Country Scientific Production

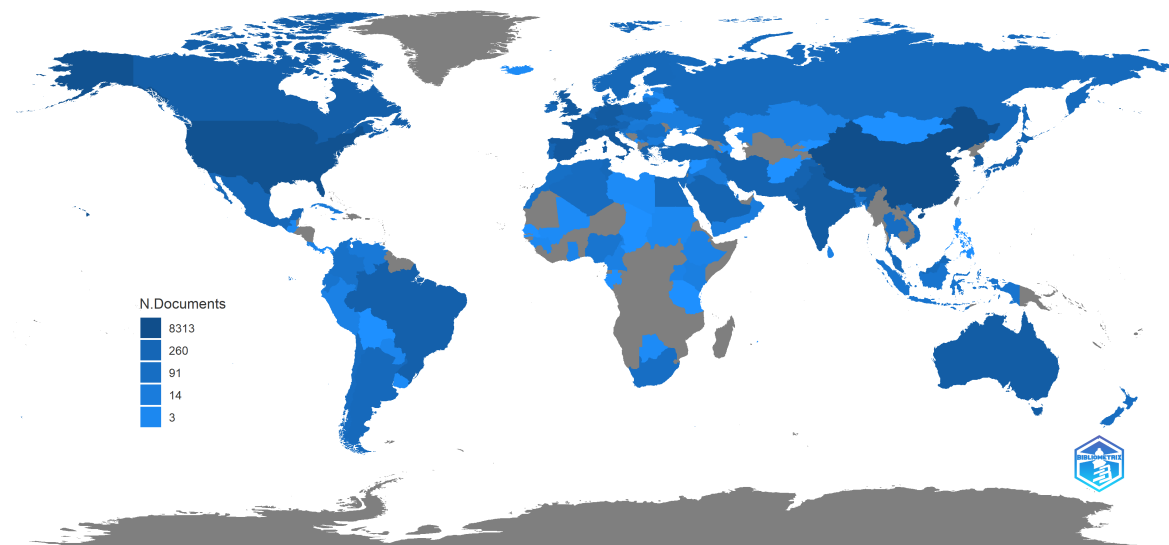


Fig. 5. Scientific publications of countries.

Country Collaboration Map

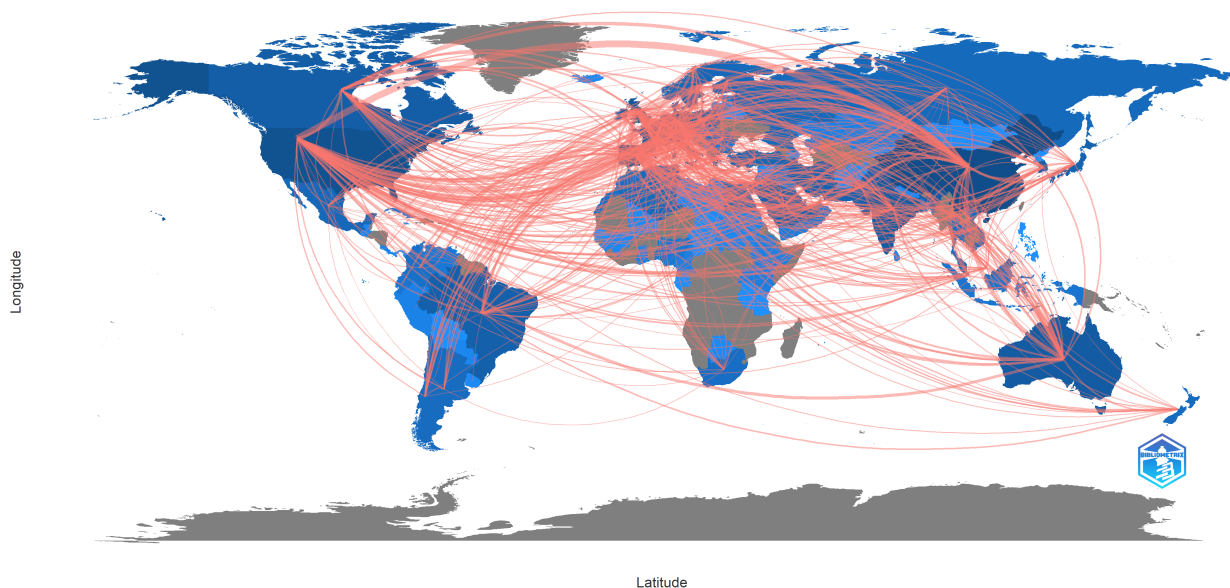


Fig. 6. Collaboration map of countries.

the network structure in the field of KGs, including the keywords visualized by VOSviewer 1.6.19. The network consisted of nine main clusters in different colors, using VOSviewer's algorithms and analyses on 28,298 total co-occurrences, 800,200 total link strength, and 5818 total links.

Fig. 9 illustrates the overlay visualization of the network in this field. The colors of this map were determined by their weight in the network. Blue shows the lowest score, green indicates an average score, and yellow represents the highest score.

Thus, moving from blue to yellow indicates more importance and weight due to the keyword's greater score and significance in the network (Van Eck and Waltman, 2014). The years 2017, 2018, and 2019, followed by 2020 and 2021, included more prominent and major keywords in the overlay network related to KGs. Tables 8,9,10,11 represents the characteristics of clusters based on keywords' co-occurrence.

The high-frequency keywords were subjected to MCA to analyze factorial analysis in the field of KGs. Large-scale

Table 7. The most cited documents globally.

Rank	Journal	Title of Paper	Author	Total Citations	TC per Year
1	Bioinformatics	Causal analysis approaches in Ingenuity Pathway Analysis	KRAMER A. <i>et al.</i> , (2014)	2799	279.90
2	IEEE Communications Surveys & Tutorials	Context-Aware Computing for The Internet of Things: A Survey	PERERA C. <i>et al.</i> , (2014)	1457	145.70
3	Semantic Web	DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia	LEHMANN J. <i>et al.</i> , (2015)	1231	136.78
4	MIS Quarterly	Positioning and presenting design science research for maximum impact	GREGOR S. <i>et al.</i> , (2013)	1062	96.55
5	IEEE Transactions on Knowledge and Data Engineering	Knowledge Graph Embedding: A Survey of Approaches and Applications	WANG Q. <i>et al.</i> , (2017)	857	122.43
6	Bioinformatics	CluePedia Cytoscape plugin: pathway insights using integrated experimental and <i>in silico</i> data	BINDEA G. <i>et al.</i> , (2013)	704	64.00
7	Bioinformatics	ShinyGO: a graphical gene-set enrichment tool for animals and plants	GE SX. <i>et al.</i> , (2020)	633	158.25
8	Knowledge-Based Systems	A survey on opinion mining and sentiment analysis: Tasks, approaches and applications	RAVI K. <i>et al.</i> , (2015)	628	69.78
9	Bioinformatics	UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches	SUZEK BE. <i>et al.</i> , (2015)	623	69.22
10	IEEE Transactions on Parallel and Distributed Systems	Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement	FU ZJ. <i>et al.</i> , (2016)	600	75.00

Table 8. Characteristics of clusters based on keywords' co-occurrence analyzed by VOSViewer.

The first cluster: 142 keywords / total co-occurrences: 5563 / links: 5685 / total link strength: 12364	
	Keywords in the Cluster
Ontology-based interoperability of knowledge	<p>Ontology/ interoperability/knowledge management/social network ontology engineering/conceptual model/semantic interoperability/social media/algorithm/domain ontology/evaluation/collaboration/knowledge/information system/software engineering/data quality/context/healthcare/case-based reasoning/unified modeling language/data visualization/modeling/visual analytics/design/social network analysis/twitter/graph database/web of data/simulation/usability/information/ontology evaluation/human computer interaction/knowledge organization/knowledge sharing/performance/network analysis/data science/semantic network/process mining/epistemology/formal ontology/information management/ontology design/case study/software/architecture/information/integration/language/methodology/multimedia/framework/foundational ontology/reuse/sustainability/energy efficiency/pattern mining/philosophy/telemedicine/decision-making/information technology/model-driven engineering/ontological analysis/theory/bim/business process modeling/ehealth/enterprise architecture/experimentation/information visualization/mhealth/ontology reuse/representation/virtual reality/content analysis/data transformation/gis/interaction/knowledge reuse/metrics/ontology construction/owl ontology/quality/semantic representation/business process/computer-aided design/construction/engineering informatics/information science/ information theory/model/process modeling/quality assessment/search/analysis/assessment/concepts/design science/integration/knowledge retrieval/maintenance/navigation/requirements/reverse engineering/seminternet of things/user experience/conceptual design/data/design science research/documentation/domain-specific language/engineering design/function/kdd/knowledge model/onto unified/modeling language/product development/product lifecycle management/socio materiality/systematic review/tacit knowledge/ufo/buildings/community detection/conversational agent/critical realism/enterprise ontology/experiment/first-order logic/health informatics/health information system/hierarchical clustering/human factors/knowledge-based engineering/metamodel/metamodeling/ontology design patterns/public health/semantic data/standardization/supply chain/user interface/web</p>
The second cluster: 124 keywords / total co-occurrences: 4322 / links: 5138 / total link strength: 9704	
	Keywords in the Cluster
Knowledge representation and reasoning	<p>semantic web/ knowledge representation/owl/description logic/reasoning/decision support system/expert system/web services/fuzzy logic/multi-agent system/semantic web service/fuzzy ontology/decision making/knowledge-based system/bayesian network/context awareness/provenance/context modeling/context awareness/decision support/service discovery/quality of service/activity recognition/query answering/semantic model/clinical decision support/cyber-physical system/service/composition/ambient intelligence/similarity/swrl/knowledge modeling/semantic reasoning/big data analysis/building information modeling/ontology evolution/web service/agent/complexity/argumentation/computational complexity/intelligent system/ontology-based data/access/rules/data management/domain knowledge/ontology modeling/user profile/diagnosis/semantic web technologies/workflow/trust/knowledge representation and reasoning/logic programming/ontology reasoning/planning/service-oriented architecture/belief revision/inconsistency/business intelligence/inference/situation awareness/ubiquitous computing/web service composition/fuzzy description logic/learning/smart home/verification/answer set programming/middleware/query rewriting/semantic matching/ambient assisted living/discovery/mobile computing/model checking/visual knowledge discovery/automated reasoning/cloud/dl-lite/information fusion/reproducibility/smart environments/augmented reality/cloud manufacturing/datalog/matching/probabilistic reasoning/risk assessment/robotics/rule-based reasoning/e-commerce/alignment/causality/context awareness computing/intelligent agent/non monotonic reasoning/ontology quality assurance/scientific workflow/semantic enrichment/composition/constraints/ disaster management/explanation/graph theory/information modeling/obda/ontological reasoning/ontology debugging/patterns/pervasive computing/petri nets/preferences/semantic knowledge/stream reasoning/validation/accessibility/actionable knowledge discovery/educational technology/group decision making/medical ontology/network security/security ontology/virtual enterprise</p>

Table 9. Cont. Characteristics of clusters based on keywords' co-occurrence analyzed by VOSViewer.

