# Scholar Collaboration Network Analysis Using Neo4j Graph Database: A Case Study in Digital Humanities (Postprint)

**Authors:** Xiong Huixiang, Huang Xiaojie, Chen Ziwei, Li Xinran

**Date:** 2023-10-08T00:00:00+00:00

## Abstract

[Purpose/Significance] Against the backdrop of deep digital development, digital humanities has become a field of deep interdisciplinary integration, with scholarly research collaborations becoming increasingly frequent. There is a need to analyze and mine these increasingly complex collaborative relationships to help scholars identify potential collaboration opportunities and promote academic exchange. [Method/Process] Taking scholars, institutions, and keywords as node data and co-authorship, citation, affiliation, and research topics as relationship data, we construct a scholar collaboration graph, store it using the graph database Neo4j, and utilize the Cypher query language and GDS algorithm library to analyze collaboration community discovery, core scholar identification, and collaboration trend prediction for scholars in the digital humanities field. [Results/Conclusion] Experimental results demonstrate that the Neo4j database effectively realizes the construction and graph analysis of scholar collaboration networks in the digital humanities field, enabling scholars to quickly identify interdisciplinary researchers highly associated with their own research interests and directions among numerous researchers, thereby promoting scholar collaboration and disciplinary development in the digital humanities field.

## Full Text

### Analysis of Scholar Collaboration Map Based on Graph Database Neo4j: A Case Study of the Digital Humanities Field

**Xiong Huixiang, Huang Xiaojie, Chen Ziwei, Li Xinran**

School of Information Management, Central China Normal University, Wuhan 430079

**Abstract:** [Purpose/Significance] In the context of deep digital development, digital humanities has emerged as a field of interdisciplinary integration where scholarly research collaboration is becoming increasingly frequent. Analyzing and mining these increasingly complex collaborative relationships is essential to help scholars identify potential cooperation opportunities and promote academic exchange. [Method/Process] This study constructs a scholar collaboration graph using scholars, institutions, and keywords as node data, and co-authorship, citation, affiliation, and research topics as relationship data. The graph is stored using the Neo4j graph database, and Cypher query language together with the Graph Data Science (GDS) algorithm library are employed to analyze cooperation community discovery, core scholar identification, and cooperation trend prediction among scholars in the digital humanities field. [Results/Conclusion] Experimental results demonstrate that Neo4j effectively facilitates the construction and graph analysis of scholar collaboration networks in digital humanities. It enables scholars to quickly identify interdisciplinary researchers whose interests and directions closely align with their own, thereby promoting scholarly cooperation and disciplinary development in digital humanities.

**Keywords:** Digital Humanities; Scholar Cooperation; Relationship Map; Neo4j

---

Digital humanities, as an interdisciplinary field bridging computer science and the humanities, encompasses a broad range of disciplines including linguistics, literature, library and information science, and computer science, evolving from the field of humanities computing [1]. In today' s era of deep digitalization, research on digital humanities has gained increasing momentum. Although Chinese academia started relatively late in this area, it has developed rapidly and shows promising prospects [2]. Currently, research on digital humanities in China primarily focuses on three aspects: investigation and analysis of foreign digital humanities projects, visualization of humanities and artistic resources using digital technologies, and applications of digital humanities in library, information, and archival science. Additionally, China has made breakthroughs in digital humanities education, with research institutions such as Shanghai Library, Renmin University of China Digital Humanities Research Center, Wuhan University Digital Humanities Research Center, and Peking University Information Management Department KVision Laboratory deeply promoting the integrated development of digital humanities and library/information/archival science [3].

Within this broad development platform, an increasing number of scholars in digital humanities have emerged, forming a vast and complex academic research network with diverse topics and frequent collaborations. However, a key focus of recent research on scientific cooperation prediction has been how to accurately identify collaborators with relevant research directions from the vast ocean of academic resources, scholars, and institutions. Therefore, analyzing scholar cooperation networks helps uncover patterns and trends in scholarly collaboration and understand cooperative relationships within core research teams [13].

