Heliyon 9 (2023) e19651

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Review article

Knowledge diffusion of Geodetector: A perspective of the literature review and Geotree

Yuting Liang [a,b,c], Chengdong Xu [a,b,*]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[b] Department of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China
[c] Department of Human Geography and Planning, Utrecht University, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Geodetector
Spatial heterogeneity
Network analysis
WoS
Geotree

ABSTRACT

Spatial heterogeneity is a fundamental research topic in the field of geography, and Geodetector is a widely used tool for studying this phenomenon. To understand the research advancements and knowledge diffusion trends surrounding Geodetector, we constructed an author evolutionary tree structure fusing its 847 core citations in the Web of Science database and Geotree model for the first time. The results of our literature statistics indicated that Geodetector has garnered the attention of 3123 authors from 48 countries since its publication in the Ecological Indicators journal in 2010, who have published core papers concerning ten important topics. The majority of these studies focused on spatial heterogeneity and its influencing factors. Our analysis of Geotree data revealed a significant correlation between the publication rate of scholars in large teams and their academic activities. Our analysis of the knowledge diffusion chain shown that only 2% of the total number of authors have contributed to over 20% of the scientific collaborations and knowledge diffusion, and they were recognized as experts in Geodetector research. To provide a comprehensive reference for future scholars, we have summarized the citing countries, five classical articles, main scientific domains, and core teams of Geodetector research.

1. Introduction

Geography is an interdisciplinary field that combines natural science and social science and has the qualities of being comprehensive, interdisciplinary, and regional. The study of geospatial attributes is mainly divided into two aspects: spatial autocorrelation and spatial heterogeneity. Spatial heterogeneity includes spatial stratified heterogeneity and spatial local heterogeneity, and Geodetector is widely used to detect the former [1]. Geodetector has been applied in various research fields by scholars around the world including China, America, the United Kingdom, Sweden, Australia and Italy, covering subjects such as cancer [2], global warming [3], eco-geographical division [4], vegetation index change [5], urban park usage [6], population distribution [7], PM2.5 [8], forest carbon monitoring [9,10], air pollution [11,12], urban livability [13], air quality index [14], renewable energy industry [15], soil moisture changes [16], and environmental metrology [17,18]. The transnational and interdisciplinary diffusion of Geodetector knowledge provided technical conditions and theoretical practices for the cross-fertilization of geography with other fields and contributed to the

* Corresponding author. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China.
E-mail address: xucd@lreis.ac.cn (C. Xu).

https://doi.org/10.1016/j.heliyon.2023.e19651
Received 8 June 2023; Received in revised form 28 August 2023; Accepted 29 August 2023
Available online 30 August 2023

emergence and development of related emerging research directions.

Despite benefiting from interdisciplinary and international multidimensional knowledge dissemination, Geodetector has been widely adopted and applied in various scenarios, conditions, and domains. However, the overall trends of these applications and the developmental trajectory of the underlying theory have yet to be summarized to explore further prospects for advancement. Scientometrics and bibliometrics offer a framework for the analysis of scientific literature and provide a database for the same. Scholars have used these tools to study the knowledge domains and emerging trends in various fields, including nanoparticle drug delivery technologies, echinococcosis research, GIS, organic photovoltaic technology, vulnerability assessment in the context of climate change, and sustainable development research [19–27]. Other researchers have utilized advanced techniques and visualization methods to deepen the analysis of scientific literature. For example, White (2010) and others introduced the use of correlation theory [28]. Several studies have also analyzed research hotspots and their evolving trends through bibliometric analysis and GIS techniques [29–31]. Additionally, Duan (2020) et al. applied an evolutionary tree structure to visualize the evolution of various research areas [32]. Despite these advancements, the field of literature review still lacks an in-depth analysis of the underlying knowledge diffusion processes present in the literature. These studies, primarily conducted through bibliometric analysis and visualized in various forms such as networks, maps, and evolutionary trees, provide valuable insights for researchers and policymakers across multiple fields.

Since the 1980s, economists have studied the diffusion of knowledge through social networks such as telephone, fax, and mail [33]. Studies have demonstrated that the tacit knowledge obtained through direct interaction is more profound, leading to a growing focus on the social space of knowledge diffusion [34]. Initially, researchers used knowledge networks to depict scientific collaborations, including rule networks, random networks, and small-world networks [35–38]. However, these models lacked consideration for individual heterogeneity, resulting in the introduction of weighted networks, complex networks, and collaboration hypernetwork [39,40]. Additionally, some scholars used epidemic models to illustrate knowledge diffusion, comparing the spread of knowledge to that of viruses [41,42]. Despite the differences between individuals' behavior in the face of knowledge and viruses, some researchers have innovatively improved epidemic models based on knowledge diffusion mechanisms, such as environmental feedback, self-learning [41,43], friendship incentive [44], review [45], and individual perception [42]. Hence, it is crucial to take into account both the social space and individual heterogeneity when analyzing knowledge diffusion through literature review, meaning that knowledge diffusion modeling should be carried out at the individual level.