The third cluster: 87 keywords / total co-occurrences: 2732 / links: 5273 / total link strength: 9412	
Application of knowledge graph in big data	<p>Keywords in the cluster</p> <p>internet of things/ big data/artificial intelligence/data model/cloud computing/licenses/standards/open data/semantic technology/privacy/measurement/taxonomy/industry 4/security/tools/sensors/data analysis/software/map reduce/smart city/blockchain/monitoring/analytical models/education/vocabulary/xml/survey/data analytics/smart cities/manufacturing/bibliometrics/requirements engineering/reliability/digital twin/edge computing/Hadoop/medical services/wireless sensor network/linguistics/performance evaluation/syntactics/organizations/cybersecurity/object recognition/automation/data sharing/innovation/resource management/robots/companies/distributed database/e-government/technological innovation/trajectory/data privacy/parallel computing/protocols/safety/semantic integration/smart manufacturing/convergence/games/government/graph mining/query optimization/risk management/smart grid/data warehouse/distributed computing/open government data/semantic rules/geographic information system/history/information security/ /smart contracts/systematics/cybernetics/digital forensics/fog computing/human-robot interaction/industries/market research/peer-to-peer computing/psychology/scientometrics/stakeholders/systematic literature review</p>
The fourth cluster:86 keywords / total co-occurrences: 4644 / links: 6673 / total link strength: 18824	
Knowledge engineering for task analysis	<p>Keywords in the Cluster</p> <p>knowledge graph/knowledge engineering/task analysis/deep learning/feature extraction/training/neural network/visualization/knowledge graph embedding/computational/modeling/ink prediction/predictive models/convolutional neural network/representation learning/cognition/attention mechanism/graph neural network/knowledge graph completion/reinforcement learning/transfer learning/adaptation models/graph convolutional network/correlation/learning systems/knowledge distillation/deep neural network/mathematical model/computer architecture/explainable artificial intelligence/computer vision/biological system modeling/entity alignment/multitask learning/robustness/encoding/image classification/attention/training data/benchmark testing/probabilistic logic/computer science/knowledge transfer/generators/graph attention network/prediction algorithm/interpretability/logic gates/topology/convolution/knowledge base completion/network topology/data models/deep reinforcement learning/embedding/knowledge reasoning/routing/visual question answering/decoding/electronic mail/generative adversarial network/temporal knowledge graph/zero-shot learning/few-shot learning/knowledge graph reasoning/knowledge representation learning/prototypes/solid modeling/meta-learning/three-dimensional displays/community question answering/continual learning/domain adaptation/estimation/neurons/noise measurement/parallel processing/shape/testing/adaptive systems/adversarial learning/document classification/face recognition/model compression/multi-hop reasoning/perturbation methods/relation prediction</p>

Table 10. Cont. Characteristics of clusters based on keywords' co-occurrence analyzed by VOSViewer.

The fifth cluster: 81 keywords / total co-occurrences: 3705 / links: 4131 / total link strength: 8451	
Knowledge graph querying	<p>Keywords in the cluster</p> <p>linked data/rdf/knowledge base/sparql/information retrieval/semantic similarity/data integration/metadata/question answering/electronic health record/semantic annotation/ontology matching/semantic search/cultural heritage/entity linking/wordnet/dbpedia/ontology alignment/Wikipedia/biomedical ontology/snomed ct/annotation/digital library/digital humanities/similarity measure/ranking/ontology mapping/data fusion/benchmark/query expansion/image retrieval/relational database/medical informatics/query processing/mapping/terminology/indexing/instance matching/semantic relatedness/information content/skos/quality assurance/thesauri/entity resolution/keyword search/query/graph/rdf graph/thesaurus/diversity /ontology integration/ svm/unified medical language system /evolution/natural language generation/semantic relations/text analysis/entity disambiguation/geosparql/knowledge fusion/music/wikidata/background knowledge/heterogeneous data/mesh/natural language understanding/phenotype/query language/record linkage/semantic ontology/controlled vocabularies/design patterns/knowledge organization systems/lexicon/linguistic linked data/pharmacovigilance/rare diseases/semantic mapping/sparql query/storage/web search.</p>
The sixth cluster: 68 keywords / total co-occurrences: 2611/ links: 3913/ total link strength: 7677	
Knowledge graph analysis based on ML methods	<p>Keywords in the cluster</p> <p>data mining/machine learning/clustering/gene ontology/database/optimization/bioinformatics/covid-19/genetic algorithm/ clustering algorithm/proteins/unsupervised learning/diseases/clinical decision support/system/gene expression/anomaly detection/informatics/graph embedding/search engine/artificial neural network/protein-protein interaction/breast cancer/genomics/protein function prediction/indexes/molecular dynamics simulation/evolutionary algorithm/medical diagnostic imaging/electronic medical record/statistics/drugs/real-time systems/random forest/entropy/knowledge embedding/cancer/cluster analysis/molecular docking/multi-objective optimization/network pharmacology/protein-protein interaction network /classification algorithm/biclustering/differentially expressed genes/medical diagnosis/search problems/sociology/fuzzy clustering/intrusion detection/principal component analysis/particle swarm optimization/biomedical informatics/drug repurposing/alzheimer's disease/concept extraction/machine learning algorithm/outlier detection/precision medicine/ant colony optimization/clinical trials/meta-analysis/missing data/pathway/personalized medicine/radial distribution function/sars-cov-2/statistical analysis/traditional Chinese medicine</p>
The seventh cluster: 60 keywords / total co-occurrences: 2118 / links: 2988/ total link strength: 5514	
Knowledge graph extraction	<p>Keywords in the Cluster</p> <p>natural language processing/text mining/information extraction/knowledge extraction/sentiment analysis/knowledge acquisition/relation extraction/named entity recognition/ontology learning/crowdsourcing/semantic analysis/fault diagnosis/text classification/word embedding/scalability/access control/bert/topic modeling/ontology development/word sense disambiguation/pattern recognition/transformer/neural/opinion mining/bit error rate/distant supervision/semi-supervised learning/entity recognition/supervised learning/web mining/ontology population/active learning/event extraction/lstm/time series/dimensionality reduction/folksonomy</p> <p>hidden markov models/latent dirichlet allocation/named entity disambiguation/topic model/labeling/maintenance engineering/multi-label classification/time series analysis/collective intelligence/ontology enrichment/tagging/emotion recognition/genetic programming/motion pictures/relationship extraction/affective computing/citation analysis/crf/dataset/emotion</p> <p>lda/online social network/software agent</p>

Table 11. Cont. Characteristics of clusters based on keywords' co-occurrence analyzed by VOSViewer.

The eighth cluster: 42 keywords/ total co-occurrences: 1816/ links: 2184/ total link strength: 4478	
	Keywords in the cluster
Knowledge discovery and data mining	knowledge discovery/ classification/recommendation system /feature selection/formal concept analysis/uncertainty/collaborative filtering/association rules/recommendation/decision tree/e-learning/granular computing/personalization/ rough set/rough set theory/prediction/attribute reduction/incremental learning/association rule mining/fuzzy sets/ensemble learning/matrix factorization/computational intelligence/data engineering/event detection/heuristic algorithm/complex network/concept lattice/knowledge integration/rule extraction/three-way decision/frequent pattern mining/diabetes/educational data mining/knowledge structure/singular value decomposition/concept drift/data reduction/industrial internet of things/online learning/ontology model/ temporal data.
The ninth cluster: 17 keywords / total co-occurrences: 787/ links: 1429/ total link strength: 3776	
Label	Keywords in the cluster
Application of knowledge-based systems	Semantic/knowledge-based systems/Internet/natural language/recurrent neural network/object detection/complexity theory/libraries/Bayes methods/encyclopedias/remote sensing/proposals/image annotation/text categorization/ electronic publishing/heterogeneous information network/image segmentation.

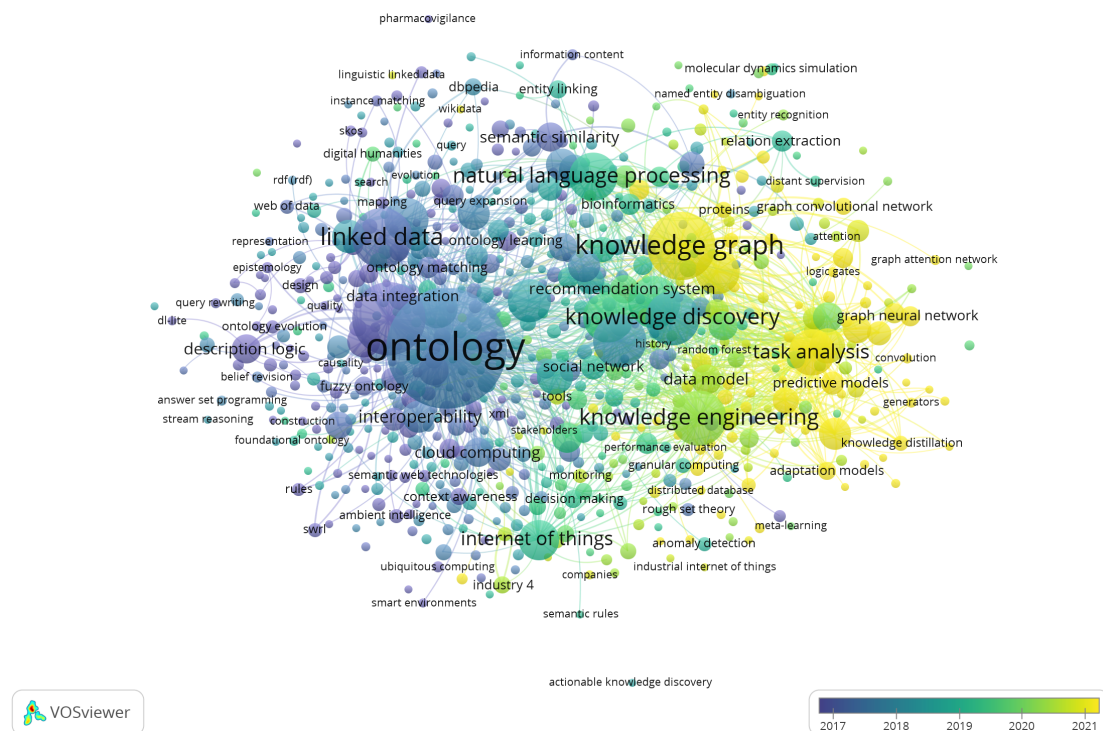


Fig. 9. Overlay visualization in the field of KG by authors' keywords (VOSViewer).

words, keywords approaching the center point indicate the greatest attention in recent years (Xie et al, 2020).

Fig. 10 was adjusted with set parameters, such as the number of terms 50, the number of clusters auto, label size 10, and the number of documents 5. The conceptual structure map included 2 clusters. **Cluster 1** (red), the most significant cluster including 31 keywords, comprised papers on SPARQL, RDF, OWL, linked data, semantic web, reasoning, ontology engineering, knowledge management, conceptual model, semantic similarity, clustering, information extraction, text mining, natural language processing, link prediction, deep learning, social network, knowledge representation, recommendation system, cloud computing, the Internet of Things, data integration, interoperability, metadata, description logic, ontology, big data, knowledge base, information retrieval, KG, and classification. **Cluster 2** (blue) consisted of studies on task analysis, semantic, visualization, data model, feature extraction, computational modeling, predictive models, neural network, and training.

Fig. 11 shows the thematic map of authors' keywords, including 250 words, 3 labels, and 5 min cluster frequency (per 1000 docs), based on the Walktrap clustering algorithm. The thematic map includes four quadrants containing different degrees of density and centrality. In other words, density and centrality are indicators of the SNA approach. The y-axis stands for density, measuring the internal strength of a cluster. The x-axis indicates the importance of the topic and represents centrality. High centrality (high degree of relevance) means the cluster has a more important position in the field. As explained in Fig. 1 and

shown in Fig. 11, Quadrant II implies low centrality and high density, representing niche themes. Clusters ‘knowledge graph’, ‘knowledge engineering’, and ‘knowledge discovery’ were located in this quadrant. They were not central themes; they are isolated but well-developed in the field of KG.

Themes located in Quadrant IV indicate basic themes due to high centrality and low density (Khasseh et al, 2017). Therefore, ‘ontology’, ‘semantic web’, and ‘linked data’ represented central but not developed clusters.