Some scholars have begun exploring the use of graph databases for social network analysis. Guo Kunming [14] leveraged Neo4j' s analytical advantages for social relationships in heterogeneous networks, storing basic information and relationships of individuals crawled from Baidu Baike, and applied the Common Neighbors algorithm to calculate network structural similarity while using node attribute similarity to predict social relationships in the constructed heterogeneous network. M. Kolomeets et al. [15] used the graph database OrientDB to construct a VKontakte social network and employed PageRank to evaluate the most influential opinion leaders in social groups. Ding Hongli [16] built a multi-dimensional relationship network based on personnel information and call records using Neo4j, and conducted visualization, utilizing Neo4j' s query and analysis functions to mine personnel relationships, significantly improving experimental efficiency. Compared to traditional social network analysis tools, graph databases can display large-scale, constantly updated complex relationships between entities while enabling simpler and faster queries between network nodes and relationship values, offering inherent advantages in mapping real-world entities and relationships [17].

## Neo4j and Its Application Advantages

With the continuous development of the internet, graph databases have attracted significant attention from enterprises and scholars for their unique advantages in handling today's high-concurrency massive big data and real-time application scenarios. Graph databases store entities and their interrelationships using graphical data structures composed of nodes, properties, and edges, where nodes represent data entities, properties are attached information of nodes, and edges represent relationships between nodes. This structure is suitable for storing and managing large-scale data with complex and dynamic relationships [18]. Compared to traditional relational databases, graph databases process unstructured and unpredictable data, better aligning with the characteristics of explosive data growth and users' personalized needs. They effectively support associative relationships between entities, and when new tags and relationships are added, there is no need to adjust previous structures. Graph databases offer various extended functions including multi-layer association, shortest path, and centrality measurement, and are widely applied in social networks, recommendation systems, and relationship maps, representing a new tool for the big data era.

Common graph databases include Neo4j, FlockDB, GraphDB, and AllegroGraph, among which the open-source Neo4j has become one of the most widely used native graph databases due to its high performance, high stability, and strong scalability [19]. Neo4j uses native graph storage and processing, reflecting the essential nature of entity connections in relationship networks. It can return associated data through fast paths during queries, demonstrating highly efficient query performance. It supports the storage of unstructured data and large-scale data growth, adapts well to changing requirements, and offers great flexibility. Additionally, Neo4j can analyze and reason about complex relationships between entities, supporting logical language analysis and constraint-based reasoning. Neo4j has its own query language—Cypher—a declarative, expressive, and user-friendly graph query language [20] that uses simple keywords such as CREATE (for creating graph nodes, relationships, and properties), MATCH (for matching target information in existing graph databases), WHERE (as conditions for MATCH functions), and RETURN (for returning specified values after matching is completed). Based on these query statements, Neo4j enables analysis and reasoning of graph data.

### Reflecting Complex Scholar Cooperation Relationships

Cooperation relationships refer to collaborative behaviors among scholars during academic research. Common forms include co-authorship and citation relationships. In academic networks, more frequent co-authorship between two scholars suggests higher likelihood of shared interests and mutual trust. Furthermore, a scholar' s co-authors also collaborate with other scholars, forming a scholar co-authorship network based on these cooperative relationships. This relationship can be stored using graph structures, upon which social network analysis methods and graph mining algorithms can be applied to analyze and cluster relationships among scholars, thereby identifying the most suitable collaborators and research teams.

Another cooperation relationship among scholars is citation relationships, which include citing and being cited. Based on these two citation behaviors, scholars form a citation network that serves as a link between citing literature and cited achievements, reflecting the citer' s references, affirmations, and in-depth research on related issues. Typically, such citation relationships are used to achieve resource aggregation and scholar aggregation, with scholars as nodes and citation relationships between documents as edges connecting nodes, thereby constructing citation networks among relevant documents to better mine core scholars or core teams from citation relationship networks. Regardless of the cooperation relationship type, as research on related issues becomes more diversified, relationships among scholars grow increasingly complex. Neo4j is precisely capable of storing and reflecting such large, complex, and changing associative data, supporting large-scale data growth and updates while clearly presenting the relationships between various nodes.