To reach that goal, this study proposed a novel approach by incorporating the concept of social space in scientific knowledge graphs

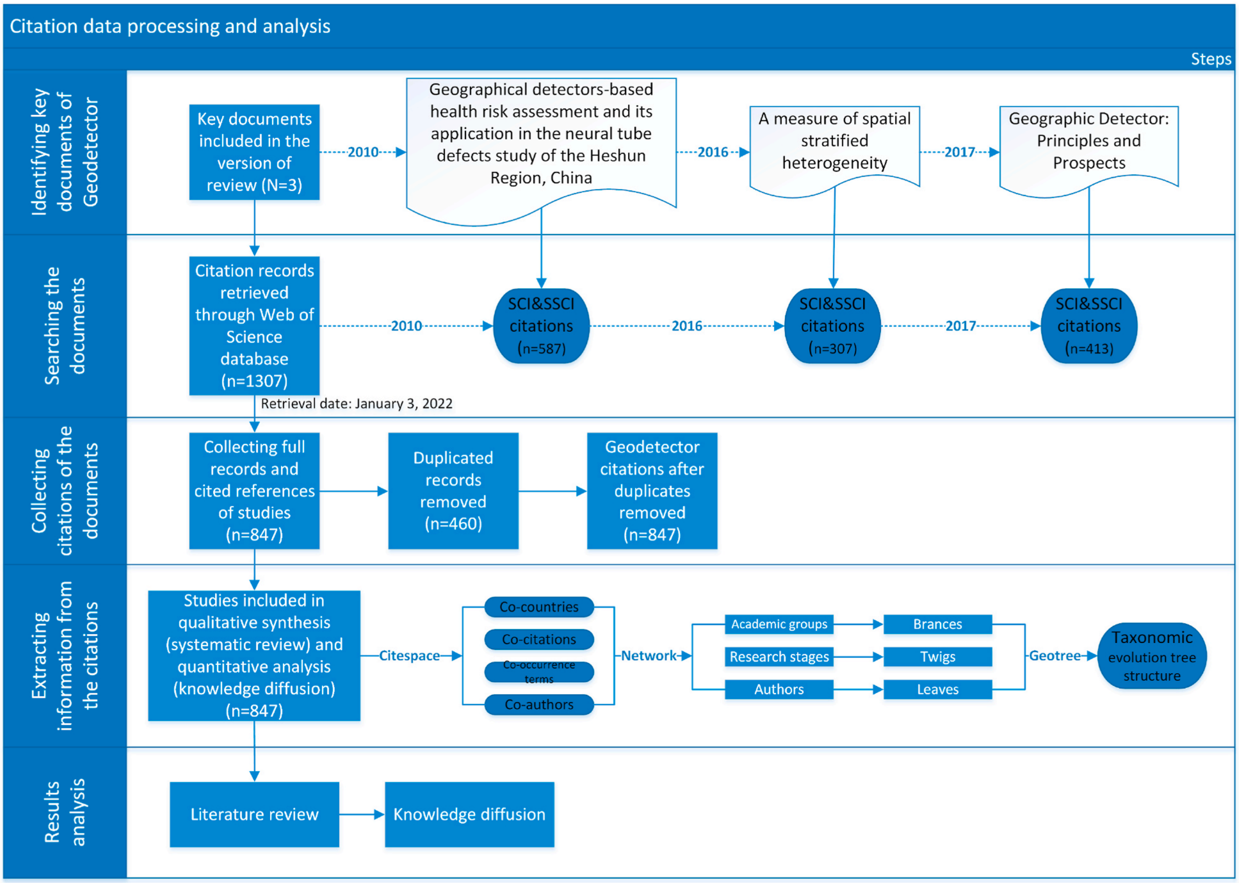

Fig. 1. The collection and processing of studies involved in a systematic review and knowledge diffusion analysis.

into a spatiotemporal evolution tree model, resulting in a knowledge diffusion evolution tree model that considers individual heterogeneity. The spatiotemporal evolution tree model was used to analyze the heterogeneity and evolution of research objects. A concern for this approach is that because modeling results are influenced by input parameters, so the application of the model requires scholars to have a deep understanding of the relationship between the classification and development trends of the research objects. For instance, the evolution tree of land expansion can be used to study the dynamic change in urban construction land based on city types and urban development stages [46]. The evolution tree of a comprehensive evaluation of urban compactness consists of four secondary indicators and twelve tertiary indicators [47]. The evolution tree of the accidental disability population is based on social security levels and industrial structures [48]. Other examples of the domain-specific knowledge necessary to use the evolution tree model effectively include works from Zhang [49], Wang and Liu [50], He [51], Hu [52], Wang and Wang [53], Duan [32], etc.

Based on the information discussed above, we conducted a search in the Web of Sciences core citation database, utilized the widely-used citation analysis tool Citespace to analyze the distribution of Geodetector knowledge across various fields worldwide, and employed the Geotree to model the evolution of the knowledge taxonomy. For the first time, this study utilized a bibliometric method to identify research hotspots and dynamic trends of Geodetector research from the past ten years, offering new insights into the main progress of spatial stratified heterogeneity detection research. Moreover, the study introduced a novel author evolution tree model that considers both social space and individual heterogeneity, providing a new perspective on the knowledge diffusion process. Different from many regional geographic research and bibliometric analyses, this study proposed a new idea of quantifying, spatializing, and individualizing the knowledge diffusion process through citations using geoscientific methods. The author evolution tree model developed in this study can be applied to various fields of knowledge diffusion research.

2. Materials and methods

2.1. Research data

The data in this article was downloaded from the databases of the Science Citation Index (SCI) and Social Science Citation Index (SSCI) created by the American Institute for Scientific Information (Fig. 1). The team of researcher JF Wang from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, whose three publications laid the theoretical foundation of spatial heterogeneity statistics, proposed the Geodetector model: the first published spatial heterogeneity for statistical attribution [54]. The second published the probability density function of Geodetector q statistic [55]. The third published the geoscience principle of Geodetector [1]. Accordingly, we searched the above three articles in the Web of Science (WoS) scientific citation index database (retrieval date: January 3, 2022), and obtained the SCI&SSCI citation data of Geodetector for 12 consecutive years (2010–2021), with total of 1307 valid records downloaded. Among them, 1283 records were articles, 15 records were proceedings, and 9 records were reviews. After checking the duplicates, 847 unique records were obtained.