The strategic diagram (SD) was graphed using SNA metrics such as centrality and density. This approach helps examine cluster advancement and growth. To this end, a co-occurrence matrix (created using Python code) was created, followed by a correlation matrix (created using Microsoft Excel); UCINET was also used to measure the centrality and density of each cluster (Table 12).

The SD was visualized (Fig. 12) based on the origin of the centrality average (0.41) and density average (0.34), respectively. Quadrants for the clusters located in the SD are mentioned in Table 13 based on the quadrants' type.

Table 13 summarizes the identified and classified clusters in the SDs. Cluster 9, ‘application of knowledge-based systems’, was recognized as a motor theme in Quadrant I. It was placed at the core of the field. In other words, it stood in a dynamic and dominant position and had strong internal linkage and maturation. Clusters one (*ontology-based interoperability of knowledge*) and 4 (*knowledge engineering for task analysis*) in Quadrant II implied high density and

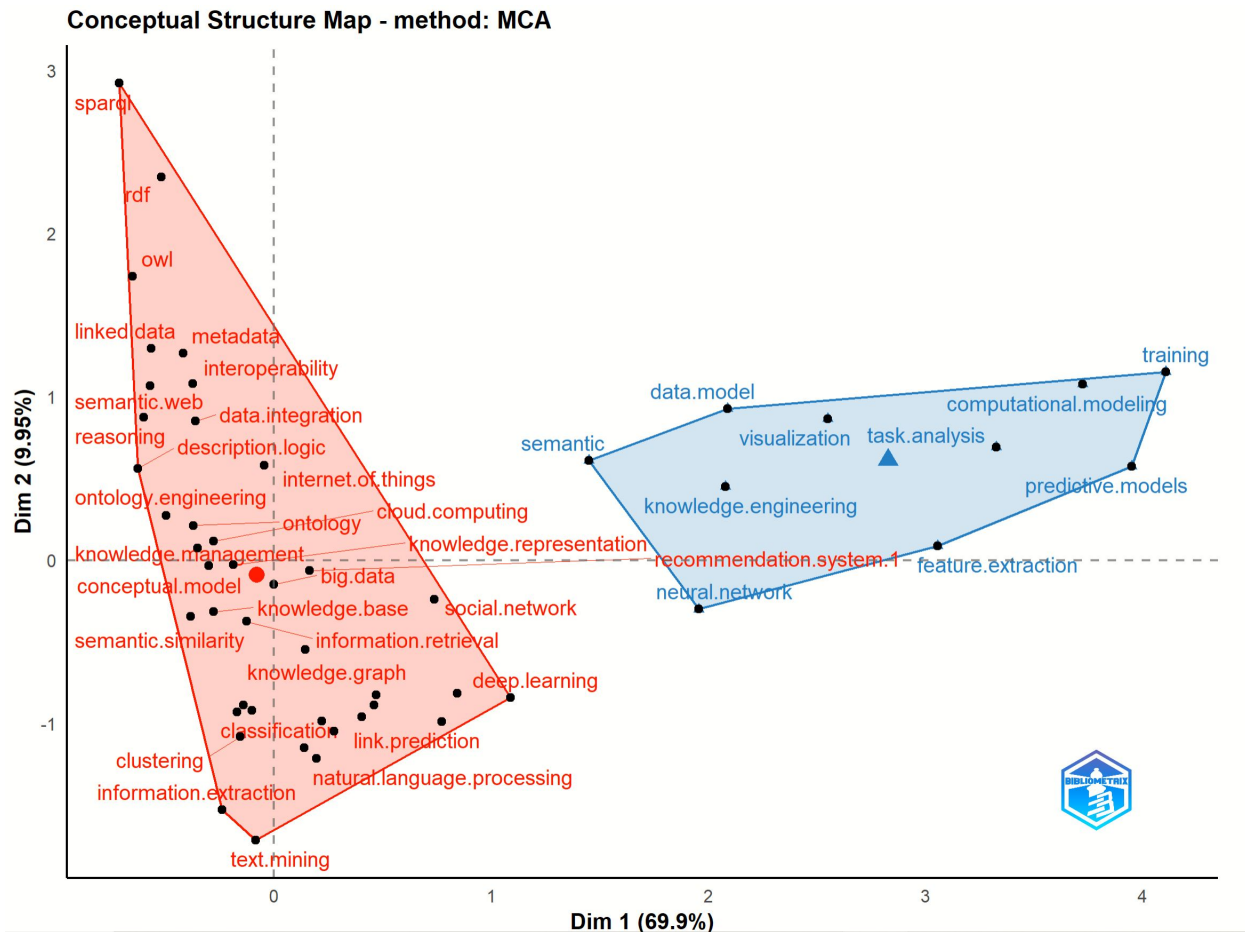


Fig. 10. Factorial analysis of the conceptual structure map method: MCA of high-frequency keywords. MCA, Multiple correspondence analysis

Table 12. Degree of Centrality and Density of the Clusters from UCInet.

Cluster	Density	Centrality
1	0.66	0.24
2	0.22	0.51
3	0.22	0.36
4	0.77	0.18
5	0.25	0.47
6	0.23	0.48
7	0.20	0.49
8	0.19	0.41
9	0.33	0.61

low centrality, indicating niche themes, which are isolated yet well-developed themes in KGs.

Themes in Quadrant III represented emerging, not axial, or declining themes (Khaseh et al, 2017). The cluster entitled ‘application of knowledge graph in big data’ was located here as a marginal topic, not axial, with low attention, low centrality, and low density. In other words, it was an underdeveloped theme with largely discontinuous structures.

Five basic themes, namely ‘knowledge representation and reasoning’, ‘knowledge graph querying’, ‘knowledge graph analysis based on machine learning (ML) methods’, ‘knowledge graph extraction’, and ‘knowledge graph classification’ were located in Quadrant IV, indicating central but not developed clusters due to high centrality and low density.

Fig. 13 analyzes the evolution of the themes considering their keywords across time. It is depicted with authors’ keywords, cutting point 2 (cutting year 1 2017, cutting year 2 2020), and the Walktrap clustering algorithm. This figure includes the 2013–2017, 2018–2020, and 2021–2023 periods as a Sankey diagram.

Some trends like ‘data mining’, ‘natural language processing’, ‘formal concept analysis’, and ‘ontology’ were revealed as thematic growth in 2013–2018. An evolution was found to ‘knowledge graph’, ‘ontology’, and ‘gene ontology’ in 2018–2020. Moreover, six main thematic evolutions during 2021–2023 were identified: ‘ontology’, ‘natural language processing’, ‘machine learning’, ‘protein’, ‘knowledge engineering’, and ‘knowledge graph’. Accordingly, the research in this field presents dramatic cohesion; the detected themes are categorized within a discovered

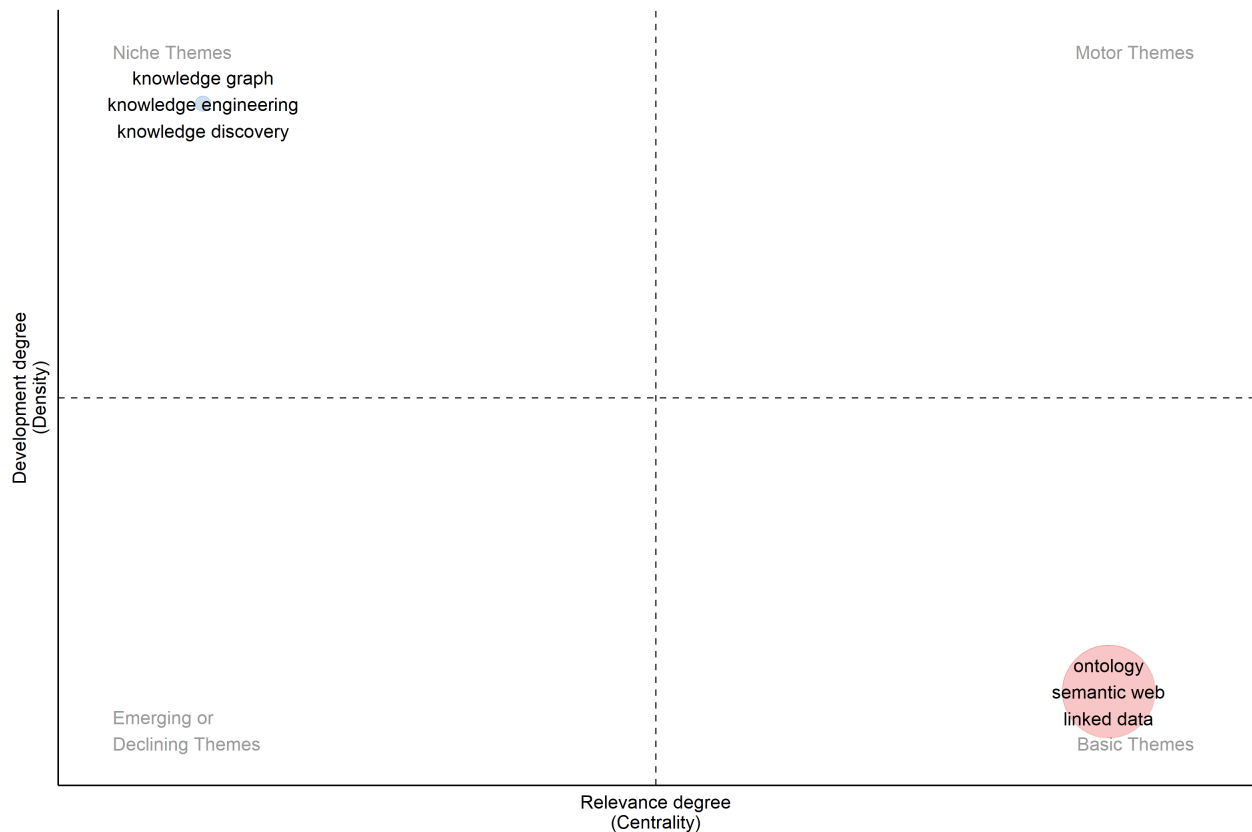


Fig. 11. Thematic map of authors' keywords.

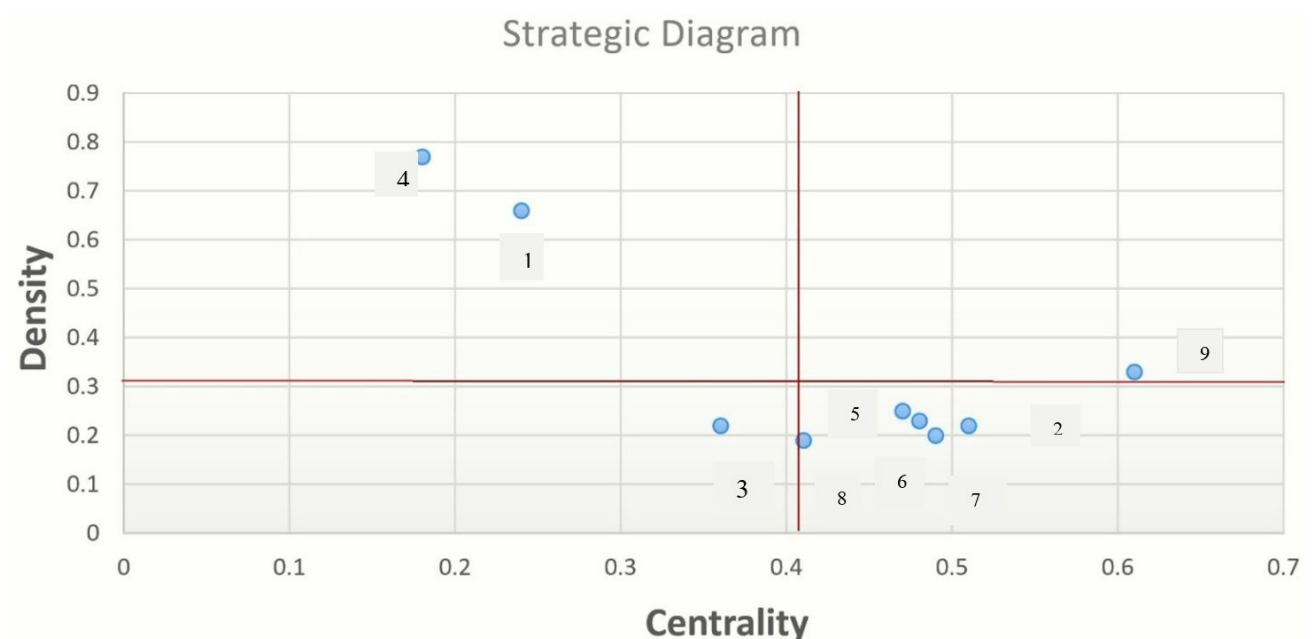


Fig. 12. Strategic Diagram of the Clusters by VOSViewer.

cluster taken from the thematic map (Fig. 11) like 'knowledge engineering', 'knowledge discovery', and 'knowledge graph' as niche themes, and 'ontology' as a basic and traversal theme. Besides, Table 14 shows the flow of thematic evolution within words and occurrences.