### Real-time Querying of Target Scholars' Cooperation Relationships

Beyond storage capabilities, Neo4j' s retrieval function is exceptionally powerful, relying on the Cypher query language—a declarative graph database query language that is concise, expressive, highly efficient, and extensible, allowing users to customize their query methods. Cypher queries consist of four parts: START, MATCH, WHERE, and RETURN. START specifies one or more starting nodes in the graph, obtained through index lookup or directly through node IDs. MATCH is used for graph pattern matching and is a crucial part of instantiation. WHERE provides conditions to filter pattern matching results. RETURN specifies which nodes, relationships, and properties among the matched data should be returned to the client. Through this traversal-based lookup process, it is easy to locate and focus on desired scholar nodes, and through conditional matching, obtain the cooperation relationships of target scholars for targeted analysis. Additionally, Neo4j supports real-time updates to the graph database without affecting existing data structures, enabling continuous expansion of the current relationship graph to display increasingly complete and complex cooperation networks.

### Predicting Potential Cooperation Trends Among Scholars

Currently, there are two main methods for character relationship reasoning: ontology-based methods and graph database-based methods [21]. Ontology-based character relationship reasoning has high time complexity, and reasoning speed decreases rapidly with increasing amounts of character relationship data, making it difficult to meet the demands of character relationship reasoning in the big data era. In contrast, graph database-based character relationship reasoning represents a new trend in character relationship data analysis. The data storage structure and query methods of graph databases are based on graph theory, making them suitable for CRUD operations on character relationship data containing numerous connections. Graph database-based character relationship reasoning methods first convert character relationship data into graph database storage formats, then use graph database query languages for character relationship analysis [22]. As a highly efficient and scalable declarative graph query language with rich development patterns, Neo4j offers incomparable advantages for storing scholar relationship knowledge graphs. Complex relationship links also endow it with reasoning capabilities to predict potential cooperation trends among scholars, providing possible research directions for scientific cooperation across different fields and disciplines.

## Construction of Scholar Collaboration Relationship Map Based on Neo4j Graph Database

### Data Selection and Acquisition

This study selected CSSCI journals from CNKI Academic Resource Database as the data source, retrieving literature with "digital humanities" or "humanities computing" as themes. As of April 3, 2021, a total of 615 documents were retrieved. Data preprocessing was conducted using NoteExpress reference management software, removing duplicate literature, conference calls for papers, and documents not closely related to digital humanities themes, resulting in 334 valid documents. For documents with multiple authors, this study uniformly selected the top three authors as research objects. After removing duplicates, 410 scholar nodes, 244 institution nodes, and 636 keyword nodes were obtained, with example data processing results shown in [Figure 1: see original paper]. Python was then used to acquire co-authorship and citation relationships between scholars, employment relationships between scholars and institutions, and research topic relationships between scholars and keywords. This study primarily constructed the scholar cooperation graph based on these three types of nodes and four types of relationships, with the data model shown in [Figure 2: see original paper].

### Data Import

To store graph data in a graph database, a specific graph data model must be adopted, which concerns the implementation method for storing graph data. Common graph data models include property graphs, hypergraphs, and triples. Since the property graph model is intuitive and easy to understand, capable of describing the vast majority of graph usage scenarios, Neo4j adopts the most popular property graph model. First, both node and relationship data Excel files were saved as ".csv" files. Then, using Cypher language' s CREATE statement, node files and relationship files were input into the code editing area according to the code example shown in [Figure 3: see original paper]. The final operation result is shown in [Figure 4: see original paper], clearly displaying the number of nodes, number of relationships, and the scholar cooperation relationship graph. For a specific node, taking the Institute of Literature, Chinese Academy of Social Sciences as an example, querying this node reveals two scholars working at this institution, and their cooperating scholars, research topics, and other related relationships are clearly presented, as shown in [Figure 5: see original paper].

## Analysis of Scholar Cooperation Network

Faced with massive and complex unstructured relationship data, Neo4j provides an effective solution for technical applications. However, reviewing domestic literature reveals that current research utilizing Neo4j' s embedded graph algorithms and Cypher query language for data analysis and processing remains relatively limited. This study fully leverages Neo4j' s powerful graph algorithm capabilities to analyze scholar cooperation networks in the digital humanities research field. Neo4j' s Graph Data Science (GDS) library can implement various complex social network analyses, including centrality algorithms, community detection algorithms, path finding algorithms, and link prediction algorithms. This study employs relevant graph algorithms to realize scholar cooperation community discovery, core scholar identification, and scholar cooperation trend prediction, providing references for scholars in the digital humanities field to find their cooperation partners and resources from different perspectives.