2.2. CiteSpace

CiteSpace (Version: 5.8. *R3c*) is an information visualization tool developed using Java language [56]. Its input includes S&E citation databases such as WoS, Scopus, Derwent, CSSCI, and CNKI. Citespace has been widely used in the analysis of research hotspots, fronts, and trends [19,23]. Based on this, the co-citation networks of documents, authors, and journals the co-citation networks, the collaboration networks of authors, institutions, and countries, and the co-occurrence networks of keywords, terms, and domains are drawn. The interpretation of the scientific knowledge graphs usually has the following points: network structure, network clusters, inter-cluster associations, key nodes, and paths. Betweenness centrality, burst detection, and citation tree-rings are three critical terms used to measure article impact, identify research fronts, and display citation history, respectively [19]. Literature with high betweenness centrality is often a key hub connecting two different fields, so it is also called a turning point. The calculation formula of this indicator is as follows [57]:

$$BC_i = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}} \tag{1}$$

In formula (1), g_{st} is the number of shortest paths from node s to node t, and n_{st}^i is the number of shortest paths passing through node i among the g_{st} shortest paths from node s to node t. In short, the betweenness centrality metric describes the importance of a node in terms of the number of shortest paths passing through a node.

2.3. Spatiotemporal evolution tree

The spatiotemporal evolution tree model was proposed by the research team of JF Wang [46,58,59], which can simply and clearly represent the evolution process and hidden mechanism of the research object through a multi-dimensional tree coordinate system. This model is based on the evolutionary theory of biology and is often used in the analysis of spatial and temporal data in geography [59]. During the development and practice of Geodetector theory, the authors of the citations play a major role in promoting knowledge diffusion. Treating the complexity of knowledge diffusion as one would an organic individual, the multi-dimensional process of authors has the characteristics of life evolution.

Therefore, this paper adopted the Geotree software to construct the knowledge diffusion evolution tree of Geodetector (http://

www.sssampling.cn/geotree/) (Fig. 2). The academic groups formed by the author cluster and the research stage divided according to the node degree centrality were used to represent the author classification and evolution stage respectively, which constituted the first-level branch and the second-level branch of the tree model in turn, and the leaves represent the authors. The final result was beautified with draw. io software (http://draw.io/).

3. Results

3.1. The distributions of knowledge based on network analysis

3.1.1. The geographical space of Geodetector knowledge

Mapping the co-countries network helps to examine the cooperative relationship between countries, analyze the status of different countries in the cooperative network, and explore the core position of spatial heterogeneity research and the development of Geodetector theory. And it helps to explain the macro scientific research and production mode. Of the 847 articles collected in this study, 833 have the country/region field, and 14 have this field missing. After merging the documents of regions into their countries, authors are distributed in 48 countries around the world, and each country publishes 17.35 papers on average. As shown in Fig. 3, there are a total of 48 nodes and 92 links in the Geodetector co-countries network, and the network density is 0.0816. This shows that the spatial diffusion of Geodetector knowledge is very extensive, but the cooperative relationship among countries is unbalanced, that is, some countries cooperate closely and some countries cooperate very little. The burst detection result (red circles in Fig. 3) shows a surge in citations in the United States. The spotlight detection result shows that China, the United States, the United Kingdom, Italy, and the Netherlands are the core nodes, which is closely related to the distribution of many famous geography-related master's and doctoral degrees in these places.

To understand the spatial distribution pattern of knowledge, we draw its world citation distribution heatmap (Fig. 4) based on cartography and geographic information system technology and convert the virtual scientific knowledge graph into geographic space. In order of the number of citations, the TOP ten countries are: China, the United States, Australia, the United Kingdom, the Netherlands, Germany, Canada, Italy, South Africa and Sweden. Except for Italy and Sweden, the node betweenness centrality values of eight countries are all greater than 0.1, indicating that these countries carry out more exchanges and cooperation in spatial heterogeneity research. According to the cooperation time nodes, China and the United States are the first to establish cooperative relations with other countries in Geodetector research (2010), followed by Australia and the United Kingdom (2013), Canada (2014), then the Netherlands and Sweden (2015), Germany (2016), and finally Italy (2018) and South Africa (2019).

3.1.2. The citation space of Geodetector knowledge

Mapping the document co-citation network helps to analyze the role of documents in the collaborative network. The nodes in the network represent documents, and the larger the node, the greater the reference value of this research in the application of Geodetector. In the document co-citation network of the Geodetector (Fig. 5), the number of nodes is 656, the number of links is 2304, and the network density is 0.0107. The Modularity Q (MQ) and Mean Silhouette (MS) values are 0.8104 and 0.92, respectively, indicating that the network density is small, but the clustering effect is good and the co-citation degree is high. This forms many sub-clusters, with frequent intra-sub-cluster cooperation and less inter-sub-cluster cooperation.