Table 15 indicates the topical evolution and conceptual structure of various time-based and sequential sub-periods of diverse themes. It reveals the thematic progression of KGs, exposing the paths and trajectories of various topics throughout time. The first path covers the 2013–2017

Table 13. Quadrants for the clusters located in the strategic diagram.

Quadrant II: Niche Themes	Quadrant I: Motor Themes
C1: Ontology-based interoperability of knowledge	C9: Application of knowledge-based systems
C4: Knowledge engineering for task analysis	
Quadrant III: Chaos and emerging	Quadrant IV: Basic and transversal
C3: Application of knowledge graph in big data	C2: Knowledge representation and reasoning
	C5: Knowledge graph querying
	C6: Knowledge graph analysis based on ML methods
	C7: Knowledge graph extraction
	C8: Knowledge discovery and data mining

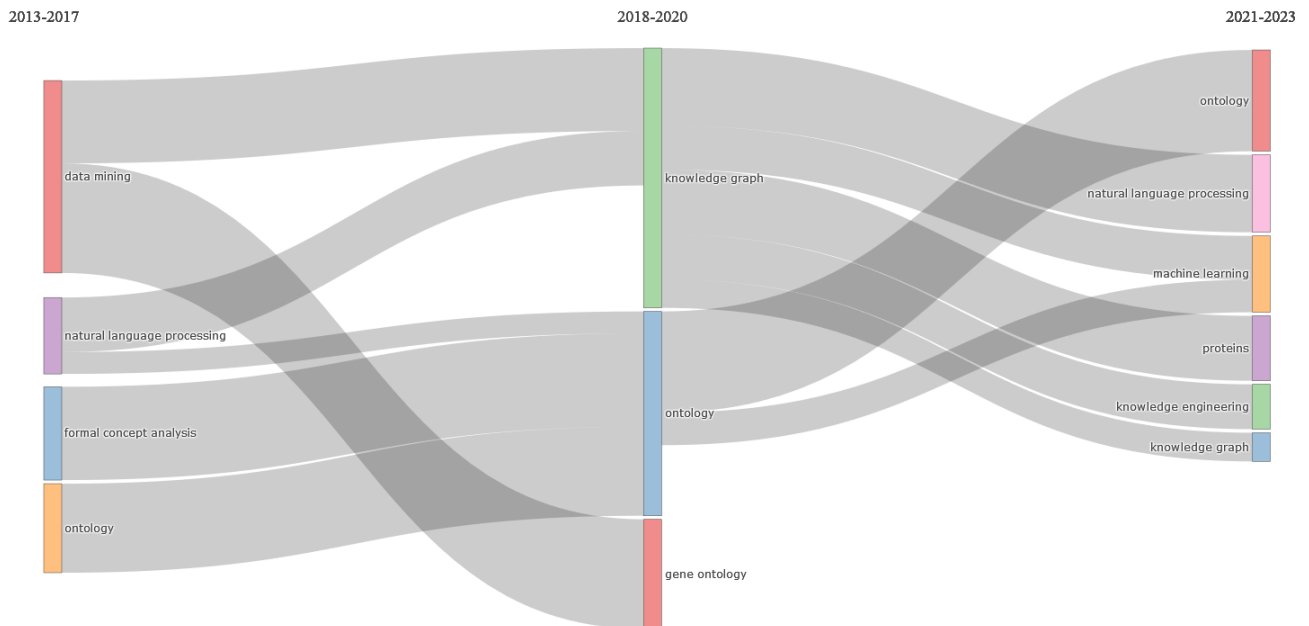


Fig. 13. Sankey diagram of thematic evolution of the KG research field (2013–2023).

period and includes some thematic clusters: ‘formal concept analysis, algorithm design, fuzzy ontology’ and ‘semantic web, linked data, ontology, data mining, knowledge discovery, and machine learning’. These clusters are classified as niche themes and basic themes, respectively. The second trajectory (2018–2020) features two thematic clusters: ‘gene ontology’ and ‘knowledge graph, knowledge discovery, and machine learning’. Emerging themes and motor themes are the respective categories assigned to these clusters. The third progression pertains to the 2021–2023 time frame, comprising several thematic clusters, namely ‘protein, bioinformatics, feature selection’, ‘knowledge engineering, task analysis, semantics’, categorized as the motor theme. Moreover, ‘semantic web, linked data, ontology’ is classified as the basic theme, as represented in Fig. 11. ‘Natural language processing, knowledge base’ is grouped as an emerging theme during time slice 3.

5. Discussion

This paper highlighted outcomes and showcased findings from a topic-driven bibliometric analysis of academic

papers on KGs from 2013 to 2023. The growth of published papers on KGs over the past decade was reported as a negative percentage rate of 18.51% annually. This means that the growth rate was not constant but temporary and low. This trend is depicted in Fig. 3. Considering the negative growth, it is necessary to examine this trend in the long term to know whether the increase in publications was temporary or had stable growth. There was also a slight decrease in publication numbers from 2013 to 2015, followed by a considerable increase since 2015. Moreover, times cited (citations received per paper) equaled 13.47 on average, and each article was written by 4 authors (4.92) on average. The *IEEE Access* journal had the top position and was the most relevant resource. The study by [Chen et al \(2021b\)](#) also revealed that *IEEE Access* is the most prolific journal in the field.

Journals *IEEE Access*, *Expert System with Applications*, *Knowledge-Based System*, *Bioinformatics*, and *Semantic Web* published a wealth of knowledge in various subjects of KGs and its related domains. China, the USA, and the UK ranked as leading countries, respectively. As

Table 14. The flow of thematic evolution within words and occurrences.

From	To	Words	Occurrences
Data mining—2013–2017	Gene ontology—2018–2020	gene ontology	48
Data mining—2013–2017	Knowledge graph—2018–2020	data mining; knowledge discovery; machine learning; semantic similarity; classification; clustering	183
Formal concept analysis—2013–2017	Ontology—2018–2020	formal concept analysis	27
Natural language processing—2013–2017	Knowledge graph—2018–2020	natural language processing; knowledge base; text mining; knowledge engineering; information extraction; knowledge extraction; social media; recommender systems; crowdsourcing; sentiment analysis	86
Natural language processing—2013–2017	Ontology—2018–2020	information retrieval; big data; knowledge acquisition; artificial intelligence; fuzzy logic	74
Ontology—2013–2017	Knowledge graph—2018–2020	Semantics	71
Ontology—2013–2017	Ontology—2018–2020	ontology; semantic web; linked data; RDF; knowledge representation; interoperability; OWL; knowledge management; cloud computing; SPARQL; reasoning; linked open data; data integration; ontology engineering; internet of things; metadata	1034
Knowledge graph—2018–2020	Knowledge engineering—2021–2023	knowledge discovery; knowledge engineering; deep learning; semantics; task analysis; feature extraction; classification; data models; visualization; training; neural networks; computational modeling; knowledge-based systems; licenses; optimization; predictive models.	152
Knowledge graph—2018–2020	Knowledge graph—2021–2023	knowledge graph; recommender systems; knowledge graph embedding; reinforcement learning; recommender system	161
Knowledge graph—2018–2020	Machine learning—2021–2023	machine learning; data mining; clustering	160
Knowledge graph—2018–2020	Natural language processing—2021–2023	natural language processing; knowledge base; text mining; information extraction; sentiment analysis; knowledge extraction; question answering; relation extraction; social media	110
Knowledge graph—2018–2020	Ontology—2021–2023	semantic similarity	42
Knowledge graph—2018–2020	Proteins—2021–2023	bioinformatics; feature selection	26
Ontology—2018–2020	Knowledge graph—2021–2023	KGs	22
Ontology—2018–2020	Machine learning—2021–2023	big data; internet of things; cloud computing; artificial intelligence; security; Internet of things (IoT)	84
Ontology—2018–2020	Natural language processing—2021–2023	information retrieval	46
Ontology—2018–2020	Ontology—2021–2023	ontology; semantic web; linked data; knowledge representation; SPARQL; linked open data; knowledge management; data integration; owl; interoperability; reasoning; metadata; ontology engineering; resource description framework.	667

Table 15. Topical evolution of various themes across sub-period time slices.

Period	Thematic Cluster	Kinds of Themes
Time slice 1	Formal concept analysis, algorithm design, fuzzy ontology	Niche theme
2013–2017	Semantic web, linked data, ontology, data mining, knowledge discovery, machine learning	Basic theme
Time slice 2	Gene ontology	Emerging theme
2018–2020	Knowledge graph, knowledge discovery, machine learning	Motor theme
Time Slice 3	Protein, bioinformatics, feature selection	Niche theme
2021–2023	Knowledge engineering, task analysis, semantics	Motor theme
	Semantic web, linked data, ontology	Basic theme
	Natural language processing, knowledge base	Emerging theme

Chen et al (2021b) accentuated, China plays an active role in KG research. Iran, despite being a developing country, ranks 16th.

China, the USA, and Spain held the top three ranks in terms of the number of citations by country, respectively. China is a prominent player in global research networks on KGs. Moreover, China and the USA formed a top-ranked scholarly communication network. Tsinghua Univ was recognized as the top-ranked affiliation in China, followed closely by some Chinese affiliations. It implies that China has made significant and valuable contributions to the field of KGs. The most cited articles belonged to fields such as bioinformatics and computer sciences. It confirms that these fields were under consideration and had a high capacity to increase productivity and make valuable contributions to the progression of understanding and the evolution of knowledge on KGs. The results of Niknia and Mir-taheri (2015) and Hosseini et al (2021) also revealed that the top scientific publications in the WoS classification on linked data belong to the sub-categories of computer sciences. Chen et al (2021b) demonstrated that the KG domain will be deeply integrated with the computer science domain. As a significant domain, bioinformatics had a significant portion of scientific publications and was a significant leader in the field, as accentuated by Hosseini et al (2021). Moreover, Chen et al (2021b) strongly advised that scholars should emphasize this area.

Moreover, 15% of all the articles had zero citations. This finding suggests that despite being available for some time, a notable proportion of the published articles in the field have not received citations from other researchers. This could be due to factors such as limited visibility, niche subject matter, or lack of relevance to current research trends. The absence of citations for these articles raises questions about their impact and potential contribution to the research landscape in the field.

