Journal of Computers Vol. 35 No. 4, August 2024, pp. 159-174
doi: 10.53106/199115992024083504012

# Research on Key Technologies for Constructing Knowledge Graphs of Spatialization Experts

HengYao Jin[1], Chuan Yin[1*], MingYi Du[1], QiXuan Gao[1],
Xu Yan[2], and Yang Zhang[3]

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
jinhengyao2022@163.com, {yinchuan, dumingyi}@bucea.edu.cn,
gaoqixuan001@126.com

[2] Hezhou City Babu District Bureau of Natural Resources,
Hezhou 542899, China
13557144334@163.com

[3] School of Electronic and Information Engineering, Beijing Jiaotong University,
Beijing 100044, China
zhang.yang@btju.edu.cn

*Received 19 July 2024; Revised 28 July 2024; Accepted 15 August 2024*

**Abstract.** The construction of knowledge graphs for spatialization experts represents a significant research domain within the field of spatial knowledge service platforms. This study aims to resolve the prevalent issues of low entity disambigation accuracy and the inadequate representation of spatial knowledge in conventional expert knowledge graphs. Focusing on the discipline of surveying and mapping, the paper introduces a novel approach to entity disambiguation that synergizes methods for unknown institutional entity resolution with community detection predicated on co-authorship networks. Additionally, by integrating principles pertinent to spatial knowledge organization, the paper enhances the traditional expert knowledge graph with a spatial dimension and formulates a strategic framework for the visualization of knowledge maps. The proposed techniques for entity disambiguation and expert knowledge graph visualization furnish viable references for the development and deployment of knowledge graphs within the survey.

**Keywords:** expert knowledge graph, entity disambiguation, map visualization representation

## 1 Introduction

In recent years, with the rapid development of big data, artificial intelligence 2.0 and other technologies, it is imperative to move from traditional geographic data information services to spatialized knowledge services, in which the automatic construction technology of domain knowledge graph is one of the key technologies to realize knowledge services [1]. Expert knowledge graph is an important part of domain knowledge graph, which has important application value to improve the multidimensional semantic association of knowledge graph and enrich the application of knowledge service [2]. Expert knowledge graph is a knowledge graph formed by a triad consisting of experts as entities in the graph, co-authorships of currently published papers as relationships between entities, and research institutes and geographic locations of the research institutes as attributes of the entities [3]. At present, most of the knowledge graph construction methods are based on the network data of the construction of the relationship network from different perspectives. For example, Baozhen Li et al. constructed the knowledge graph of related fields based on specialized multi-source network data [4]. Dianchao Sun et al. combed the role of policy relationship network by studying the impact of innovative talent policy relationship network on talent introduction [5]. Massive scientific and technological literature resources contain rich scientific research entities and their relationships, and these massive resources can provide more information basis for the discrimination of scientific research entities [6]. On the basis of rich data, expert relationships can be cross-validated so that the accuracy of obtaining expert relationships can be improved [7]. In previous studies, most of the

* Corresponding Author

relationship networks are constructed using author names + sequence numbers, which is convenient for obtaining data and constructing the network quickly, but the shortcomings are that the experts are not recognized, which leads to too many nodes with the same name, affecting the accuracy of the expert relationship [8].

At present, as a common method of constructing expert knowledge graph based on literature author co-authorship, disambiguation brings great challenges due to the limited amount of information carried in the literature. The entity disambiguation of literature-based expert co-authorship knowledge graph mainly includes two key techniques: expert entity relationship extraction and homonym disambiguation. Expert entity relationship extraction is currently mainly based on expert co-authorship information, such as the GHOST framework are used only co-authorship information to construct expert entity relationships. The algorithm of this expert relationship extraction method is simple and easy to implement, but this method also has certain defects. Since the number of author nodes in the constructed knowledge graph is the same as the number of their published papers, when the authors' names are in common use, a large number of nodes with the same name will exist in the knowledge graph, which will bring a large amount of workload to the subsequent homonym disambiguation process. However, the institutional information in the literature can distinguish experts with the same name from different organizations, and if the institutional information can be integrated into the relationship extraction process, the number of nodes with the same name in the knowledge graph can be greatly reduced to reduce the workload of homonym disambiguation. For example, the CoAAND model [9] constructs an entity-relationship graph based on the cooperative relationship between authors and the affiliation relationship between authors and organizations. Constructing relationships through expert-organization relationship can reduce manual annotation workload and improve the display intensity of expert relationships, which in turn improves the extraction effect [10]. Overall, only a few studies have taken this kind of approach, mainly because at present, in addition to WOS and other databases that contain the correspondence between authors and organizations in their metadata, the metadata of most other public databases (e.g., Zhi.com, Wanfang) only save authors and organizations as two independent fields with no correspondence, so that the researcher needs to manually carry out the work with the help of cooperative relationship annotation information in the original text of the literature. Researchers need to manually extract experts' affiliations with the help of partnership labeling information in the original literature, and it takes a large amount of human resources and time cost to construct large-scale expert mapping, so there is an urgent need for a method that can automatically extract experts' affiliations with the help of the existing metadata information at the stage of experts' relationship extraction.

In the research of homonymous disambiguation of expert knowledge graphs based on co-authorship, the prevailing idea is to solve the disambiguation problem by transforming it into a clustering problem based on the similarity between authors with the same name [11-13]. The similarity calculation methods can be roughly divided into two categories, one is mainly based on the author's research field, affiliation, title and other attributes to calculate the similarity, the accuracy is relatively high, but the required attribute information is more, and the universality is low; the other method is mainly based on the number of effective paths between the authors with the same name and the length of the information to calculate the similarity, such as the classic CoAAND model, the method requires fewer attribute information and is easy to realize, but the accuracy is relatively low. realize, but the accuracy is relatively low [10]. With the development of network construction technology, based on the construction of complex networks and on this basis to analyze the community structure, not only makes the researcher more in-depth understanding of the characteristics of the nodes, and at the same time can further discover the evolution of the network law [14], so more and more research on the community discovery technology applied to the field of homonymous disambiguation. Community discovery is to cluster the network into a collection of nodes with close internal connections and sparse connections to the outside, and the community discovery algorithm focuses on the rapid discovery of the closely connected subgraphs in the network, generally does not take into account the attribute information of the nodes, so the computational efficiency is generally higher than the traditional clustering algorithms, and it is suitable for homonym disambiguation of large-scale expert knowledge graphs. Commonly used algorithms include k-clique percolation algorithm [15], Louvain algorithm [16], LPA algorithm [17], etc. Among them, k-clique percolation algorithm is mainly applied to delineate the complete subgraphs of a complete community that exists with a certain feature, so as to carry out the problems of evolution, deduction, etc. For the relational network constructed with co-authorship, the process of co-authorship is affected by the implicit information of expert entities, and the k-clique percolation algorithm can numericalize the features. Extracting expert entities based on literature data has the problem of name ambiguity, and the core problem is how to divide literature data of the same author in a collection. And Louvain's algorithm is a method to describe the community by modularization, which has high delineation effect on the network with high interconnection within the community and high sparsity between the communities. The expert relationship network constructed by co-authorship relationship satisfies this condition, and the modular division can effectively improve the disambiguation accuracy. In terms of spatialization of knowledge maps, scholars have already conducted research, and Junnan Liu et al. proposed a knowledge map construction process based on spatial relations, focusing on the extraction and fusion of spatial relations [18]. Fuqiang Wang proposed that the semantic character of spatial knowledge map illustrates that spatial knowledge map is more in line with people's cognitive habits of the geographic world compared with traditional geographic entities.