### Cooperation Community Discovery

In recent years, digital humanities technology has developed rapidly, attracting more and more scholars to conduct extensive and in-depth research on related issues, thus forming a complex scholar network. Community structure is an important property in complex networks, manifested as densely connected nodes within communities and sparsely connected nodes between different communities [23]. It can cluster scholars with similar characteristics or common attributes, helping scholars discover and find peers with similar interests or interdisciplinary collaborators for mutual exchange.

Among community detection algorithms such as Louvain, Label Propagation, and Infomap, Louvain performs well in both efficiency and effectiveness and can discover hierarchical community structures. Guo Li et al. [24] used the classic dataset American College Football to compare the Louvain algorithm with commonly used overlapping community detection algorithms CPM, LFM, and COPRA, with results showing Louvain significantly outperformed other algorithms. G. Drakopoulos et al. [25] constructed two social network graphs on Twitter for controversial topics and ordinary topics respectively in Neo4j, and evaluated four community detection algorithms including Louvain, Edge Betweenness, Walktrap, and CNM, empirically finding that Louvain produced communities with higher cohesion and more closely connected members. Therefore, this study selected the Louvain method to detect communities in the constructed scholar cooperation network for modular clustering of scholars, thereby better analyzing the characteristics of scholar aggregation partitions and their trends of strengthening or dispersion.

Applying the Louvain algorithm in GDS discovered 100 scholar cooperation communities, with partial results shown in [Figure 6: see original paper] presented in descending order of community size. The largest community contains 26 scholars, among whom Deng Jun, Wang Ruan, Zhong Chuyi, Song Xianzhi, and Sun Shaodan have high co-authorship frequencies, conducting analysis and research on historical projects from a digital humanities perspective; He Chenzhi and Xu Xiaojuan conducted practical research on crowdsourcing projects for digital humanities in libraries; Li Daoxin analyzed application paths of digital humanities from the perspective of film art, among others. It is evident that in modularized communities, there are scholars with high co-authorship frequencies as well as scholars from different disciplines citing each other, with scholars in the same community having high degrees of association and shared research directions and hotspots, exhibiting high similarity. Meanwhile, [Figure 7: see original paper] the scholar cooperation relationship graph also clearly shows the distribution and closeness of scholars in different communities. Nodes of the same color represent that they are in the same community. While research themes are similar, different scholars cite each other, further strengthening the degree of association between scholars and providing a learning platform for knowledge exchange and sharing.

### Core Scholar Identification

Core scholars refer to those who have produced numerous research achievements, possess significant academic influence, and have contributed to the development of a particular research field—they are the backbone driving academic progress in that field [26]. Analysis of core scholars facilitates research by helping scholars comprehensively locate core scholar groups of interest and quickly access core scientific literature in the field, thereby rapidly understanding the current state and gaps in research to lay a solid foundation for in-depth study.

The Betweenness Centrality algorithm is a classic indicator for measuring network centrality. This study utilizes the Betweenness Centrality algorithm in GDS to measure the importance of different nodes in the scholar network, i.e., to detect the influence of a node on information flow in the graph. The algorithm calculates all unweighted shortest paths between all node pairs in a network, with each node receiving a score based on the number of shortest paths passing through it. Nodes that more frequently lie on shortest paths between other nodes receive higher scores.

In GDS, the Betweenness Centrality algorithm scored the shortest paths of 410 scholars, assigning each scholar a ranking number while sorting by score in descending order, with identification results shown in . Scholar Liu Wei received the highest score, followed by Zhao Yuxiang. Higher scores indicate these scholars are more active in digital humanities research and have made substantial contributions with significant academic influence in this field. Based on the identification results, a scatter plot was drawn as shown in [Figure 8: see original paper], where an obvious cliff-like decline appears after node 16, leading to the preliminary conclusion that the top 16 scholars can be identified as core scholars in digital humanities research. Among these core scholars, Liu Wei and Xia Cuijuan work at Shanghai Library, Zhu Xuefang and Ye Ying at Nanjing University, Zhao Yuxiang at Nanjing University of Science and Technology, and Wang Xiaoguang at Wuhan University, among others. To some extent, this reflects that these scholars' institutions are the main bases for their scientific research, and they represent core teams in the field, leading their students and collaborators to conduct in-depth digital humanities research with numerous achievements. Among them, Shanghai Library hosts national philosophy and social science fund projects on digital humanities. To more clearly reflect core scholars, the size of scholar nodes in the graph presented by Neo4j can be used to reflect their positions in digital humanities research, as shown in [Figure 9: see original paper], where larger nodes indicate greater academic influence. This provides important reference value for relevant researchers to locate core scholars in the field in a more convenient, fast, and clear manner.