The mainstream topic of KGs (*application of knowledge-based systems*) was located in Quadrant I and was defined as a motor cluster (Cluster 9 in Table 11). This cluster contained the most extensive themes within this field. The concepts were the focal points of the domain. It was the most robust and mature cluster in a central position in the field. Cluster 9 implied that KGs could be regarded

as a complementary technology for knowledge-based systems (Chicaiza and Valdiviezo-Diaz, 2021). Systems can improve their effectiveness, ability, and accuracy in problem-solving tasks by using KGs as a source of knowledge, as well as rules and reasoning techniques to generate new knowledge. Recently, KGs have been extensively used in various AI systems (Mohamed et al, 2021). This finding aligns with the outcome of Ji et al (2021) research, which emphasized that exploring knowledge-aware applications is a significant area of study within the KG field.

Knowledge-based systems like recommendation systems, question-answering, and expert systems can reason about the relationships between entities in the KG to answer complex queries. Rule-based systems generally apply ontology-based rules as common knowledge representations to build intelligent systems that can perform complex tasks by reasoning about domain-specific knowledge (Tian et al, 2022).

Clusters 1 (*ontology-based interoperability of knowledge*) and 4 (*knowledge engineering for task analysis*) were located in Quadrant II as niche themes. These clusters were considered an ivory tower; not axial but well-developed, prominent, and isolated. This kind of isolation is due to switching, maturation, and developing into higher-level subjects based on some trends or research frontiers. Regarding Cluster 1, exchanging information according to rules, meanings, and shared structure of concepts in an ontology between various knowledge-based or information systems can be considered ontology-based interoperability. Ontologies can represent structured data as common vocabulary to provide a better understanding of entities and relationships with a KG background. Ontologies also can share knowledge across different domains, contexts, and systems. Therefore, by using the facilities of ontologies, various systems can interact and share information more efficiently (Trausan-Matu, 2009). In other words, ontology-based knowledge interoperability can resolve problems like inconsistency, semantic interoperability, and knowledge reusability (Fraga et al, 2020). As a result, ontologies can be regarded as a robust tool to strengthen semantic and ontology-based interoperability (Liyanage et al, 2015).

Cluster 4 indicates that knowledge engineering considers representing and utilizing knowledge in knowledge-based, decision-support, cognitive-driven, and intelligence systems (Hoffman et al, 1998). A knowledge engineer elicits its knowledge from domain experts and transfers it into rule languages, semantic models, and structured tools like ontologies or KGs (Simsek et al, 2023). Knowledge engineering for task analysis focuses on capturing, using, and representing knowledge to perform and enhance a particular task. It has a significant role in recognizing, organizing, and improving tasks and subtasks to augment task optimization and performance (Averbukh, 1995; Klein and Militello, 2001).

Regarding Cluster 3 (*application of knowledge graph in big data*), KGs can be used as a robust tool to analyze big data and process large and complex datasets. Mapping different datasets (e.g., large big data) onto a KG can resolve inconsistencies, redundancies, and semantic gaps (Yu et al, 2021). Utilizing KGs from a big data perspective confers benefits such as data integration, discovering hidden relationships, data exploration, context-based entity matching, pattern recognition, complex reasoning, enriching semantic search, enhancing context-aware recommendations, and designing domain-specific applications (Bellomarini et al, 2017; Hussey et al, 2021; Tasnim et al, 2020). Therefore, KGs are crucial in assisting various applications in the big data era, such as aiding decision-making, providing personalized recommendations, supporting question-answering, and dialogue (Janev et al, 2020).

Quadrant IV shows the basic, underdeveloped, immature, central, and transversal themes, e.g., Clusters 2, 5, 6, 7, and 8. These themes will be extended more in the future as revolutionary ideas aiming to consider paradigms such as *knowledge graph analysis based on ML methods*.

As for Cluster 2, knowledge representation and reasoning (KRR, KR&R, KR^2) as a subfield of AI concerns representing knowledge by various techniques and formalisms like logic-based languages (e.g., description logic, first-order logic), ontologies, frames, and KGs (Tiwarei et al, 2024). As a result, KGs and KRR are closely interconnected. KGs utilize KRR techniques to enable reasoning and inference knowledge. In summary, reasoning is a major KG technology, which represents the process of inferring (Li et al, 2022a; Peng et al, 2023) to resolve complicated tasks, such as medical diagnosis and natural language dialogue. This finding aligns with the outcomes of Chen et al (2021b) study, emphasizing that knowledge reasoning and KG representation are important subjects within the field.

Regarding Cluster 5, *knowledge graph querying* conveys techniques and procedures for extracting information, extracting valuable attributes from interconnected knowledge representations, and retrieving the most relevant data from entities and relationships of a KG (Opdahl et al, 2022). Querying KGs is essential for various applications, including semantic search, fact-checking, question answering

(QA), semantic search, recommendation systems, and personal assistants (Khan, 2023). The queries are expressed in the semantic context in various query languages like SPARQL (the RDF query language) (Liang et al, 2020), Cypher (the graph query language) (Francis et al, 2018), and GraphQL (a flexible query language) (Ni et al, 2022). By querying the KG, knowledge discovery and semantic data exploration become more efficient, and scholars can access a semantic network of interconnected and meaningful information (Fahd et al, 2022).

Concerning Cluster 6, *knowledge graph analysis based on ML methods* includes ML techniques to extract purposeful information from KGs. Different tasks, such as link prediction (Kanakaris et al, 2021), KG completion (Lan et al, 2021), graph embedding (Mohamed et al, 2021; Tu et al, 2022; Zhu et al, 2019), entity classification (Li et al, 2009), reinforcement learning for KGs (Sakurai et al, 2022), and graph convolutional networks (GCNs) (Chang et al, 2021) can be applied to KGs according to ML methods. These methods can analyze the structure of KGs and entity attributes to generate knowledge, enhance context-aware recommendations, learn patterns and features from the graph, make predictions, and learn embedding by leveraging the graph topology and entity attributes (Bakal et al, 2018; Wang et al, 2021). As Tiddi and Schlobach (2022) emphasized, large-scale KGs could be integrated into explainable ML systems as a new generation to present more meaningful explanations in the era of combined symbolic and sub-symbolic methods. Furthermore, Chen et al (2021a) revealed that KG embedding is the most researched topic by authors.

As for Cluster 7, *knowledge graph extraction* is a process of collecting and integrating information from structured and unstructured sources to craft a KG. Some steps, like entity extraction, relation extraction, semantic annotation, and KG construction, are considered in the process. Some named entity recognition (NER) and NLP techniques and algorithms are utilized to achieve the goal of extraction (Li et al, 2022b; Mahon, 2020; Turki et al, 2021; Wei et al, 2023; Zuo et al, 2022).

Cluster 8 signifies *knowledge discovery and data mining* (KDD) as an interdisciplinary area that, mainly in the context of computer sciences, refers to extracting knowledge from data by focusing on data mining techniques (Cao, 2012). Moreover, data mining tries to extract useful information and meaningful patterns from large and complex datasets to make knowledge-driven decisions and predictions (Shu and Ye, 2022). Data mining is the core of the KDD process, which explores data and discovers previously unidentified patterns (Maimon and Rokach, 2010). Knowledge discovery can be enhanced by semantic approaches like semantic graph embeddings on large KGs (Dörpinghaus and Jacobs, 2020). Moreover, some advanced data-mining approaches, such as the graph-based approach as a complementary technique, can help improve

knowledge discovery (Holzinger et al, 2014). Similar to the findings of Salatino et al (2021), data mining emerged as a prominent anticipated research direction.

Based on Fig. 11, niche themes recognized by the Walktrap algorithm (located in Quadrant II) were labeled ‘*knowledge graph*’, ‘*knowledge engineering*’, and ‘*knowledge discovery*’. They may not be the most studied topics in KGs, but they have evolved and advanced distinctly. Their well-developed nature signifies that scholars find value in deeply exploring these topics. Their placement in Quadrant II highlights their distinctiveness and specialization within the KG field. Therefore, KG professionals can concentrate on particular research areas such as KG construction, engineering techniques, and discovery algorithms. They can make a special contribution to understanding the diverse aspects of KGs. As Zhong et al (2023) emphasized, KG construction and feature engineering are prominent in empowering human intelligence within AI applications. In addition, Chen et al (2021b) noted that constructing KGs is a popular study area.

‘*Ontology*’, ‘*semantic web*’, and ‘*linked data*’ were introduced as basic themes by the Walktrap algorithm. They were considered foundational concepts within the KG domain. Their placement (Quadrant IV) highlighted their foundational nature and pivotal role in the KG field. These fundamental themes provide the conceptual infrastructure for fruitful knowledge representation. As a result, KG scholars can uncover potential topics for improvements and advancements through these basic themes. According to Saeed et al (2017), semantic web technologies are founded on integrating linked data principles and semantic tools, e.g., ontologies, to foster data organization and integration. Shao et al (2021) also support the idea that ontology and other semantic technologies guarantee semantic developments, such as context-aware recommendation systems.

Some reasons led to the emergence of specific trends, such as ‘*data mining*’, ‘*natural language processing*’, ‘*formal concept analysis*’, and ‘*ontology*’ during 2013–2018 in the field of KGs. Semantic developers and practitioners extensively highlighted the availability of large-scale data and the growing need for techniques and tools to mine and extract information and knowledge, like text mining, NLP understanding, sentiment analysis, and information extraction to represent domain-specific knowledge in a formal and machine-understandable. Moreover, formal concept analysis presented a mathematical framework to help discover hierarchical relationships and pattern recognition within KGs (DeMaio et al, 2012; Gardiner and Gillet, 2015).

The evolution of ‘*knowledge graph*’, ‘*ontology*’, and ‘*gene ontology*’ during 2018–2020 reflects the maturation and utilization of KG-based approaches in various fields (Li et al, 2020; Zhou et al, 2021), domain ontology development for knowledge representation (Sattar et al, 2020), and the growth of gene-related knowledge focusing on gene on-

tology as a standard tool in the bioinformatics arsenal and biomedical domains (Pomaznoy et al, 2018).

Topical evolution during 2021–2023 indicated six themes, some of which were previously discussed. The trends of ML and KGs have been closely intertwined (Kose et al, 2023). KGs can be enhanced and enriched technically using ML algorithms by automatically performing entity disambiguation, improving entity classification, and predicting missing or incomplete information within the graphs (Abu-Salih, 2021; Ji et al, 2021). On the other hand, ML models can be optimized for better performance by structured knowledge provided by KGs and graph-based learning tasks. This convergence can empower capabilities for leveraging the abilities of both to extract new knowledge and improve the quality of knowledge representation and reasoning (Chai et al, 2023; Kose et al, 2023).

In this period, the emphasis on ‘*protein*’ indicates a particular focus on modeling and integrating protein-related knowledge, like protein-protein interactions (PPIs). PPIs play a fundamental role in biological processes and are a major means of function and signaling in biological systems. The results of some ML methods, such as deep learning techniques (sequence-based), were promising and informative (Yang et al, 2020) in PPIs. The prediction of interactions in protein networks, such as hierarchical graph learning (Gao et al, 2023), prediction of PPIs using graph neural networks (Jha et al, 2022), link prediction by attributed graph embedding (Nasiri et al, 2021), graph-based deep learning methods for PPI prediction, and graph representation learning methods (Yang et al, 2020) were prospective and encouraging, leading the trends in KG in the domain of bioinformatics. According to Turki et al (2023), topics such as entity linking, KG embedding, and graph neural networks have received considerable interest.