To summarize, expert knowledge graph is an important part of domain knowledge service platform, and homonym disambiguation is still a difficult problem that needs to be solved in the construction of expert knowledge graph, and at present, there are mainly problems such as low efficiency of knowledge graph construction, and poor accuracy of cross-community entity disambiguation. To address the above problems, this paper proposes entity disambiguation based on the co-authorship relationship of unknown institutions and the homonymous merger disambiguation method based on community discovery algorithm for entity disambiguation, and on the basis of which the spatial visualization expression of expert knowledge graph is realized, i.e., through the introduction of geo-coding technology, the location information of expert institutions is standardized and structured, and the abstract expert relationship network is mapped to the specific geographic space, which forms an intuitive and easy-to-understand expert knowledge graph. Forming an intuitive and easy-to-understand expert knowledge map. This spatial visualization enhances the intuitiveness and readability of the map, and at the same time provides researchers with a new perspective to examine and analyze the distribution, flow and cooperation mode of expert resources, which provides powerful decision-making support for scientific research cooperation, introduction of talents, and policy formulation.

## 2 Technical Route of Constructing Spatial Expert Knowledge Graph

This paper's technical roadmap encompasses three core components: the construction of a knowledge graph from bibliographic metadata, the diminution of noisy data through entity disambiguation techniques, and the employment of spatial visualization techniques to represent and analyze the knowledge graph spatially. Fig. 1 graphically depicts the technical roadmap.

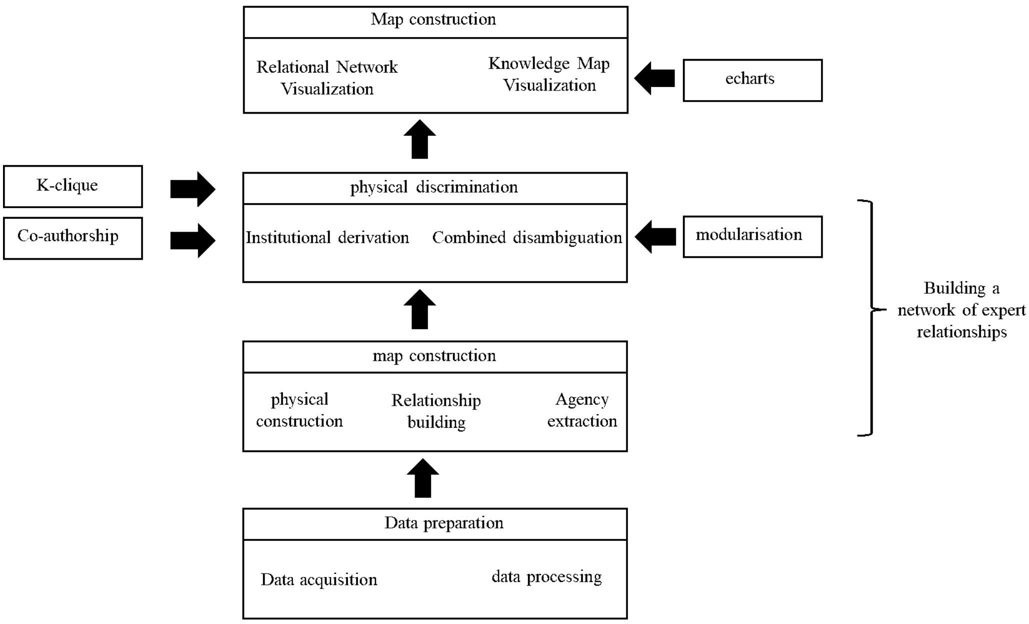



**Fig. 1.** Structure of the paper

The expert data is extracted with literature data, and the expert co-authorship relationship graph model constructed based on the expert data provides a data source for the construction of the knowledge graph and guarantees the smooth progress of the experiment. The preprocessing technology is used to deal with the problems of incomplete, noisy, and inconsistent data crawled from the network. The main tasks of data preprocessing are data regularization, deleting erroneous data, filling in missing data, and integrating duplicate data. Expert relationship construction is generally based on expert co-authorship of literature to establish contact, but due to the large number of scientific research institutions in China, the large base of literature authors, the promotion and reassignment of researchers, the phenomenon of author homonymy is unavoidable, and the same author will also appear between multiple institutions, in order to ensure the accuracy of the study, it is necessary to carry out the entity of disambiguation between the experts. The expert knowledge map is expressed in the form of abstract data in the form of relational model, and it is difficult to express the implicit information. Spatial visualization technology maps the expert relationship model to a spatial database, which forms spatial entities and visualizes the implicit information.

## 3 Construction and Design of Spatial Expert Knowledge Graph

### 3.1 Design of Expert Knowledge Graph Model

The expert co-authorship graph model based on journals can generally be represented as $G = \{V, E, W\}$, where $V$ represents the set of expert nodes, $E$ represents the set of edges representing co-authorship relationships among experts, and $W$ represents the edge weight representing the number of co-authored articles between experts. In previous studies, most of the authors' names + sequence number (e.g., "Zhangsan1", "Zhangsan2", "Li41", "Li42") are used as the experts' names.") as the unique identification of experts to build the co-authorship network, the method is simple and easy to realize, but the shortcoming is that the number of nodes with the same name in the network is too many, and the workload of homonym elimination is large at a later stage [10]. The author's affiliation is an attribute information with important disambiguation value, so this paper adopts the author's name + organization (Zhang San {Beijing University of Architecture}) as the expert's unique identifier, which can greatly reduce the number of nodes with the same name in the network, and the effect of clustering and community discovery will be improved, but if there is a change of the author's work or study, the same author will still have different unique identifiers of the same node. However, if the authors have job or study changes, the same author will still have different uniquely identified nodes with the same name, which still need to be merged and disambiguated.

The graph model of expert co-authorship relationship defined in this paper is represented as $G = \{V, E, W\}$, $V$ denotes the set of expert nodes, $E$ denotes the set of edges of expert co-authorship relationship, $W$ denotes the edge weights of expert co-authorship articles, and $V_i$ denotes the attribute information of the expert including: *Number* represents the expert's identifier. *Name* represents the expert's name, *Institution* represents the expert's affiliated institution, *Location* represent the coordinates of the institution, Using a graph model as in Equation (1).

$$\begin{aligned} &G = \{V, E, W\} \\ &V = \{v_1, v_2, v_3, \ldots\} \\ &E = \{e_{1i}, e_{2i}, e_{3i}, \ldots\} \\ &W = \{w_{1i}, w_{2i}, w_{3i}, \ldots\} \\ &V_i = \{Number, Label, Location, \ldots\} \\ &Label_i = \{Name(Institution_i)\} \end{aligned} \quad . \tag{1}$$

### 3.2 Expert Knowledge Graph Entity Disambiguation

The problem of unknown affiliations and multiple affiliations of authors can be categorized into four types of relationships between authors and institutions: 1:1, 1: n, n:1, and n:n. 1:1 Relationship: Represents a scenario where one author is affiliated with only one institution, and this institution exclusively has this author. This is the simplest case. 1:n Relationship: Indicates a situation where one author is affiliated with multiple institutions, but each institution has only this author. This reflects collaboration across multiple institutions by the author. n :1 Relationship: Signifies multiple authors collectively affiliated with the same institution, and this institution has each of these authors. This scenario is common in team collaboration research. n:n Relationship: The most complex scenario where multiple authors are affiliated with multiple institutions, and each institution has multiple authors. This relationship reflects intricate collaboration networks and organizational structures. Handling these different types of relationships requires careful consideration of entity resolution and disambiguation strategies to accurately reflect the affiliations between authors and institutions, ensuring the accuracy and completeness of the constructed knowledge graph.