Table 1 lists the five documents most frequently cited by users in the Geodetector co-citation network. They interpret Geodetector

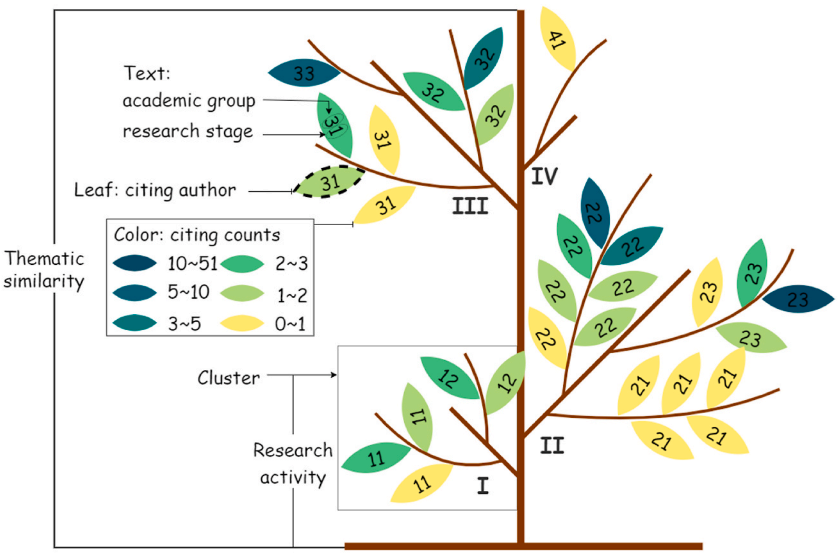

Fig. 2. The Geotree structure of Geodetector authors.

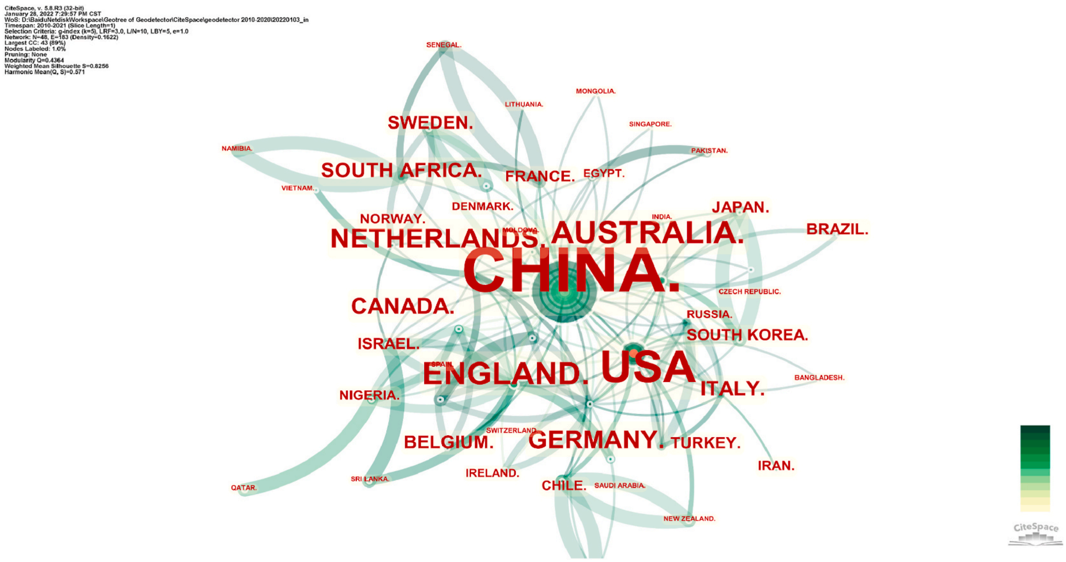

Fig. 3. The co-countries network in the field of Geodetector.

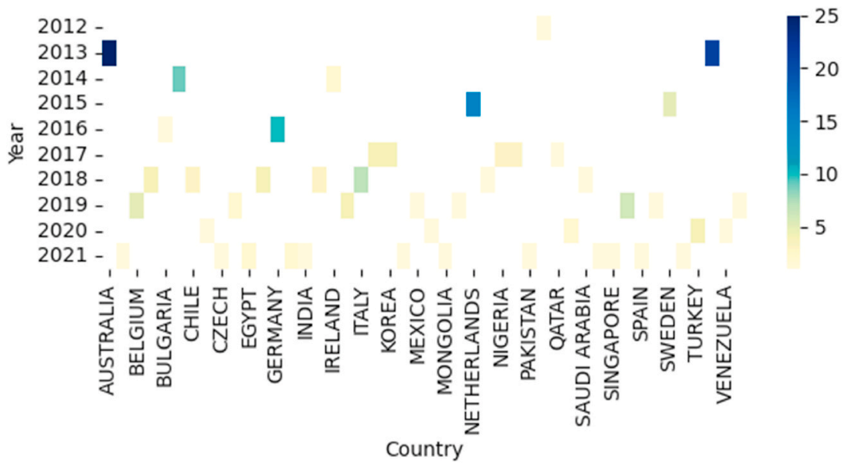

Fig. 4. Heatmap of citations by country and first cited year. Excluding China (766 citations in 2010) and the United States (110 citations in 2010).

from the perspectives of principles, technologies, and applications. Among them, the most popular document proposed the q-statistic method to measure the spatial stratified heterogeneity [55]. The second document describes the geoscience principles of Geodetector and provides a method guide for users [1]. The third document is based on a Geodetector analysis of the physical geographical zoning characteristics of the United States [60]. The fourth document is the first to propose a Geodetector and apply it to the study of impact factors of the incidence of neural tube defects in Heshun County, China [54]. The fifth document applies Geodetector to the study of built-up land expansion [61]. Therefore, when applying Geodetector for the research of spatial stratified heterogeneity or factors detection, the above five documents can be prioritized for reference.