In terms of practical implications, semantic scholars, ontology developers, designers of KGs, and policy-makers can establish collaboration by leveraging the results of the present study as a thematic policy map to avoid repeated research and make informed decisions. Moreover, practitioners can harness the power of KGs to drive innovation and advancement. For example, financial services can be improved through KGs. KGs can detect complex patterns and relationships within financial data, making them valuable for fraud detection and prevention.

6. Conclusion

This study aimed to summarize recent advancements and potential directions to support future research steps related to KGs. Predictably, due to its interdisciplinary nature, the field of KGs continues to contribute novel approaches and applications in the future.

Nine major topic clusters were identified based on the co-occurrence network. The most mature and mainstream thematic cluster was the *application of knowledge-based systems*. The most well-developed themes were *ontology*-

based interoperability of knowledge and knowledge engineering for task analysis. The most undeveloped and chaotic theme was the *application of KGs in big data*.

Five basic themes with high centrality and low density, which indicated central but not developed themes, were labeled as *knowledge representation and reasoning*, *knowledge graph querying*, *knowledge graph analysis based on ML methods*, *knowledge graph extraction*, and *knowledge graph classification*.

Six clusters experienced significant developments between 2021 and 2023: ‘ontology’, ‘natural language processing’, ‘machine learning’, ‘protein’, ‘knowledge engineering’, and ‘knowledge graph’.

The Walktrap algorithm was utilized to identify six clusters within the network structure. ‘Knowledge graph’, ‘knowledge engineering’, and ‘knowledge discovery’ clusters were recognized as niche themes. The basic clusters were ‘ontology’, ‘semantic web’, and ‘linked data’.

Various findings by different methods indicated significant coherence in the field of KGs. It means that the identified themes and trends were aligned towards a common direction or frontier.

This study contributes to clarifying the current state and boundaries of knowledge in the field of KGs. Regarding theoretical implications, the findings highlight that innovative methods, such as integrating machine learning with knowledge engineering, can significantly enhance tools for analyzing complex data. Additionally, identified themes such as query-based KG analysis and KG classification pave the way for advancements in knowledge representation and reasoning theories. The results of this study have practical applications across various domains. For instance, themes related to NLP and medical data demonstrate high potential for developing advanced systems for medical text analysis and disease prediction. They hold significant potential for improving advanced text analytics in healthcare and disease prediction. For example, in medical imaging, KGs are being integrated with AI to enhance diagnostic capabilities. For example, by linking imaging data with textual knowledge from clinical records, KGs help AI systems make more informed and explainable diagnoses (Wang et al, 2022). In biomedical research, KGs have been instrumental in connecting disparate datasets to uncover relationships between genes, proteins, and diseases (Tanaka et al, 2025). Moreover, KG analysis can improve recommendation systems by enabling more accurate and context-aware suggestions. These applications can inform the design of more effective recommendation systems and support better decision-making in big data environments.

This study had certain limitations. The data were restricted to the WoS Core Collection from the Clarivate Analytics database and three research areas. A limited timeline (2013–2023) was chosen for the investigation. There is also limited understanding regarding the limitations and disadvantages of visualization software.

Despite the contributions of this study, significant research gaps remain. Therefore, future studies are advised to examine data from scholarly citation databases, such as Scopus and Dimensions, in various subject areas over a longer period. It is also suggested that topic trends be explored using other topic modeling algorithms, such as LDA and STM, to uncover hidden trend patterns. Predicting topic trends using deep neural networks for modeling past research and forecasting future topics is also recommended. Future studies should analyze topic clusters with an interdisciplinary approach in subfields of KGs, such as KG embedding, KG completion, KG propagation, multimedia KG construction, KG analytics, spatial KG, visual KG, KGs and explainable AI (Shaikh et al, 2024), probabilistic KGs, graph-based data science, representation of knowledge in images (Rosa et al, 2024), KGs and deep semantics, semantic AI, KGs in medical imaging analysis (Yang et al, 2024), and KGs and large language models (LLMs) (Soos and Haroutunian, 2023). In particular, the use of KGs for analyzing multimodal data and developing explainable models in AI requires further exploration. Future research is encouraged to investigate interdisciplinary applications of KGs in subfields such as KG embedding, probabilistic KGs, and their integration with LLMs. Additionally, leveraging advanced algorithms like deep neural networks for forecasting topic trends is highly notable for future directions. This research not only clarifies the current trends in KGs but also outlines actionable directions for future studies. By addressing the identified gaps and leveraging the proposed methodologies, researchers can contribute to the theoretical, methodological, and practical advancements in this rapidly evolving field.

Availability of Data and Materials

The data used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

EH: Conceptualization, Data Curation, Formal Analysis, Software, Supervision, Methodology, Visualization, Project Administration, Writing—original draft, Writing—review & editing. AS: Data curation, Validation, Writing—review & editing, Formal Analysis. AD: Resources, Writing—review & editing, Investigation, Validation. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

The authors declares no conflict of interest. EH serves as editorial board member of this journal. EH declares that she was not involved in the processing of this article and has no access to information regarding its processing.

References

- Abu-Salih B. Domain-Specific KGs: A Survey. *Journal of Network and Computer Applications*. 2021; 185: 103076. <https://doi.org/10.1016/j.jnca.2021.103076>
- Aria M, Cuccurullo C. Bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. *Journal of Informetrics*. 2017; 11: 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Averbukh E. Multidisciplinary Task Knowledge Engineering. *IFAC Proceedings*. 1995; 28: 161–166. [https://doi.org/10.1016/S1474-6670\(17\)46611-0](https://doi.org/10.1016/S1474-6670(17)46611-0)
- Back C, Doleck T. Educational Data Mining: A Bibliometric Analysis of an Emerging Field. *IEEE Access*. 2022; 10: 31289–31296. <https://doi.org/10.1109/ACCESS.2022.3160457>
- Bakal G, Talari P, Kakani E, Kavuluru R. Exploiting semantic patterns over biomedical KGs for predicting treatment and causative relations. *J Biomed Inform*. 2018; 82: 189–199. <https://doi.org/10.1016/j.jbi.2018.05.003>
- Bellomarini L, Gottlob G, Pieris A, Sallinger E. ‘Swift Logic for Big Data and Knowledge Graphs’, Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne, Australia, August 19–25, 2017. Publisher: IJCAI. 2017. <https://doi.org/10.24963/ijcai.2017/1>
- Borgatti S, EverettUcinet M, Freeman L. UCINET for Windows: Software for social network analysis. 2002. Available at: http://www.researchgate.net/publication/216636663_UCINET_for_Windows_Software_for_social_network_analysis(Accessed:22July2023).
- Börner K, Chen C, Boyack K. Visualizing Knowledge Domains. *Annual Review of Information Science and Technology*. 2003; 37: 179–255. <https://doi.org/10.1002/aris.1440370106>
- Brachman R, Schmolze J. An Overview of the KL-ONE Knowledge Representation System. *Cognitive Science*. 1985; 9: 171–216. [https://doi.org/10.1016/s0364-0213\(85\)80014-8](https://doi.org/10.1016/s0364-0213(85)80014-8)
- Brusco M, Steinley D, Watts A. A Comparison of Spectral Clustering and The Walktrap Algorithm for Community Detection in Network Psychometrics. *Psychological Methods*. 2022. <https://doi.org/10.1037/met0000509>
- Cao L. Actionable Knowledge Discovery and Delivery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2012; 2: 149–163. <https://doi.org/10.1002/widm.1044>
- Chai C, Wang J, Luo Y, Niu Z, Li G. Data Management for Machine Learning: A Survey. *IEEE Transactions on Knowledge and Data Engineering*. 2023; 35: 4646–4667. <https://doi.org/10.1109/TKDE.2022.3148237>
- Chang D, Lin E, Brandt C, Taylor RA. Incorporating Domain Knowledge Into Language Models by Using Graph Convolutional Networks for Assessing Semantic Textual Similarity: Model Development and Performance Comparison. *JMIR Med Inform*. 2021; 9: e23101. <https://doi.org/10.2196/23101>
- Chaudhri V, Baru C, Chittar N, Dong XL, Genesereth M, Hendler J, et al. Knowledge Graphs: Introduction, History, and Perspectives. *AI Magazine*. 2022; 43: 17–29. <https://doi.org/10.1002/aaai.12033>
- Chen K, Xie B, Deng S. Theme Evolution of Research on KGs Based on Visualization Analyses of Data. *Journal of Physics Conference Series*. 2021a; 1813: 012039. <https://doi.org/10.1088/1742-6596/1813/1/012039>
- Chen X, Xie H, Li Z, Cheng Gary. Topic Analysis and Development in Knowledge Graph Research: A Bibliometric Review on Three Decades. *Neurocomputing*. 2021b; 461: 497–515. <https://doi.org/10.1016/j.neucom.2021.02.098>
- Chen X, Li J, Sun X, Wu D. Early Identification of Intellectual Structure Based on Co-Word Analysis from Research Grants. *Scientometrics*. 2019; 121: 349–369. <https://doi.org/10.1007/s11192-019-03187-9>
- Chicaiza J, Valdiviezo-Diaz P. A Comprehensive Survey of Knowledge Graph-Based Recommender Systems: Technologies, Development, and Contributions. *Information*. 2021; 12: 232. <https://doi.org/10.3390/info12060232>
- Cho J. Intellectual Structure of The Institutional Repository Field: A Co-Word Analysis. *Journal of Information Science*. 2014; 40: 386–397. <https://doi.org/10.1177/0165551514524686>
- Clarivate. Web of Science. 2023. Available at: <https://clarivate.libguides.com/directlinks> (Accessed: 5 April 2023).
- Cobo M, López HAG, Herrera VE, Herrera Francisco. Science Mapping Software Tools. *Journal of The American Society for Information Science and Technology*. 2011; 62: 1382–1402. <https://doi.org/10.1002/asi.21525>
- DeMaio C, Fenza G, Loia V, Senatore S. Hierarchical Web Resources Retrieval by Exploiting Fuzzy Formal Concept Analysis. *Information Processing & Management*. 2012; 48: 399–418. <https://doi.org/10.1016/j.ipm.2011.04.003>
- Ding Y. Semantic Web: Who is Who in The Field—a Bibliometric Analysis. *Journal of Information Science*. 2010; 36: 335–356. <https://doi.org/10.1177/0165551510365295>
- Dörpinghaus J, Jacobs M. Knowledge Detection and Discovery Using Semantic Graph Embeddings on Large Knowledge Graphs Generated on Text Mining Results. 2020 Federated Conference on Computer Science and Information Systems. Sofia, Bulgaria, September 6–9, 2020. IEEE: Poland. 2020. <https://doi.org/10.15439/2020F36>
- DosSantos BS, Steiner MT, Fenerich AT, Lima RHP. Data Mining and Machine Learning Techniques Applied to Public Health Problems: A Bibliometric Analysis from 2009 to 2018. *Computers & Industrial Engineering*. 2019; 138: 106120. <https://doi.org/10.1016/j.cie.2019.106120>