### Cooperation Trend Prediction

In the era of big data, the trend toward collaborative academic research is increasingly evident. As an important component of scientific research activities, cooperation forms play an extremely important role in enhancing research efficiency and promoting research output. Research shows that over the past 20+ years, the number of collaborative studies in various disciplines has shown a significant growth trend, with scholars sharing the same research field and similar research directions being more likely to cooperate in the future [30]. However, due to temporal and spatial barriers, it is difficult for scholars to accurately identify researchers with similar research directions from the vast scholar community. Therefore, analyzing and mining potential cooperation partners can effectively improve research efficiency. This study uses link prediction algorithms in GDS to calculate proximity between nodes, thereby helping scholars find potential cooperation opportunities.

Link prediction algorithms predict the likelihood of edge formation between pairs of nodes not yet connected, based on known node feature information and network topology. Common link prediction algorithms include neighbor-based link prediction and common neighbor-based link prediction. Neighbor-based algorithms include total neighbors and preferential attachment, while common neighbor-based algorithms include common neighbors, resource allocation, and Adamic-Adar (AA) algorithm [31]. D. Liben-Nowell et al. [32] and T. Zhou et al. [33] compared multiple link prediction algorithms experimentally and found the AA algorithm to be relatively superior. The AA algorithm calculates the closeness between two nodes based on the adjacent node sets of common neighbors, with nonlinear normalization of set quantities. The formula for predicting scholar cooperation links in the network is as follows:

$$A(x,y) = \sum_{z \in N(x) \cap N(y)} \frac{1}{\log |N(z)|}$$

In this formula, when the calculated value is 0, it indicates the two nodes are not close; when the value is larger, it indicates the nodes are closer. In the above scholar cooperation community analysis, compared to different communities, scholars within the same community have tighter cooperation relationships, though even within the same community, their cooperation varies in density. This study selected the fourth largest scholar cooperation community, with core scholar "Liu Wei" as the research object, using the above formula and Cypher query language "MATCH (s1:author{Author: 'Liu Wei' }),MATCH (s2:author{Author:' *' }),RETURN gds.alpha.linkprediction.adamicAdar(s1, s2) AS score" to calculate and present the possible link degree between Liu Wei and other scholars in the same community. The prediction scores are shown in . Among them, Liu Wei and Zhao Yuxiang have the highest score for potential link formation, indicating the greatest possibility of cooperation, while Liu Wei and Wang Li have the lowest possibility of cooperation.

Meanwhile, the scholar cooperation relationship graph of Liu Wei' s community was presented using Cypher query statements, as shown in [Figure 10: see original paper]. This graph shows that scholars in the same community are closely associated, though there are also a few scholars who have not established direct cooperation relationships, such as Liu Wei with Cen Jiongliang, Zeng Hui, Liu Hong, and Wang Li, corresponding to their lower cooperation link prediction values. By analyzing and [Figure 10: see original paper], it is not difficult to find that among scholars who have already established direct connections, Song

Shijie has the lowest score, which can be determined as the minimum threshold for new link formation. That is, when the score between two scholars without direct connection exceeds this threshold, it indicates a higher likelihood of link formation and greater cooperation possibility. Thus, it can be seen that Liu Wei is more likely to have effective academic exchanges and cooperation with Cen Jiongliang, Zeng Hui, and Liu Hong, showing a clear cooperation trend.

In summary, Neo4j' s statement querying and algorithm analysis functions are effective tools for predicting scholar cooperation trends, saving time for scholars in finding cooperation partners and improving cooperation benefits. Against the backdrop of increasingly frequent scholarly exchanges, research cooperation has become a necessary form for scholars to promote academic research development. The more diverse and multidisciplinary the cooperation among scholars, the more active and efficient the academic exchange atmosphere in the field, with different thinking collisions driving the diversified and interdisciplinary development of digital humanities.