3.1.3. The application domains of geodetector knowledge

The co-occurrence terms network help to analyze the main content and research hotspots of application cases. The nodes in this network (Fig. 6) are terms, and the size of each node indicates the co-occurrence frequency of terms. Colored lines between nodes indicate the first co-occurrence year of terms: green for the earliest and yellow for the latest. In Geodetector studies, "spatial distribution" has the highest co-occurrence frequency (132), followed by: "influencing factor (115)", "spatial heterogeneity (100)", "driving factor (97)", "spatial pattern" (89)", "geographical detector (88)", and "dominant factor (66)", etc. It shows that the main contents of the application of Geodetector are influencing factors and spatial heterogeneity, and there is little research to distinguish between spatial stratified heterogeneity and spatial local heterogeneity. The cluster labels in the network are based on Latent Semantic Indexing (LSI), so research hotspots of Geodetector are land use (105), urban agglomeration (96), human activities (76), soil erosion (105),

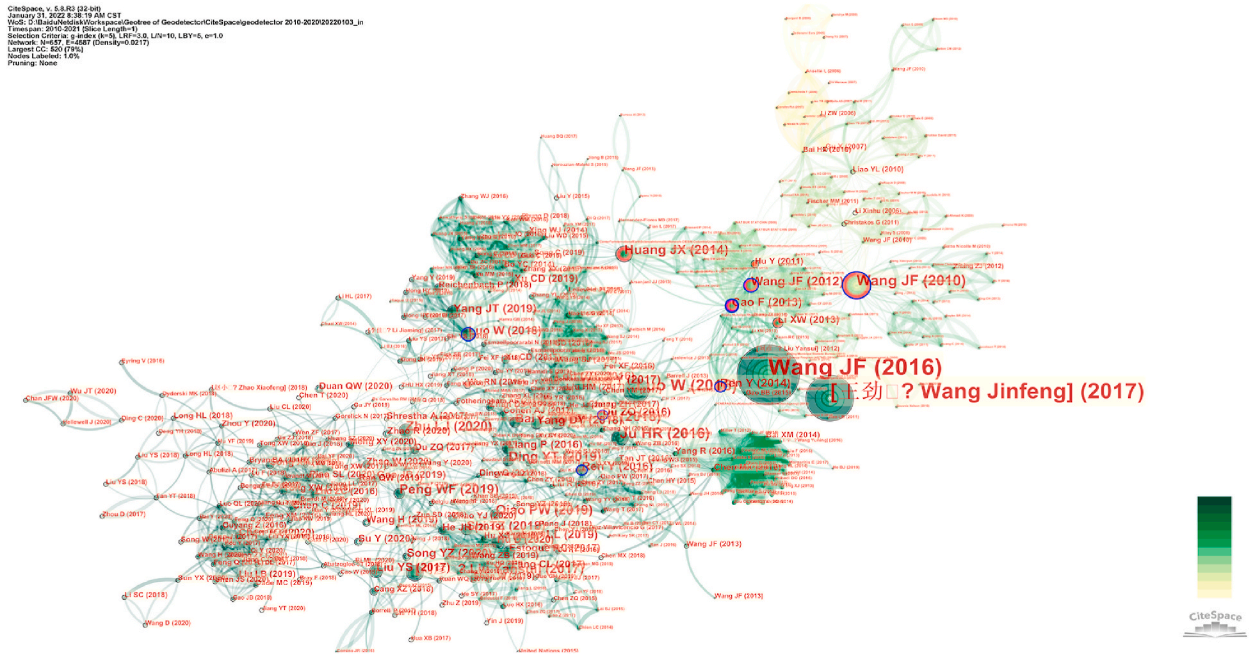

Fig. 5. The co-citation network in the field of Geodetector. Blue circles indicate high abruptness; red circles indicate high betweenness centrality.

Table 1
Top five most co-cited articles in the Geodetector field.

Co-cited Frequency	[a]Google Scholar	Title	Author	Year	Half Life	Journal
400	872	A measure of spatial stratified heterogeneity	JF Wang et al.	2016	3.5	Ecological Indicators
291	681	Geodetector: Principle and prospective	JF Wang et al.	2017	3.5	Acta Geographica Sinica
54	116	Spatial association between dissection density and environmental factors over the entire conterminous United States	W Luo et al.	2016	3.5	Geophysical Research Letters
38	1363	Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China	JF Wang et al.	2010	3.5	International Journal of Geographical Information Science
38	112	Driving forces and their interactions of built-up land expansion based on the geographical detector – a case study of Beijing, China	HR Ju et al.	2016	3.5	International Journal of Geographical Information Science

[a] Google Scholar: citations retrieved on Google Scholar on February 13, 2023.

rainstorm waterlogging (51), and landslide susceptibility (41), etc. The numbers in parentheses indicate the documents included in the cluster.

Burst detection reflects the short-term change in the frequency of a word or phrase in the Geodetector citing documents. Terms appear at different times and durations, reflecting emerging trends in different periods and the development trajectories of Geodetector theory. The results show that the emerging trends of geographic detectors in 2010 were northern China, birth defects, and neural tube defects. In 2013, it was risk factor. In 2015, it was the geographical detector. In 2016, they were spatial association and population density. In 2017, they are hfmd disease, geographical detector model, and relative risks. In 2018, they are urban planning, geographical detector technique, cold spots, and spatial stratified heterogeneity. In 2019, they are the main influencing factors, relative humidity, eastern China, hfmd disease, and distribution pattern. Among them, the longest outbreak of "Northern China" studies has lasted for nine years, from 2010 to 2018.