- Fahd K, Yuan M, Miah SJ, Venkatraman S, Ahmed KEU. Knowledge Graph Model Development for Knowledge Discovery in Dementia Research Using Cognitive Scripting and Next-Generation Graph-Based Database: A Design Science Research Approach. *Social Network Analysis and Mining*. 2022; 12: 61. <https://doi.org/10.1007/s13278-022-00894-9>
- Fan C, Meiling C, Xinghua W, Jiayuan W, Bufu H. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in energy research*. 2021; 9: 652801. <https://doi.org/10.3389/fenrg.2021.652801>
- Fraga A, Vegetti M, Leone H. Ontology-Based Solutions for Interoperability Among Product Lifecycle Management Systems: A Systematic Literature Review. *Journal of Industrial Information Integration*. 2020; 20: 100176. <https://doi.org/10.1016/j.jii.2020.100176>
- Francis N, Green A, Guagliardo P, Libkin L, Lindaaker T, Marsault V, et al. Cypher: An Evolving Query Language for Property Graphs. *International Conference on Management of Data*. Houston, TX, USA, June 10–15, 2018. ACM: New York, NY, USA. 2018. <https://hal.science/hal-01803524>
- Gao Z, Jiang C, Zhang J, Jiang X, Li L, Zhao P, et al. Hierarchical Graph Learning for Protein–Protein Interaction. *Nature Communications* 2023; 14: 1093. <https://doi.org/10.1038/s41467-023-36736-1>
- Gardiner E, Gillet V. Perspectives on Knowledge Discovery Algorithms Recently Introduced in Chemo Informatics: Rough Set Theory, Association Rule Mining, Emerging Patterns, and Formal Concept Analysis. *Journal of Chemical Information and Modeling*. 2015; 55: 1781–1803. <https://doi.org/10.1021/acs.jcim.5b00198>
- Greenacre M, Blasius J. Multiple Correspondence Analysis and Related Methods. CRC Press: NewYork. 2006.
- Guizzardi G, Nicola G. Explanation, semantics, and ontology. *Data & Knowledge Engineering*. 2024; 153: 102325. <https://doi.org/10.1016/j.datak.2024.102325>
- Hall W, Tiropanis T. Web Evolution and Web Science. *Computer Networks*. 2012; 56: 3859–3865. <https://doi.org/10.1016/j.comnet.2012.10.004>
- Hoffman R, Crandall B, Shadbolt N. Use of The Critical Decision Method to Elicit Expert Knowledge: A Case Study in The Methodology of Cognitive Task Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 1998; 40: 254–276. <https://doi.org/10.1518/001872098779480442>
- Holzinger A, Dehmer M, Jurisica I. Knowledge Discovery and Interactive Data Mining in Bioinformatics-State-of-The-Art, Future Challenges and Research Directions. *BMC Bioinformatics*. 2014; 15: 1–9. <https://doi.org/10.1186/1471-2105-15-S6-II>
- Hosseini E, Ghaebi A, Baradar R. Bibliometrics and Mapping of Co-Words in The Field of Linked Data. *Scientometrics Research Journal*. 2021; 7: 91–116. <https://www.sid.ir/paper/394845/en>
- Hosseini E, Taghizadeh MK, Sabetnasab M. Development and Maturity of Co-Word Thematic Clusters: The Field of Linked Data. *Library Hi Tech ahead-of-print*. (ahead-of-print) 2023. <https://doi.org/10.1108/LHT-10-2022-0488>
- Hu C, Hu J, Deng S, Liu Y. A Co-Word Analysis of Library and Information Science in China. *SpringerLink*. 2013; 97: 369–382. <https://doi.org/10.1007/s11192-013-1076-7>
- Hussey P, Das S, Cert S, Ledger L, Spencer A. A Knowledge Graph to Understand Nursing Big Data: Case Example for Guidance. *Journal of Nursing Scholarship*. 2021; 53: 323–332. <https://doi.org/10.1111/jnu.12650>
- IBM Corp. IBM SPSS Statistics for Windows, Version 22.0. IBM Corp, Armonk NY. 2013. Available at: [https://www.scirp.org/reference/referencespapers?referenceid=2010524\(A](https://www.scirp.org/reference/referencespapers?referenceid=2010524(A) accessed:7June2023).
- Isfandyari MA, Saberi MK, Tahmasebi LS, Mohammadian S, Naderbeigi F. Global Scientific Collaboration: A Social Network Analysis and Data Mining of The Co-Authorship Networks. *Journal of Information Science*. 2023; 49: 1126–1141. <https://doi.org/10.1177/01655515211040655>
- Janev V, Pujić D, Jelić M, Vidal ME. Chapter 9 Survey on Big Data Applications. In Janev V, Graux D, Jabeen H, Sallinger E (eds.) *Knowledge Graphs and Big Data Processing* (pp. 149–164). 1st ed. Springer: Cham. 2020. https://doi.org/10.1007/978-3-030-53199-7_9
- Jaya I, Fizal M, Nincarean D. Systematic Description of The Internet of Things: A Bibliometric Analysis. *Journal of Theoretical and Applied Information Technology*. 2022; 100: 2835–2853.
- Jha K, Saha S, Singh H. Prediction of Protein-Protein Interaction Using Graph Neural Networks. *Scientific Reports*. 2022; 12: 8360. <https://doi.org/10.1038/s41598-022-12201-9>
- Ji S, Pan S, Cambria E, Marttinen P, Philip SY. A Survey on KGs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*. 2021; 33: 494–514. <https://doi.org/10.1109/TNNLS.2021.3070843>
- Kanakaris N, Giarelis N, Siachos I, Karacapilidis N. Shall I Work with Them? A Knowledge Graph-Based Approach for Predicting Future Research Collaborations. *Entropy*. 2021; 23: 664. <https://doi.org/10.3390/e23060664>
- Khan A. Knowledge Graphs Querying. *ACM SIGMOD Record* 2023. 2023; 52: 18–29. <https://doi.org/10.48550/arXiv.2305.14485>
- Khasseh A, Soheili F, Moghaddam HS, Chelak AM. Intellectual Structure of Knowledge in iMetrics: A Co-Word Analysis. *Information Processing & Management*. 2017; 53: 705–720. <https://doi.org/10.1016/j.ipm.2017.02.001>
- Klein G, Militello L. Some Guidelines for Conducting a Cognitive Task Analysis. In Salas E. (ed.) *Advances in Human Performance and Cognitive Engineering Research*. 163–199. Elsevier Science/JAI Press: Leeds. 2001. [https://doi.org/10.1016/S1479-3601\(01\)01006-2](https://doi.org/10.1016/S1479-3601(01)01006-2)
- Kong H, Wang K, Qiu X, Cheung C, Bu N. 30 Years of Artificial Intelligence (AI) Research Relating to The Hospitality and Tourism Industry. *International Journal of Con-*

- temporary Hospitality Management. 2023; 35: 2157–2177. <https://doi.org/10.1108/IJCHM-03-2022-0354>
- Kose OD, Shen Y, Mateos G. Fairness-Aware Graph Filter Design. arXiv. 2023; 2303: 11459. (preprint)
- Kumar A, Steyvers M, Balota D. A Critical Review of Network-Based and Distributional Approaches to Semantic Memory Structure and Processes. *Topics in Cognitive Science*. 2022; 14: 54–77. <https://doi.org/10.1111/tops.12548>
- Kumari R, Jeong JY, Lee BH, Choi KN, Choi K. Topic Modeling and Social Network Analysis of Publications and Patents in Humanoid Robot Technology. *Journal of Information Science*. 2019; 47: 016555151988787. <https://doi.org/10.1177/0165551519887878>
- Kyaw ATZ, Wang Z. Mapping The Intellectual Structure of The Linked Data Field: A Co-Word Analysis and Social Network Analysis. *International Journal of Advanced Research in Science, Engineering and Technology*. 2018; 5: 6632–6647. <https://www.ijarset.com/upload/2018/august/27-IJARSET-AINT.pdf>
- Lan Y, He S, Liu K, Zeng X, Liu S, Zhao J. Path-Based Knowledge Reasoning with Textual Semantic Information for Medical Knowledge Graph Completion. *BMC Medical Informatics and Decision Making*. 2021; 21: 1–12. <https://doi.org/10.1186/s12911-021-01622-7>
- Lassila O, Swick R. Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation. 1998. <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/> (Accessed: 15 April 2023).
- Lazarides M, Lazaridou IZ, Papanas N. Bibliometric Analysis: Bridging Informatics with Science. *The International Journal of Lower Extremity*. 2023. <https://doi.org/10.1177/15347346231153538>
- Li P, Wang X, Liang H, Suzhi Z, Jiang Y, Tang Y. A Fuzzy Semantic Representation and Reasoning Model for Multiple Associative Predicates in a Knowledge Graph. *Information Sciences*. 2022a; 599: 208–230. <https://doi.org/10.1016/j.ins.2022.03.079>
- Li X, Chen CH, Zheng P, Wang Z, Jiang Z, Jiang Z. A knowledge graph-aided concept-knowledge approach for evolutionary smart product-service system Development. *Journal of Mechanical Design*. 2020; 142: 101403. <https://doi.org/10.1115/1.4046807>
- Li Y, Lin H, Yang Z. Incorporating Rich Background Knowledge for Gene-Named Entity Classification and Recognition. *BMC Bioinformatics*. 2009; 10: 1–15. <https://doi.org/10.1186/1471-2105-10-223>
- Li Z, Zhong Q, Yang J, Duan Y, Wang W, Wu C, et al. DeepKG: An End-to-End Deep Learning-Based Workflow for Biomedical Knowledge Graph Extraction, Optimization, and Applications. *Bioinformatics*. 2022b; 38: 1477–1479. <https://doi.org/10.1093/bioinformatics/btab767>
- Liang S, deFarias TM, Anisimova M, Gil M. Querying Knowledge Graphs in Natural Language. *Journal of Big Data*. 2020; 8: 3. <https://doi.org/10.1186/s40537-020-00383-w>
- Lissandrini M, Gianmarco P, Gianmaria S. The ESW of Wikidata: Exploratory search workflows on Knowledge Graphs. *Journal of Web Semantics*. 2025; 85: 100860. <https://doi.org/10.1016/j.websem.2024.100860>
- Liu X, Liu Z. Knowledge Engineering Research Topic Mining Based on Co-Word Analysis. *The 31st International Conference on Software Engineering and Knowledge Engineering*. Lisbon, Portugal, July 10–12, 2019. KSI Research Inc. and Knowledge Systems Institute Graduate School: USA. 2019. <https://doi.org/10.18293/SEKE2019-047>
- Liyanage H, Krause P, de LS. Using Ontologies to Improve Semantic Interoperability in Health Data. *Journal of Innovation in Health Informatics*. 2015; 22: 170. <https://doi.org/10.14236/jhi.v22i2.170>
- Löebner S. Frames at The Interface of Language and Cognition. *Annual Review of Linguistics*. 2021; 7: 261–284. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3778512
- Mahon L., Giunchiglia E., Li B., Lukasiewicz T. ‘Knowledge Graph Extraction from Videos’, 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). Miami, FL, USA, December 14–17, 2020. IEEE: Piscataway, NJ. 2020. <https://doi.org/10.1109/ICMLA51294.2020.00014>
- Maimon O, Rokach L. Introduction to Knowledge Discovery and Data Mining. *Data Mining and Knowledge Discovery Handbook* (pp. 1–15). Springer US: Boston, MA. 2010.
- McGuinness D, VanHarmelen F. OWL Web Ontology Language Overview. W3C Recommendation. 2004; 10. Available at: <https://www.w3.org/TR/owl-features/> (Accessed: 15 April 2023).
- Minsky M. A Framework for Representing Knowledge. In Winston P (ed). *The Psychology of Computer Vision* (pp. 27–211). McGraw Hill: New York. 1975.
- Mohamed S, Nounu A, Nováček V. Biological Applications of Knowledge Graph Embedding Models. *Briefings in Bioinformatics*. 2021; 22: 1679–1693. <https://doi.org/10.1093/bib/bbaa012>
- Mokhtarpour R, Khasseh AA. Twenty-Six Years of LIS Research Focus and Hot Spots, 1990–2016: A Co-Word Analysis. *Journal of Information Science*. 2021; 47: 794–808. <https://doi.org/10.1177/0165551520932119>
- Morsey M, Lehmann J, Auer S, Stadler C, Hellmann S. DBpedia and The Live Extraction of Structured Data from Wikipedia. *Program Electronic Library and Information Systems*. 2012; 46: 157–181. <https://doi.org/10.1108/00330331211221828>
- Nasiri E, Berahmand K, Rostami M, Dabiri M. A Novel Link Prediction Algorithm for Protein-Protein Interaction Networks by Attributed Graph Embedding. *Computers in Biology and Medicine*. 2021; 137: 104772. <https://doi.org/10.1016/j.compbiomed.2021.104772>
- Ni P, Okhrati R, Guan SU, Chang V. Knowledge Graph and Deep Learning-Based Text-to-GraphQL Model for Intelligent Medical Consultation Chatbot. *Information Systems Frontiers*. 2022; 26: 1–19. <https://doi.org/10.1007/s10796-022-10295-0>
- Niknia M, Mirtaheri SL. Mapping a Decade of Linked Data