## Conclusion

With the deepening development of the digital era, "digital humanities" has important practical significance for implementing document rescue and protection, providing public cultural services, and promoting excellent traditional Chinese culture. In China, digital humanities as a professional academic research field has begun to accelerate its development, and research results born from this interdisciplinary research paradigm will be presented through more cooperation forms. For researchers, cooperation can generate new ideas and research directions, increasing collaborators' output and influence. For disciplinary development, cooperation can promote the formation of new knowledge systems, broaden scholars' knowledge horizons, and update their knowledge structures, thereby promoting the deep integrated development of digital humanities across multiple disciplines while helping scholars quickly and efficiently find interdisciplinary researchers highly associated with their research interests and directions.

This study utilizes Neo4j, a powerful tool for processing complex associative data, to store and analyze research subjects (scholars) and their relationships in China' s digital humanities field, employing the GDS algorithm library to realize scholar cooperation community discovery, core scholar identification, and cooperation trend prediction. Although social network analysis methods are common in analyzing various association network structures from perspectives such as centrality, cohesive subgroups, and core-periphery, this study not only achieves functions attainable by traditional social network analysis methods but also implements data storage, real-time updates, instant querying, and predictive reasoning functions, representing a powerful supplement to social network analysis and providing new ideas and methods.

However, this study has several limitations: (1) When acquiring relevant literature, some research results whose titles did not include "digital humanities" or "humanities computing" but whose content focused on digital humanities were overlooked, resulting in relatively small amounts of scholar node and relationship data and slight deficiencies in completeness. (2) The larger the data volume and the higher the complexity, the more obvious the advantages of Neo4j in data processing. However, this study' s application of Neo4j' s functions is relatively simple, not fully leveraging its data analysis advantages. Therefore, in future research, the authors will continue to deeply study Neo4j' s powerful data analysis functions, continuously expand and update scholar data volumes, and fully demonstrate the complex cooperation relationships among scholars to provide references for potential scientific research cooperation.

## References


[1] Ke Ping, Gong Ping. Analysis of the Evolution Path and Hot Fields of Digital Humanities Research [J]. Journal of Library Science in China, 2016, 42(6): 13-30.

[2] Pan Liangen. Rational Thinking on the Application of Digital Humanities in the Archival Field [J]. Archives and Construction, 2020(7): 6-10.

[3] Niu Li, Gao Chenxiang, Zhang Yufeng, et al. Discovery, Reconstruction, and Storytelling: Paths and Methods of Archival Research from a Digital Humanities Perspective [J]. Journal of Library Science in China, 2021, 47(1): 88-107.

[4] Liu Bei, Yuan Yi, BOUTIN E. Research on the Application of Social Network Analysis in Paper Cooperation Networks [J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(3): 407-417.

[5] Liu Zhihui, Zhang Zhiqiang. Author Keyword Coupling Analysis Method and Empirical Research [J]. Journal of the China Society for Scientific and Technical Information, 2010, 29(2): 268-275.

[6] Qiu Junping, Liu Guohui. Author Cooperation Research Based on Social Network and Keyword Analysis—Taking the Domestic Knowledge Management Field as an Example [J]. Information Science, 2014, 32(6): 3-7, 13.

[7] Xu Chenfei, Zhao Wenjuan. Analysis of Author Co-authorship Networks in China' s Digital Humanities Research Field [J]. Library Tribune, 2019, 39(11): 14-24.

[8] Gong Xue, Yang Ying. Visualization Analysis of Research Hotspots and Co-authorship Networks of Digital Humanities in China [J]. Journal of Library and Information Science, 2019, 4(6): 39-45.

[9] Graph Database GDB-Help Center-Alibaba Cloud [EB/OL].[2021-10-10]. https://help.aliyun.com/product/102714.html.

[10] HugeGraph [EB/OL].[2021-10-10]. https://hugegraph.github.io/hugegraph-doc/.

[11] Native Graph Database | Neo4j Graph Database Platform [EB/OL].[2021-10-10]. https://neo4j.com/product/neo4j-graph-database/.

[12] Graph Database | Graph Analytics Platform | TigerGraph [EB/OL].[2021-10-10]. https://www.tigergraph.com/.

[13] Liu Chunjiang, Li Shuying, Hu Hanlin, Fang Shu. Research and Application Progress of Graph Databases in Complex Network Analysis [J]. Data Analysis and Knowledge Discovery, 2022, 6(7): 1-11.