3.2. The Geotree modeling and knowledge diffusion analysis

Among the 847 papers collected in this study, 833 of which contain the author field, 14 papers are missing this field. There are 3123 unduplicated authors, with an average of 3.75 authors per paper, and the frequency of author appearances is 4293. Based on the default threshold parameters, we plot the co-authorship network as shown in Fig. 7. The network has a total of 365 nodes, 704 links,

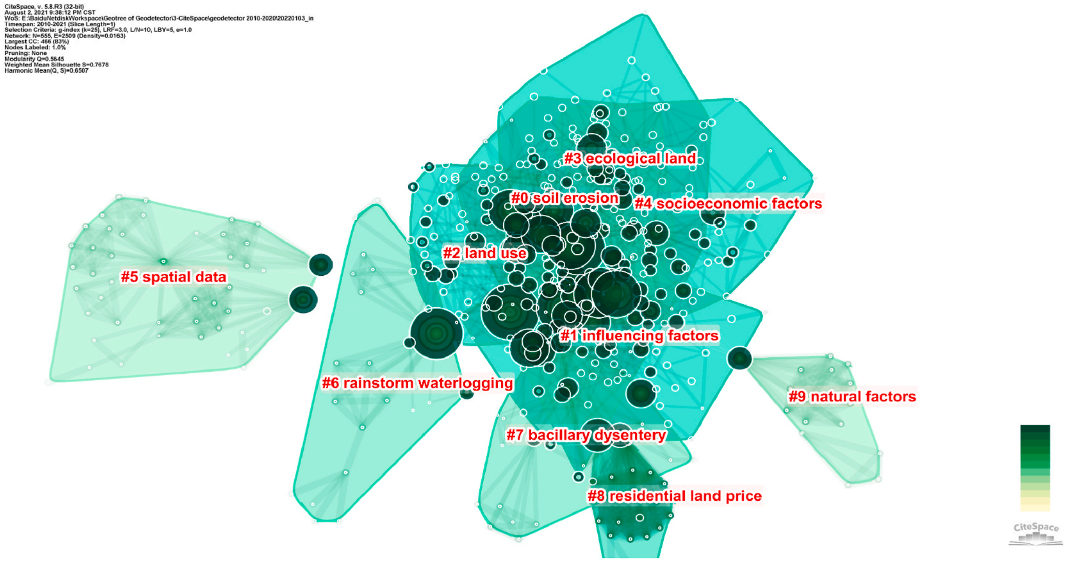

Fig. 6. The co-occurrence terms network in the field of Geodetector.

and a network density of 0.0106. It shows that there is extensive and frequent cooperation among the authors of Geodetector, but the cooperation relationship between them is not close or balanced. The MQ value of the network is 0.80, and the MS value is 0.90, indicating that the network has a high degree of modularity, indicating that there are academic groups with clear boundaries, and there is a close cooperative relationship within the divided academic groups.

3.2.1. *First-level branch—academic groups*

According to the clustering results of the co-authorship network in Table 2, there are seven main academic groups of Geodetector research (132 scholars in total, this table only shows 87 scholars who have published more than two related papers), which were formed in 2010, 2011 and 2019. Therefore, the first-level branch of the author's Geotree is divided into seven parts. These studies mainly involve three disciplines: geographic information, ecological environment, and public health. The research topics from academic group one to academic group seven are spatial data, heavy metals, schistosomiasis risk, birth defects, climate change, neural tube defects, and soil salinity.

As for the closeness of cooperation, the average level of the silhouette of these seven clusters is 0.923, so the cooperation relationship within each group is very close. From the perspective of the author's contribution, Jinfeng Wang, George Christakos, Yilan

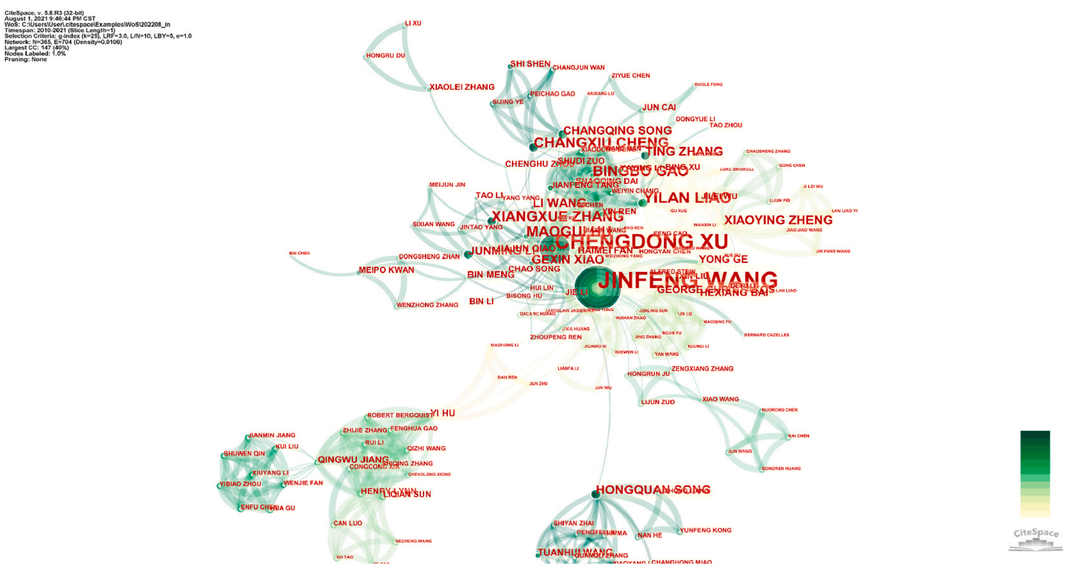

Fig. 7. The co-authorship network in the field of Geodetector.

Table 2
The core academic groups of Geodetector research.