Clearly, in the fourth type of relationship, it's challenging to establish a direct correspondence between authors and institutions, leading to author affiliations being labeled as unknown. Furthermore, when authors undergo job changes or career advancements, they may be associated with multiple institutions simultaneously. Therefore, this paper proposes an entity disambiguation method based on expert relationships to modify and merge problematic expert entities, thereby reducing the information error of expert entities. The entity disambiguation process is illustrated in Fig. 2.

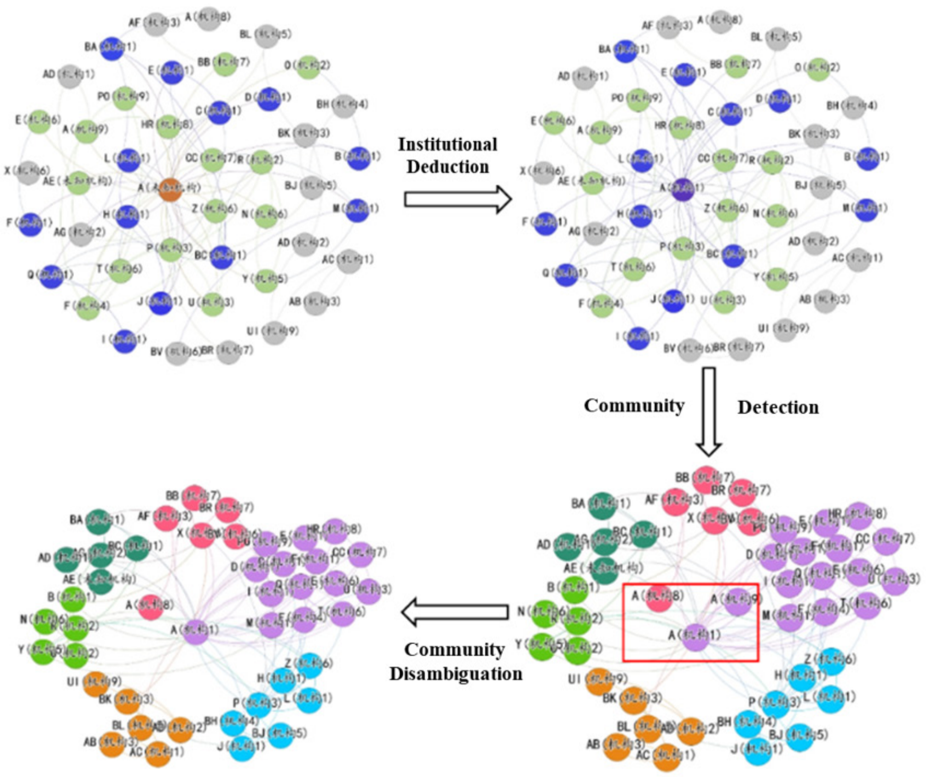



**Fig. 2.** Diagram of entity relationships

The process of entity disambiguation in the expert knowledge graph is as follows:

(1) For expert entities with unknown affiliations, the neighbor nodes' affiliation recurrence frequency is calculated. The "k-clique" percolation algorithm is then applied to the expert knowledge graph to partition communities based on the neighbor nodes of expert entities with unknown affiliations. The primary focus is on entity disambiguation based on co-authorship relationships with unknown affiliations. High recurrence frequency institutions are selected as results, supplemented by the "k-clique" percolation algorithm to determine cases where affiliation cannot be uniquely identified to complete the disambiguation.

(2) Apply the "Louvain" community detection algorithm to the expert knowledge graph, setting the resolution of the modularity. Identify experts within the same community, search for experts with the same name within the community, merge and disambiguate the experts with the same name. Compare the merged entities with manually labeled data to validate the accuracy of disambiguation and merging. Determine if the parameters are set optimally.

$$Q = \frac{1}{2m}\sum_{i,j}[sim_{i,j} - \frac{k_i k_j}{2m}]\delta(c_i, c_j)\,. \quad \textbf{(2)}$$

Where denotes the similarity between $i$ and $j$ , denotes the sum of the weights of the edges connected to vertex $i$ , is the community assigned to vertex $i$ , and the $\delta$ function $\delta$ $(u, v)$ is 1 if $u = v$ otherwise 0.

**Entity Disambiguation of Unknown Institutions Based on Co-authorship Relationships.** The relationship between authors and institutions can be broadly categorized into four types: 1:1, 1:n, n:1, and n:n. The first three types of relationships allow for easy identification of authors. However, in the fourth type of relationship, where authors are associated with multiple institutions, the relationship between authors and institutions becomes ambiguous, resulting in author affiliations being labeled as unknown. In this paper, we address the aforementioned issue using institution inference. Firstly, we utilize the neighbors of nodes in the first-order co-authorship relationship to count the frequency of institution occurrences among these neighbors. We then select the institution with the highest occurrence frequency to disambiguate the unknown institution entities. The formula is as follows:

$$C(A_i) = \sum\nolimits_{v \in N(u)} \delta(A(v), A_i)\,. \quad \textbf{(3)}$$

where $N(u)$ denotes the set of first-order neighbouring nodes of author $u$, $A(v)$ denotes the institutions to which author $v$ belongs, $\delta(A(v), A_i)$ is the indicator function, and if $A(v) = A_i$, then $\delta(A(v), A_i) = 1$; otherwise $\delta(A(v), A_i) = 0$.

To address the issue of non-uniqueness of the institution with the highest occurrence frequency, we employ the "k-clique" percolation algorithm to partition the neighboring nodes into communities. We assess the closeness between neighboring nodes and the node to be disambiguated. The technical process is illustrated in Fig. 3.