[14] Guo Kunming. Research on Relationship Reasoning and Prediction Methods Based on Heterogeneous Networks [D]. Taiyuan: Taiyuan University of Technology, 2017.

[15] KOLOMEETS M, CHECHULIN A, KOTENKO I V. Social Networks Analysis by Graph Algorithms on the Example of the VKontakte Social Network[J]. Journal of Wireless Mobile Networks, 2019, 10(2): 55-75.

[16] Ding Hongli. Personnel Relationship Mining Based on Neo4j Graph Database [J]. Telecommunication Engineering, 2020, 60(7): 771-777.

[17] CHU Z, YU J, HAMDULLA A. A Novel Deep Learning Method for Query Task Execution Time Prediction in Graph Database[J]. Future Generation Computer Systems, 2020, 112: 534-548.

[18] Li Jinyang. Research on the Application of Graph Databases in Libraries [J]. Library, 2020(11): 109-115.

[19] FLEMING J, LEVY S, NAG P, et al. Graph Database System and Method for Facilitating Financial and Corporate Relationship Analysis: US, US8674993 B1[P]. 2014.

[20] Zhang Weichong, Wang Fang, Huang Yi. Research on Knowledge Modeling of Guizhou Province' s Big Data Policy Based on Graph Database [J]. Digital Library Forum, 2020(4): 30-38.

[21] Yu Juan, Huang Hengqi, Xi Yunjiang, et al. Research on Knowledge Graph Reasoning Methods for Character Relationships Based on Graph Database [J]. Information Science, 2019, 37(10): 40-47.

[22] DRAKOPOULOS G, KANAVOS A, MYLONAS P, et al. Defining and Evaluating Twitter Influence Metrics: A Higher-Order Approach in Neo4j[J]. Social Network Analysis & Mining, 2017, 7(1): 52.

[23] GIRVAN M, NEWMAN M E J. Community Structure in Social and Biological Networks[J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7821-7826.

[24] Guo Li, Wang Jiaqi, Zhang Hengxu, et al. Louvain Overlapping Community Detection Algorithm [J]. Journal of Shihezi University (Natural Science Edition), 2020, 38(3): 384-389.

[25] DRAKOPOULOS G, GOURGARIS P, KANAVOS A. Graph Communities in Neo4j[J]. Evolving Systems, 2020, 11(3): 397-407.

[26] Wang Yuefen, Yang Xue, Yu Houqiang, et al. Cooperation Patterns and Comparative Studies of Artificial Intelligence Research Teams [J]. Library and Information Service, 2020, 64(20): 26-36.

[27] Xia Cuijuan, Liu Wei, Chen Tao, et al. Development and Practice of Genealogy Linked Data Service Platform [J]. Journal of Library Science in China, 2016(3): 27-38.

[28] Xia Cuijuan, Zhang Lei, He Chenzhi. Library Digital Humanities Project Construction for Knowledge Services: Methods, Processes, and Technologies [J]. Library Tribune, 2018(1): 1-9.

[29] Xia Cuijuan, Lin Haiqing, Liu Wei. Research and Design of Chinese Ancient Books Data Model for Evidence-Based Practice [J]. Journal of Library Science in China, 2017(6): 16-34.

[30] GLANZEL W, SCHUBERT A. Analysing Scientific Networks Through Co-authorship[M]//Handbook of Quantitative Science and Technology Research. Dordrecht: Springer, 2004: 257-276.

[31] LU L, ZHOU T. Link Prediction in Complex Networks: A Survey[J]. Physica A: Statistical Mechanics and its Applications, 2011, 390(6): 1150-1170.

[32] LIBEN-NOWELL D, KLEINBERG J. The Link-Prediction Problem for Social Networks[J]. Journal of the American Society for Information Science and Technology, 2007, 58(7): 1019-1031.

[33] ZHOU T, LU L, ZHANG Y C. Predicting Missing Links via Local Information[J]. The European Physical Journal B, 2009, 71(4): 623-630.

## Author Contributions

Xiong Huixiang: Conceptualized the overall research framework and provided guidance on the paper; Huang Xiaojie: Collected and processed data and drafted the paper; Chen Ziwei: Processed data and revised the paper; Li Xinran: Revised the paper and finalized the version.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv —Machine translation. Verify with original.*