Cluster ID	Group Members	Silhouette	Year	Clustering labels	
				LSI	LLR
1	28 members including Jinfeng Wang, Maogui Hu, Xiangxue Zhang, Bingbo Gao, Li Wang, etc.	0.786	2010	spatial data	spatial data
2	19 members including Chengdong Xu, Yilan Liao, Yin Ren, Jianfeng Tang, Yaying Li, etc.	0.841	2010	heavy metals	ecological factor;
3	12 members including Yi Hu, Qingwu Jiang, Liqian Sun, Henry Lynn, Robert Bergquist, etc.	0.933	2011	schistosomiasis risk	sandwich method
4	Xiaoying Zheng, Ting Zhang, and Jilei Wu.	0.965	2010	birth defects	birth defect
5	13 members including Hongquan Song, Tuanhui Wang, Lizhong Liang, Guangli Zhang, Nan He, etc.	0.999	2019	pm concentrations	meteorological condition
6	George Christakos and Xin Liu.	0.952	2010	ntd determinants	hongta district
7	10 members including Changxiu Cheng, Changqing Song, Shi Shen, Xiaolei Zhang, Chengfu Zhou, etc.	0.988	2019	soil salinity	soil salinity

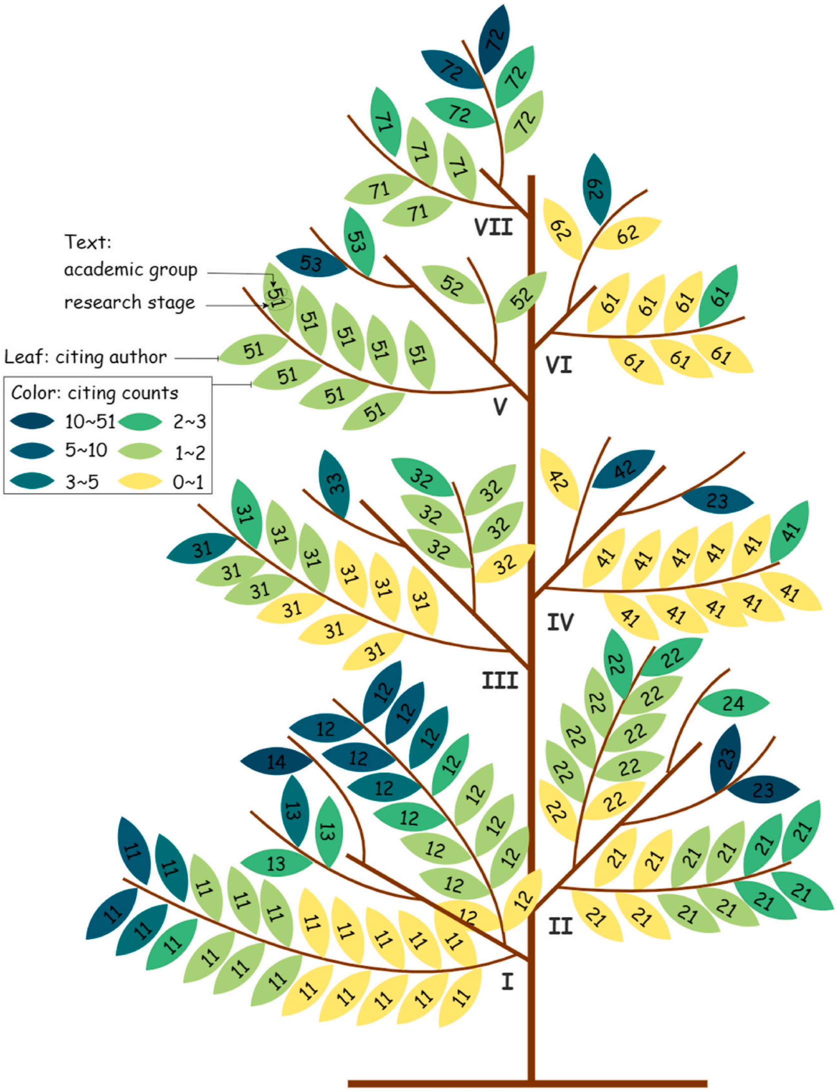

Fig. 8. The Geotree of Geodetector authors.

Liao, Ting Zhang, and Xiaoying Zheng first proposed the model and then formed an academic group with a certain scale. It has promoted the knowledge diffusion and theoretical development of Geodetector in different disciplines and different research fields.

3.2.2. *Second-level branch—research stages*

Betweenness centrality and burstiness detection can identify core and frontier scholars in the co-authorship network. Among the 365 authors of Geodetector, Jinfeng Wang's betweenness centrality value is the highest (0.2), which plays a key role in knowledge diffusion, and other scholars are all lower than 0.1. The top three authors of the burst detection values are Xiaoying Zheng (from 2010