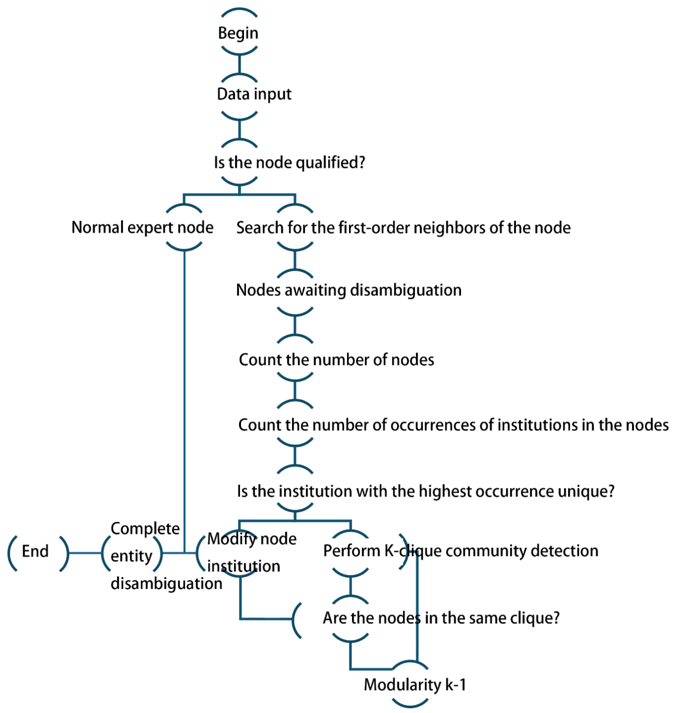



**Fig. 3.** Organization deduction flow chart

**Entity Disambiguation of Homonymous Entities Based on Community Detection Algorithm.** Extracting expert entities from literature data will lead to the problem of multiple affiliations of the same author, and the key to solving the problem lies in how to divide all the organizations of the same author into the same set. If the same author belongs to different organizations, the co-authored literature will often be divided into the same co-authorship circle due to the closer co-authorship relationship, so the key to solving the problem is how to divide the appropriate size of small co-authorship circle in the large-scale co-authorship network. Compared with traditional clustering disambiguation algorithms, community discovery algorithms are more suitable for entity disambiguation in large-scale knowledge graphs. When community discovery is applied to disambiguate entities in expert knowledge graphs, if authors with the same name correspond to different entities, they have different egocentric networks, and thus are likely to be classified into different communities; if authors with the same name correspond to the same entities, they have similar egocentric networks, and thus are likely to be classified into the same communities. community discovery algorithm based on modularization idea, suitable for expert relationship networks such as high interconnection within the community and high sparsity between the communities [19]. Louvain algorithm uses the modularity as a parameter to describe the advantages and disadvantages of the results of the network community division, and the formula is defined as follows:

$$Q = \frac{1}{2m}\sum_{i,j}[A_{i,j} - \frac{k_i k_j}{2m}]\sigma(c_i, c_j)\,. \tag{4}$$

In the formula, $A$ represents the edge weight between nodes, $A_{i,j}$ is the weight of the edge between nodes $i$ and $j$. $m$ represents the total number of edges in the network, $k_i$ and $k_j$ represent the sum of edge weights between any node and node $i$ and $j$. respectively. So $\frac{k_i k_j}{2m}$ represents the average edge weight. Specifically, when weights are introduced in the network, $A_{i,j}$ becomes 1, In this case, the network is considered as an unweighted network; $c$ represents the community, and $c_i$ and $c_j$ represent the community of any nodes $i$ and $j$ respectively. $\sigma$ denotes the same community, When nodes $i$ and $j$ are in the same community, $\sigma$ is 1, otherwise, it is 0. Therefore, the value of modularity $Q$ ranges from [0,1], the closer the value is to 1, the better the effect of community division, and the value of modularity ranges from 0.3 to 0.7 in the real network.

Louvain community detection algorithm mainly determines which community the current node belongs to by traversing each node in the network and calculating the modularity gain of each node from the community it is currently in and merging it into the community of the neighbouring node. The modularity gain is calculated by taking the current node out of its current community and merging it into another neighbouring community, the change in community modularity of that neighbouring community before and after joining this node. The formula for its modularity gain is shown in Equation (5):

$$\Delta Q = [\frac{\sum in + 2k_{i,in}}{2m} - (\frac{\sum tot + k_i}{2m})^2] - [\frac{\sum in}{2m} - (\frac{\sum tot}{2m})^2 - (\frac{k_i}{2m})^2]\,. \tag{5}$$

On the basis of formula (5), the module degree gain formula can be further simplified, and the simplified formula is shown in formula (6):

$$\Delta Q = \frac{2K_{i,in}}{2m} - \frac{2\sum tot \times K_i}{(2m)^2}\,. \tag{6}$$

In the formula, $K_{i,in}$ is the sum of the weights of the edges connecting node $i$ to the nodes in community $C$, $\sum tot$ is the sum of the weights of the edges connecting all the nodes in community $C$ , $K_i$ is the sum of the weights of all the edges connecting node $i$ , and in an undirected unweighted graph, $K_i$ is the degree of node $i$ , and $m$ is the sum of the weights of all the edges in the network.

The community discovery clustering by setting a reasonable modularity threshold can realize the homonymous elimination of most of the experts, but for the small number of well-known experts, the method has certain defects, such experts, because of the research direction and the number of published papers, may have co-authorship relationships with experts in different fields, so the community discovery algorithm will divide these nodes into different community species. There may be co-authorship with experts in different fields, and the nodes with the same name will be dispersed to different academic circles, so the community discovery algorithm will divide these nodes into different community species.

In this paper, through the introduction of Louvain's algorithm, modularized computation is performed on the expert knowledge graph established by co-authorship, the authors of the same name are classified as within the same community, and compared with the manually labeled dataset in order to obtain the iterative termination conditions, and to seek for the optimal parameter settings, and the community discovery disambiguation process is shown in Fig. 4.

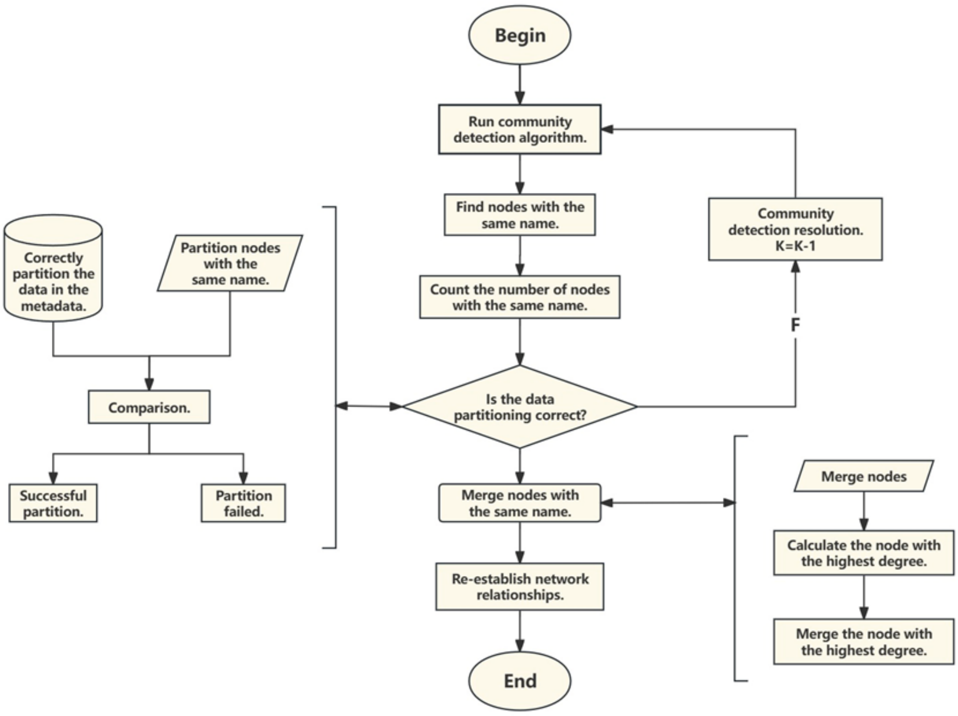



**Fig. 4.** Flow chart of merging and disambiguation of the same name

### 3.3 Spatialization and Analysis of Expert Knowledge Graphs

Spatial Expression and Analysis of Expert Knowledge Graphs mainly includes the following two research components.