- Progress Through Co-Word Analysis. *Webology*. 2015; 12. https://www.researchgate.net/publication/301633071_Mapping_a_decade_of_linked_data_progress_through_co-word_analysis
- Opdahl A, Al MT, Dang NDT, Ocaña MG, Tessem B, Veres C. Semantic KGs for the News: A Review. *ACM Computing Surveys*. 2022; 55: 1–38. <https://doi.org/10.1145/3543508>
- Oramas S, Ostuni VC, Noia TD, Serra X, Sciascio ED. Sound and Music Recommendations with Knowledge Graphs. *ACM Transactions on Intelligent Systems and Technology*. 2016; 8: 1–21. <https://doi.org/10.1145/2926718>
- Pan X, Li X, Li Q, Hu Z, Bao J. Evolving to multi-modal knowledge graphs for engineering design: state-of-the-art and future challenges. *Journal of Engineering Design*. 2024; 1–40. <https://doi.org/10.1080/09544828.2023.2301230>
- Peng C, Xia F, Naseriparsa M, Osborne F. Knowledge Graphs: Opportunities and Challenges. *Artificial Intelligence*. 2023; 3: 1–32. (ahead of print) <https://doi.org/10.1007/s10462-023-10465-9>
- Pomaznoy M, Ha Brendan, Peters Bjoern. GOnet: A Tool for Interactive Gene Ontology Analysis. *BMC Bioinformatics*. 2018; 19: 1–8. <https://doi.org/10.1186/s12859-018-2533-3>
- Python Software Foundation. Python Language Reference, version 2.7. 2023. Available at: <http://www.python.org> (Accessed: 6 May 2023).
- Rosa PC, BarizonFilho AL, Valentim RT, Tognoli N. Datafication, Artificial Intelligence and Images: The Dominant Paradigm in the Representation of Knowledge in Images. *Knowledge Organization*. 2024; 51: 117–126.
- Saeed MR, Chelmiss C, Prasanna V. Automatic Integration and Querying of Semantic Rich Heterogeneous Data: Laying The Foundations for The Semantic Web of Things. In *Managing The Web of Things* (pp. 251–273). 2017. <https://doi.org/10.1016/B978-0-12-809764-9.00012-3>
- Sakurai K, Togo R, Ogawa T, Haseyama Miki. Controllable Music Playlist Generation Based on Knowledge Graph and Reinforcement Learning. *Sensors*. 2022; 22: 3722. <https://doi.org/10.3390/s22103722>
- Salatino AA, Mannocci A, Osborne F. Detection, Analysis, and Prediction of Research Topics with Scientific Knowledge Graphs. In book: *Predicting the Dynamics of Research Impact*. 2021. https://doi.org/10.1007/978-3-030-86668-6_11
- Sattar A, Salwana E, Surin M, Ahmad MN, YarKhanand R, Ahmad M, et al. Comparative Analysis of Methodologies for Domain Ontology Development: A Systematic Review. *International Journal of Advanced Computer Science and Applications*. 2020; 11. <https://doi.org/10.14569/IJACSA.2020.0110515>
- Shaikh N, Chauhan T, Patil J, Sonawane S. Explicable knowledge graph (X-KG): generating knowledge graphs for explainable artificial intelligence and querying them by translating natural language queries to SPARQL. *International Journal of Information Technology*. 2024; 16: 1605–1615. <https://doi.org/10.1007/s41870-023-01692-y>
- Shao B, Li X, Bian G. A Survey of Research Hotspots and Frontier Trends of Recommendation Systems from The Perspective of Knowledge Graph. *Expert Systems with Applications*. 2021; 165: 113764. <https://doi.org/10.1016/j.eswa.2020.113764>
- Shu X, Ye Y. Knowledge Discovery: Methods from Data Mining and Machine Learning. *Social Science Research*. 2022; 110: 102817. <https://doi.org/10.1016/j.ssresearch.2022.102817>
- Simsek U, Kärle E, Angele K, Huaman Elwin, Opdenplatz Juliette, Sommer Dennis, et al. A Knowledge Graph Perspective on Knowledge Engineering. *SpringerLink*. 2023; 4: 1–10. <https://doi.org/10.1007/s42979-022-01429-x>
- Sirichanya C, Kesorn K. Semantic data mining in the information age: A systematic review. *International Journal of Intelligent Systems*. 2021; 36: 3880–3916. <https://doi.org/10.1002/int.22443>
- Soos C, Haroutunian L. On The Question of Authorship in Large Language Models (LLMs). *NASKO*. 2023; 9: 1–17. <https://doi.org/10.7152/nasko.v9i1.16299>
- Tanaka T, Toshiaki K, Takeshi I. Predicting the effects of drugs and unveiling their mechanisms of action using an interpretable pharmacodynamic mechanism knowledge graph (IPM-KG). *Computers in Biology and Medicine*. 2025; 184: 109419. <https://doi.org/10.1016/j.combiomed.2024.109419>
- Tasnim M, Collarana D, Graux D, Vidal Maria-Esther. Chapter 8 Context-Based Entity Matching for Big Data. In book: *Knowledge Graphs and Big Data Processing*. 122–146. 2020. https://doi.org/10.1007/978-3-030-53199-7_8
- Tian L, Zhou X, Wu Y, Zhou W, Zhang J, Zhang T. Knowledge Graph and Knowledge Reasoning: A Systematic Review. *Journal Electronic Science and Technology*. 2022; 20: 100159. <https://doi.org/10.1016/j.jnlest.2022.100159>
- Tiddi I, Schlobach S. Knowledge Graphs as Tolls for Explainable Machine Learning: A Survey. *Artificial Intelligence*. 2022; 302: 103627. <https://doi.org/10.1016/j.artint.2021.103627>
- Tiwari S, Ortiz RF, BenAbbes S, Usoro UP, Hantach R. *Semantic AI in Knowledge Graphs*. CRC Press: Boca Raton, Fla. 2024.
- Trausan-Matu Stefan. Ontology-Based Interoperability in Knowledge-Based Communication Systems. In book: *Use of Ontologies to Support Information Interoperability* Editors: Roussay, Catherine. 2009. https://doi.org/10.1007/978-0-85729-724-2_9
- Tu E, Wang Z, Yang J, Kasabov N. Deep Semi-Supervised Learning Via Dynamic Anchor Graph Embedding in Latent Space. *Neural Networks*. 2022; 146: 350–360. <https://doi.org/10.1016/j.neunet.2021.11.026>
- Turki H, HadjTaieb MA, BenAouicha M, Fraumann G, Hauschke C, Heller L. Enhancing Knowledge Graph Extraction and Validation from Scholarly Publications Using Bibliographic Metadata. *Frontiers in Research Metrics and Analytics*. 2021; 6: 694307. <https://doi.org/10.3389/firma.2021.694307>
- Turki H, Owodunni AT, Taieb MA, Bile RF, Aouicha MB. A Decade of Scholarly Research on Open Knowledge Graphs.

- arXiv preprint arXiv.2023; 13186.
- VanEck NJ, Waltman L. VOSviewer Manual. 2014. Available at: https://www.researchgate.net/publication/265012369_VOSviewer_Manual(Accessed:17June2023).
- Verma S, Rajesh B, Sandeep H, Sanjay B. Scholarly knowledge graphs through structuring scholarly communication: a review. *Complex & intelligent systems*. 2023; 9: 1059–1095. <https://doi.org/10.1007/s40747-022-00806-6>.
- Wang S, Lin M, Ghosal T, Ding Y, Peng Y. Knowledge Graph Applications in Medical Imaging Analysis: A Scoping Review. *Health Data Science*. 2022; 2022: 9841548. <https://doi.org/10.34133/2022/9841548>.
- Wang Y, Dong L, Li Y, Zhang H. Multitask Feature Learning Approach for Knowledge Graph Enhanced Recommendations with RippleNet. *Plos One*. 2021; 16: e0251162. <https://doi.org/10.1371/journal.pone.0251162>
- Wei X, Chen Y, Cheng N, Cui X, Xu J, Han W. ColabKG: A Learnable Human-Machine-Cooperative Information Extraction Toolkit for (Event) Knowledge Graph Construction. *arXiv*. 2023; 2307: 00769. (preprint) <https://doi.org/10.48550/arXiv.2307.00769>
- Wu A, Ye Y. Bibliometric Analysis on Bibliometric-Based Ontology Research. *Science & Technology Libraries*. 2021; 40: 1–19. <https://doi.org/10.1080/0194262X.2021.1920555>
- Xie H, Zhang Y, Wu Z, Lv T. A Bibliometric Analysis on Land Degradation: Current Status, Development, and Future Directions. *Lan, MDPI*. 2020; 9: 1–37. <https://ideas.repec.org/a/gam/jlands/v9y20201p28-d310457.html>
- Yang F, Fan K, Song D, Lin H. “Graph-Based Prediction of Protein-Protein Interactions with Attributed Signed Graph Embedding”. *BMC Bioinformatics*. 2020; 21: 1–16. <https://doi.org/10.1186/s12859-020-03646-8>
- Yang P, Wang H, Huang Y, Yang S, Zhang Y, Huang L, et al. LMKG: A Large-Scale and Multi-Source Medical Knowledge Graph for Intelligent Medicine Applications. *Knowledge-Based Systems*. 2024; 284: 111323. <https://doi.org/10.1016/j.knosys.2023.111323>
- Yu C, Wang F, Liu Y, An L. Research on Knowledge Graph Alignment Model Based on Deep Learning. *Expert Systems with Applications*. 2021; 186: 115768. <https://doi.org/10.1016/j.eswa.2021.115768>.
- Zhong L, Wu J, Li Q, Peng H, Wu X. A Comprehensive Survey on Automatic Knowledge Graph Construction. *ACM Computing Surveys*. 2023; 56: 1–62. <https://doi.org/10.1145/3618295>
- Zhou B, Bao J, Li J, Lu Y, Liu T, Zhang Q. A Novel Knowledge Graph-Based Optimization Approach for Resource Allocation in Discrete Manufacturing Workshops. *Robotics and Computer-Integrated Manufacturing*. 2021; 71: 102160. <https://doi.org/10.1016/j.rcim.2021.102160>
- Zhu Q, Jiang X, Zhu Q, Pan M, He T. Graph Embedding Deep Learning Guides Microbial Biomarkers’ Identification. *Frontiers in Genetics*. 2019; 10: 1182. <https://doi.org/10.3389/fgene.2019.01182>
- Zhu Q, Kong X, Hong S, Li J, He Z. Global Ontology Research Progress: a Bibliometric Analysis. *Aslib Journal of Information Management*. 2015; 67: 27–54. <https://doi.org/10.1108/AJIM-05-2014-0061>
- Zuo H, Yin Y, Childs P. Patent-KG: Patent Knowledge Graph Extraction for Engineering Design. *Cambridge University Press: United Kingdom*. 2022.