(1) Establishing corresponding spatial entities through the expert co-authorship relationship map model, storing them in the spatial database, and mapping the entities in the database to the map for spatialized expression;

(2) Calculate the basic indexes of the expert knowledge map using the statistical algorithm of the knowledge map, and use the spatial analysis technique based on the superposition analysis to carry out geographic analysis such as spatial location statistics and geographic location normalization of expert entities on the spatially-expressed expert knowledge map.

The process of spatialization and analysis of the expert knowledge graph is illustrated in Fig. 5.

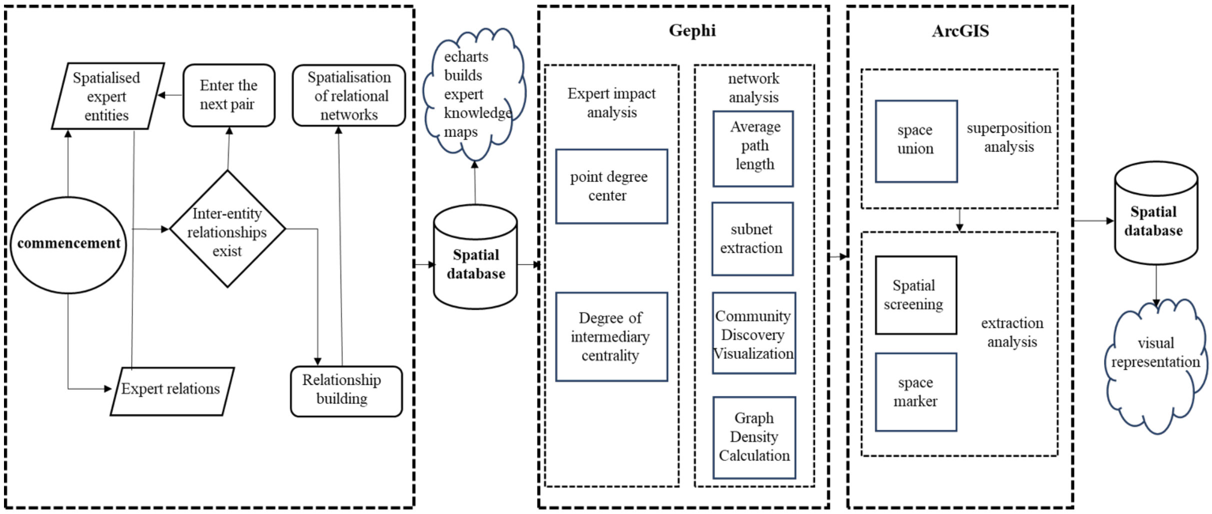



**Fig. 5.** Flow chart of spatial expression and analysis of expert knowledge map

**Analysis of Expert Knowledge Graphs.** Analysis of basic indices of expert knowledge graphs can be obtained through statistical analysis of information retrieved from the spatial database. In this study, the following indicators are mainly followed during data analysis:

(1) Degree centrality

The degree of a node in a network refers to the number of nodes directly connected to that node. In the expert relationship network studied in this paper, expert entities are network nodes, and their degree is the number of nodes with which the expert has collaborative relationships. The degree of expert node $i$ can be calculated as

$$\mathrm{k}_i = \sum_j a_{ij} \; . \tag{7}$$

$$a_{ij} \begin{cases} 1, & \text{Experts have collaborative relationships} \\ 0, & \text{Experts do not have collaborative relationships} \end{cases} . \tag{8}$$

The size of the degree of an expert node reflects the influence and importance of that expert in the network.

(2) average path length

The average path length refers to the average number of edges that must be traversed to go from one node to another in a network. The symbol $d_{ij}$ represents the distance between node $i$ and node $j$, which is expressed as the number of edges traversed when finding the shortest path between the two nodes. The average path length $L$ of the entire network is the average of the shortest path lengths between all pairs of nodes in the network.

$$L = \frac{\sum_{ij \in V, i \neq j} d_{ij}}{\frac{1}{2} N(N-1)} \; . \tag{9}$$

Here, $N$ represents the number of nodes. The value of $L$ determines the efficiency of information flow among experts and the connectivity of the network.

(3) Average clustering coefficient

The clustering coefficient is a measure of the degree to which nodes tend to cluster together, and it is defined as the proportion of a node's neighbors that are also neighbors of each other. The clustering coefficient $C_i$ is defined as the ratio of the actual number of edges between the neighboring nodes of node $i$ to the possible number of edges between them.

$$C_i = \frac{2E_i}{k_i(k_i - 1)} . \quad \textbf{(10)}$$

Where $E_i$ is the actual number of edges among the neighboring nodes of node $i$, and $k_i$ is the degree of node $i$, which is also the number of its neighboring nodes.

Due to the influence of paths between nodes on the clustering coefficient of node $i$, the average clustering coefficient is commonly used to evaluate the degree of network clustering, which is calculated as:

$$\overline{C} = \frac{1}{N}\sum_{i=1}^{N} C_i . \quad \textbf{(11)}$$

Where N is the number of nodes in the network.

The spatial distribution and implicit relationships of expert knowledge graph nodes can be visualized through spatialization. After spatialization, geographic analysis can be conducted using overlay analysis techniques. Geographic analysis involves analyzing the spatial distribution of experts in the knowledge graph and examining clustering distribution patterns. The implementation process includes:

(1) Utilizing the spatial data overlay capability, spatial joints of expert entities can be performed, followed by statistical analysis.

(2) Normalize the coordinates of expert entities to facilitate their aggregation within each province. Additionally, merge cross-province collaboration relationships with weighted aggregation to reduce the density of expert relationships across provinces.

(3) Based on the weighted aggregation of cross-province collaboration relationships, perform hierarchical representation and filtering. Use graded colors to visually display the density of collaboration relationships.

(4) Perform spatial filtering and extraction of expert entities with significant spatial distribution characteristics. Apply spatial identification to the filtered data and conduct regional statistical analysis.

**Spatial Representation of Organizational Structure in Expert Knowledge Graphs.** The spatial expression of expert knowledge graphs is constructed based on the affiliation institutions of experts. Spatial entities are established according to the geographical locations of expert affiliation institutions, with unique identifiers assigned to each expert entity. Connections are made between pairs of expert entities to establish spatial entities representing expert relationships. Spatial expression enables the reflection of regional distribution patterns of experts in the field and characteristics of inter-regional collaboration. It intuitively demonstrates the impact of spatial distribution on expert collaboration networks and co-authoring circles. The principle of implementation is as follows:

(1) The geographic location of the institution to which the expert entity belongs is obtained through the geocoding service, and the expert entity is mapped to the spatial database after one-to-one correspondence to form a spatial point entity with the following formula:

$$Geocode(Institution(V_i)) = Location(V_i) = (lat_{V_i}, lon_{V_i}) . \quad \textbf{(12)}$$

where *Geocode* is the geocoding function, $Institution(V_i)$ is the institution to which the expert $V_i$ is affiliated, and $lat_{V_i}$, $lon_{V_i}$ are the longitude and latitude of the institution to which the expert is affiliated, respectively.

(2) Will be mapped by the expert relationship to the edge table, the original information of the number of co-operation to generate the edge weight information to indicate the degree of relationship closeness; search the spatial database composed of expert relationship entities, through the loop statement will be compared with the expert relationship for the expert relationship mapped to the spatial line entity, the formula involved is as follows:

$$w_{ij} = f(C_{ij}) . \tag{13}$$

$$Edge(v_i, v_j) = (Location(v_i), Location(v_j), w_{ij}) . \tag{14}$$

where $C_{ij}$ represents the number of articles co-authored between experts $v_i$ and $v_j$, and $w_{ij}$ is the weight of side $e_{ij}$, which represents the closeness of the relationship between experts $v_i$ and $v_j$.

(3) Construct a relationship network using the completed spatialized expert entities and expert relationships in the spatial database according to the graph model described in section 2.1. Utilize third-party visualization tools in Python to invoke the Baidu API for adding geographic base maps, thus achieving spatial expression of the expert knowledge graph.

## 4 Examples and Results Analysis

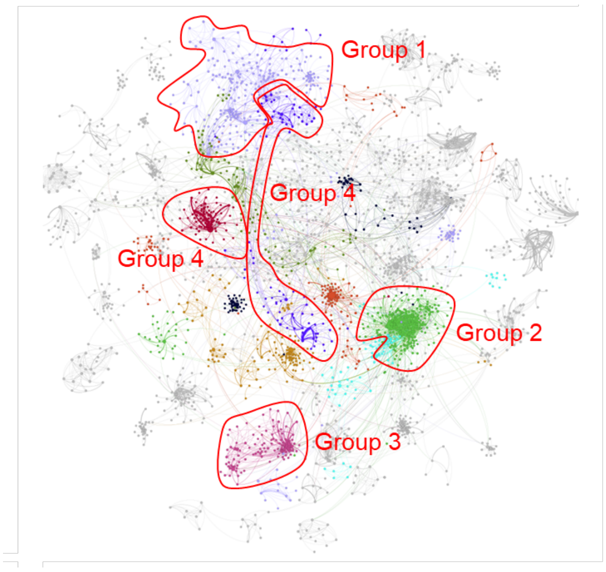
### 4.1 Data Source

The dataset for this research was obtained from the Geographic Information Professional Knowledge Service Platform (http://kmap.ckcest.cn). As of now, the platform aggregates more than 10 million records of bibliographic metadata, encompassing a range from 1991 to 2018, across disciplines including surveying and mapping, geographic information, and related fields. The analysis concentrated on extracting approximately 110,000 metadata entries from 2019 to 2020, specifically targeting research areas such as "geographic information modeling" and "mapmaking," to construct an expert knowledge graph pertinent to the field of surveying and mapping.

### 4.2 Results Analysis

**Expert networking.** The constructed expert knowledge graph is represented using gephi visualisation as shown in Fig. 6 after entity disambiguation using this paper. The expert knowledge graph contains 2970 nodes and 7048 co-operative edges, indicating that this paper presents a relatively large group of experts and that there are extensive co-operative relationships between experts in this expert network. The average degree is 4.746, which represents that each expert has collaborated with 4.746 other experts on average, indicating that the connections between experts are frequent but not very intensive. The average path length is 6.407, indicating that any two experts can know each other after about six people, which is in line with the theory of 'six degrees of separation' and reveals the multi-step nature of information transfer. The average clustering coefficient is 0.721, which is characteristic of small-world network, where the community is built around one or more core experts, and the inter-community connection relies on the connection of a few experts, which is a high clustering feature that helps the rapid dissemination of knowledge and innovation. A graph density of 0.002 indicates that the direct collaborative relationships between experts are sparse in this network despite the existence of 7048 collaborative edges. The low density indicates that although the expert network is extensive, most of the experts do not have direct collaborative relationships with each other, which may be due to the fact that the experts work in different fields or projects.

In summary, the constructed expert co-authorship network is a typical undirected network with a universal topology and small-world characteristics, which is characterised by high connectivity within communities and high sparsity between communities.




| Network Overview | |
|---|---|
| Number of nodes | 2970 |
| Number of sides | 7048 |
| Average degree | 4.746 |
| Average path length | 6.407 |
| Average clustering coefficient | 0.721 |
| Map density | 0.002 |

**Fig. 6.** Mapping expert relationship network

Statistics for large expert communities, main experts within communities, and expert degree centrality are presented in Table 1 of the expert knowledge graph.

**Table 1.** Statistics of some communities

| Community ID | Number of nodes | Main experts | Degree centrality |
|---|---|---|---|
| Group 1 | 306 | **Zhang Yi** (Information Engineering University) | 49 |
| | | **Wang Hongyan** (Shandong Agricultural University) | 33 |
| | | **Hua Yixin** (Information Engineering University) | 15 |
| Group 2 | 292 | **Chen Jun** (National Basic Geographic Information Center) | 180 |
| | | **Zhao Renliang** (National Basic Geographic Information Center) | 67 |
| | | **Liu Wanzeng** (National Basic Geographic Information Center) | 40 |
| Group 3 | 171 | **Wang Hong** (Graduate Management Brigade of Air Force Radar Academy) | 109 |
| | | **Li Ping** (Institute of Industrial Control Technology, Zhejiang University) | 17 |
| | | **Wang Yuhong** (School of Surveying and Land Information Engineering, Henan Polytechnic University) | 14 |
| Group 4 | 151 | **Deng Kazhong** (School of Environment and Surveying, China University of Mining and Technology) | 112 |
| | | **Tan Zhixiang** (School of Environment and Surveying, China University of Mining and Technology) | 30 |
| | | **Bian Zhengfu** (School of Environment and Surveying, China University of Mining and Technology) | 17 |
| Group 5 | 131 | **Zhao Yong** (TongJi University) | 75 |
| | | **Wang Lei** (School of Information Science and Technology, Jinan University) | 31 |
| | | **Yang Jing** (School of Computer Science, Beijing University of Technology) | 13 |

From the results of community statistics, the construction of co-authorship circles mainly revolves around one or two core experts, and this core expert occupies a central position in the network and has a greater influence on other experts. In the process of forming the same community, experts tend to co-author with experts from the same institution, such as (1), (2), and (4) communities, and most of the experts in these communities are from the same institution, for example, Chen Jun, Zhao Renliang, and Liu Wanzeng from the National Centre for Basic Geographic Information (NCBGI), indicating that the experts from the same institution work more closely together, and there are frequent academic exchanges and cooperation within the institution. . If there is no suitable expert for co-authorship within the same institution, experts tend to co-author with experts between the same region. For example, although experts Zhao Yong (Tongji University) and Wang Lei (Jinan University) in the community of (5) come from different institutions, their cooperation may benefit from the geographical convenience and the advantages of regional academic exchanges. This regional tendency to collaborate reflects the influence of geographic location on academic collaboration, suggesting that experts consider not only research areas and academic backgrounds, but also geographic convenience when engaging in collaborations.

**Spatialised Representation of Expert Networks.** To interrogate whether the development of co-authorship relationships among experts correlates with regional distribution, a knowledge graph reflecting the affiliations of experts was generated, as delineated in Fig. 7. This spatialized knowledge graph showcases that Beijing, along with the Yangtze River Delta and Pearl River Delta, serves as epicenters for the establishment of co-authorship ties and the congregation of experts. The dense lattice of expert connections across the central and eastern regions suggests less pronounced collaborative trends in these areas of China.

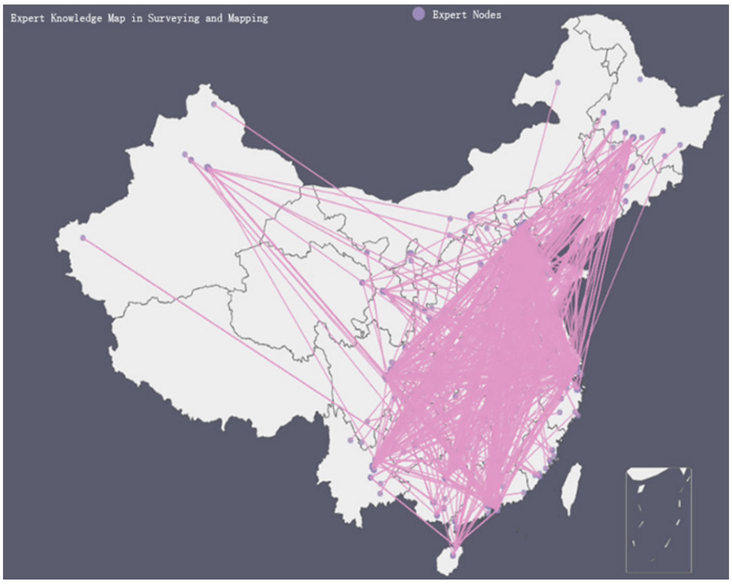
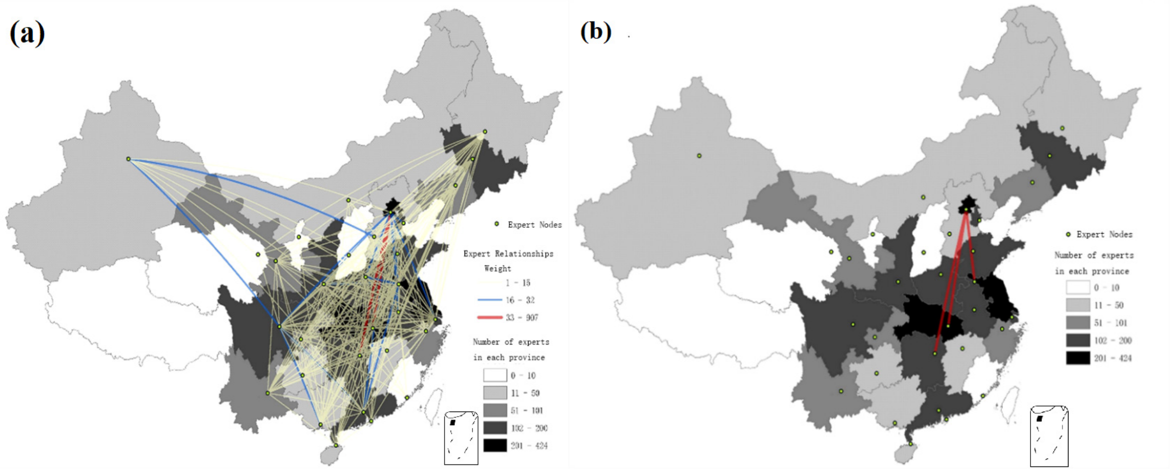



**Fig. 7.** Mapping expert knowledge map spatial results

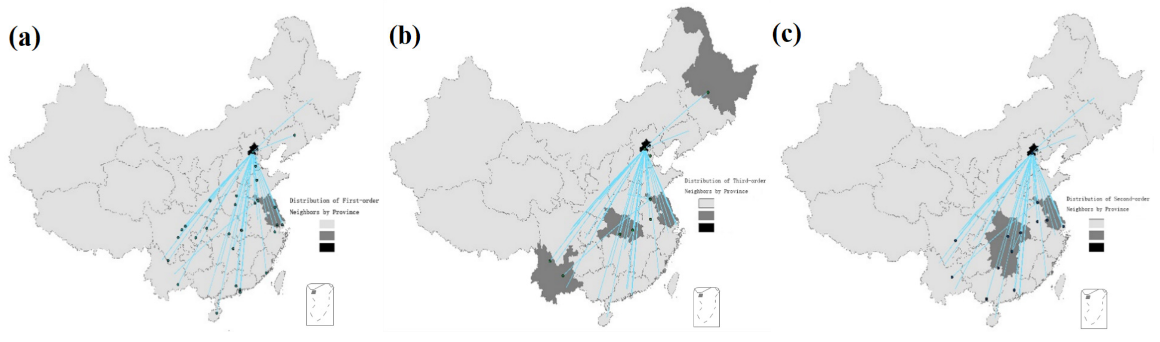
Fig. 7 shows the knowledge map of Chinese mapping experts, in which the purple nodes represent the locations of experts' distribution, and the connecting lines indicate the co-authorship among experts. The map aims to reveal the geographic distribution of mapping experts and their cooperation networks, but the large number of expert nodes and the crisscrossing of expert relationships within each province make the map unable to express the cross-regional co-authorship characteristics and inter-regional expert distribution characteristics well. Therefore, this paper adopts a more refined overlay analysis technique, which can simplify and stratify the complex expert relationships and make the cross-regional co-authorship relationships more clearly presented. This spatial analysis not only demonstrates the density of expert distribution, but also reveals the intensity of cooperation between provinces and cities. For example, Beijing, Shanghai and Guangzhou, as the key research centres, have very close collaborative relationships with the rest of the country and act as core nodes connecting experts from all over the country. Through this method, it can help identify the weak links and potential areas of scientific research cooperation. For example, the research cooperation network in the western region is relatively weak, and thus more policy support and resource input are needed to promote research development and cross-regional cooperation in the region, so as to further optimise the allocation of research resources and promote balanced development nationwide.




(a) Aggregation of experts and merging of relationships (b) Filtering of expert relationships

**Fig. 8.** ArcGIS expert knowledge map construction results

Fig. 8(a) displays an aggregated view of expert entity clustering and co-authorship networks, offering a more succinct and apparent representation of cross-provincial co-authoring dynamics than Fig. 7. The different hues signify the intensity of collaborative ties across regions and the concentration of experts in surverying and mapping. Extending this analysis, Fig. 8(b) illustrates a refined view after further filtering co-authorship links, accentuating the robust interprovincial connections and expert distribution patterns. Fig. 8(a) reveals that interspatial collaborations tend to be infrequent and localized rather than extensive and recurrent. Conversely, Fig. 7 shows that high-density areas of expert concentration are predisposed to forming interregional collaborations, suggesting that these areas are more advanced in the field and exert greater influence, fostering a clustering effect. The expert knowledge graph consists of distinct co-authoring micro-communities centered around pivotal experts. To delve into the collaboration dynamics within these groups, a core expert's co-authoring circle was analyzed by extracting and spatializing its community subnet, with the findings depicted in Fig. 9.




(a) Spatialization of first-order neighbors (b) Spatialization of second-order neighbors (c) Spatialization of third-order neighbors

**Fig. 9.** Filter extraction results

Fig. 9(a) shows the community sub-network extraction of the knowledge graph of experts in the field of surveying and mapping with Chen Jun (National Basic Geographic Information) as the core expert, and the connecting line is the co-authorship relationship of the whole community, and the first-order neighbours' distribution is counted, and the result is that Beijing has the most distribution, followed by Jiangsu, Fig. 9(b) is the second-order neighbours' distribution counted on the basis of Fig. 9(a), and the result is that Beijing has the most distribution, followed by Jiangsu, Hubei, and Hunan, Fig. 9(c) The third-order neighbour distribution on the basis of Fig. 9(a) is statistically performed, and the result is that Beijing has the largest distribution, followed by Jiangsu, Hubei, Yunnan, and Heilongjiang. Further analysis revealed that this phenomenon reveals the characteristics of knowledge dissemination and cooperation patterns in the field of mapping. Beijing, as the core region, has become the hub of the national cooperation network by virtue of its rich resources and strong academic atmosphere. The region not only attracted first-order neighbours to cooperate, but also extended the scope of cooperation through second-order and third-order neighbours, forming a knowledge network radiating across the country. Jiangsu, Hubei and Hunan, as second-order cooperation centres, have also demonstrated strong academic influence and cooperation capacity. These regions not only maintain close cooperation with core experts, but also establish extensive cooperation with experts from other regions through their own scientific research strength.

Fig. 9(a), Fig. 9(b) and Fig. 9(c) show that as the distance of the co-authorship relationship path with the core experts increases, the experts are distributed further apart in space. By comparing these three figures it can be seen that the first-order neighbours around the core expert (Fig. 9(a)) are mainly concentrated in areas closer to the core expert. As the path distance increases, the second-order neighbours (Fig. 9(b)) and third-order neighbours (Fig. 9(c)) are gradually distributed to more distant regions. It is also found that the length of the co-authorship path affects the distribution of experts to a lesser extent in regions with a large number of experts in the relevant field, and the more the number of experts is distributed, the less it is affected by this effect. This is due to the fact that in these regions, cooperation between experts is more frequent and the impact of the co-authorship path on them is relatively small.

## 5 Closings

Co-authorship networks have emerged as a vibrant area of research across multiple disciplines, aiming to reveal latent information in scholarly output. This study developed a co-authorship-based knowledge graph to form an extensive network of expert relationships. We devised a novel technique utilizing co-authorship data to disentangle affiliations of unknown institutions and minimize noise within the network. By applying community detection algorithms, we capitalized on the dense intra-community links and sparse inter-community ties to resolve ambiguities in entity identification. Spatial visualization tools were then utilized to articulate the subtle information embedded in the expert knowledge graph, thus introducing a fresh dimension to knowledge graph analysis. An experimental case study, using a subset of data from the field of surveying and mapping, validated the effectiveness of our approach in assessing co-authorship networks. This paper primarily investigates the spatialization of expert knowledge graphs, examining the distribution patterns of expert entities and the evolution of co-authorship bonds. Future research will build on these outcomes to explore the determinants of inter-regional and intra-provincial co-authorship dynamics.

## References


[1] J. Chen, W.Z. Liu, H. Wu, Z.L. Li, Y. Zhao, L. Zhang, Basic issues and research agenda of geospatial knowledge service, Geomatics and Information Science of Wuhan University 44(1)(2019) 38-47.

[2] Q. Liu, Y. Li, H. Duan, Y. Liu, Z.G. Qin, Knowledge graph construction techniques, Journal of Computer Research and Development 53(3)(2016) 582-600.

[3] Y.N. Pan, H.Z. He, Construction of expert knowledge map based on Co-Author network and citation network, Journal of Intelligence 37(8)(2018) 128-132.

[4] B.Z. Li, J. Su, Domain knowledge graph construction based on expert generated Content, Information Science 36(10) (2018) 13-19.

[5] D.C. Sun, Y. Liu, C.M. Wang, Analysis on the evolution characteristics and social network of science and technology innovation talent policy in Guangdong-Hong Kong-Macao great bay area, World Sci-Tech R & D 44(2)(2022) 199-209.

[6] J.X. Zeng, On construction of scholarly relations network based on massive digital resources, Journal of the China Society for Scientific and Technical Information 32(9)(2013) 929-935.

[7] Y. Yang, Y.Q. Sun, Z.G. Yan, Research on spatial information acquisition method of agency data based on Web crawler, Science of Surveying and Mapping 44(7)(2019) 122-127+140.

[8] Y.L. Shang, J.J. Cao, H.M. Li, Q.B. Zheng, Co-author and affiliate based name disambiguation approach, Computer Science 45(11)(2018) 220-225.

[9] J.H. Liu, M.L. Lin, J. Zhang, A heuristic local randomized feature selection algorithm, Computer Engineering and Applications 52(2)(2016) 170-174+185.

[10] Z.H. Wang, Y.R. Lu, H.T. Zhang, Research on entity disambiguation technique based on association graph and text similarity, Computer & Digital Engineering 49(12)(2021) 2469-2475.

[11] R. Moreno, C. Joel, S. Bres, Texture feature extraction and indexing by Hermite filters, in: Proc. of the 17th International Conference on Pattern Recognition, 2004.
[12] F. Jiang, G.H. Guo, X. Yue, A research method for semantic-based document feature extraction, Computer Science 43(2)(2016) 254-258.
[13] G.X. Zhou, Y. Zhang, Group component analysis for multiblock data: Common and individual feature extraction, lEEE transactions on neural networks and learning systems 27(11)(2015) 2426-2439.
[14] B. Kamal, A. BouyeR, A link-based similarity for improving community detection based on label propagation algorithm, Journal of Systems Science and Complexity 32(2019) 737-758.
[15] R.J. Bai, F.H. Leng, Knowledge innovational evolution analysis based on k-clique community network, Library and information service 57(17)(2013) 86-94.
[16] Z.R. Hu, Research on community detection in social networks based on the Louvain algorithm, Computer knowledge and technology 16(23)(2020) 197-198.
[17] S.E. Garza, S.E. Schaeffer, Community detection with the label propagation algorithm: a survey, Physica A: Statistical Mechanics and its Applications 534(2019) 122058.
[18] J.N. Liu, H.Y. Liu, X.H. Chen, X. Guo, X.M. Zhu, Construction of knowledge graph based on Geo-Spatial data, Journal of chinese information processing 34(11)(2020) 29-36.
[19] Y.Q. Chu, J.J. Ding, Research on community detection in co-authorship networks based on Louvain algorithm, High Technology Letters 31(3)(2021) 257-262.