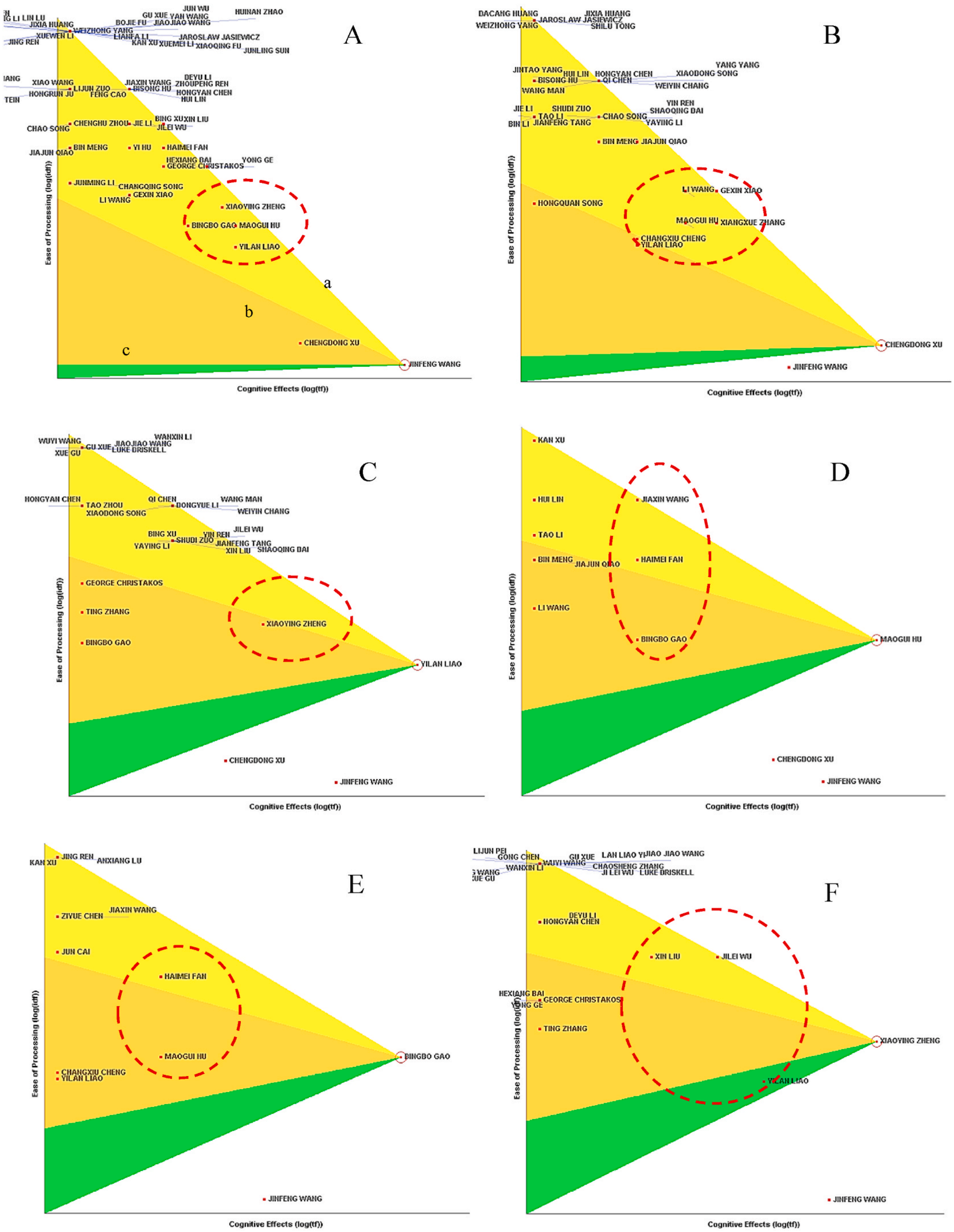

Fig. 9. The pennant maps of spreading authors. Picture A is the pennant map of the primary author, and pictures B–F are the pennant maps of the second-level spreading authors.

to 2014), Jinfeng Wang (from 2010 to 2014), and Yilan Liao (from 2010 to 2018), so their papers represent the frontiers in Geodetector research. Among them, Liao Yilan has the longest outbreak, leading the trend for nine years.

In the network, the larger the degree centrality value of a node is, the more links between this node and other nodes, and the author corresponding to the node has a wider academic communication circle. The index directly reflects the authors' activity and research depth in the academic community. Therefore, we classify the degree centrality of all nodes into three levels through the natural breakpoint algorithm to characterize the authors' research stages. Among them, stage one represents the enlightenment stage of early research (1–6), stages two and stage three represent the growth stage with the rapid increase in research results (6–12; 12–19), and stages 4 and 5 represent the mature stage of in-depth research (19–33; 33–58).

3.2.3. Leaves—authors

Based on index data and cluster analysis, the academic groups of Geodetector consist of 132 authors. Dividing them into different academic groups and research stages, all authors correspond to 132 leaves in the tree structure. The color of the leaf indicates the count of articles published by the author and is classified into 6 intervals.

3.2.4. Taxonomic evolution characteristics of Geodetector knowledge

Based on the above modeling process, the author evolution tree of the Geodetector was calculated, and its tree structure visualized by Geotree is shown in Fig. 8. The model describes the distribution of citation counts in different academic groups at different research stages and reveals the taxonomic evolution characteristics of citation knowledge. Among them, academic group 1 and academic group 2 at the bottom of the Geotree have the largest scales, scholars in group 1 are distributed in research stages 1, 2, 3, and 5, and scholars in group 2 are distributed in research stages 1, 2, 3, and 4. The authors numbered "15″ and "24″ are those with the most citations to Geodetector. Academic groups 3, 4, and 5 in the middle of the Geotree are of the same scale. Scholars in Group 3 are distributed in research stages 1, 2, and 3, scholars in Group 4 are distributed in groups 1, 2, and 5, and scholars in Group 5 are distributed in the research stages. In groups 1, 2, and 3, the authors with the most citations are numbered "42″, "45″ and "53". The academic group 6 in the middle of the Geotree and the group 7 at the top are the smallest in scale. The scholars of both groups are distributed in research stages 1 and 2, and the author with the most citations of geographic detectors is numbered "72". Therefore, in the large-scale academic group, scholars with high academic achievement output are also highly active. In the medium-scale academic group, there is no obvious correspondence between the activity of scholars and their achievement output, while in a small-scale academic group, scholars with low activity also have more academic achievements.

3.3. Authors' contribution to knowledge network

3.3.1. Authors' relevance

In addition to the diffusion process of knowledge in different research stages of different academic groups, the research also wants to further understand the contribution of scholars to the knowledge network. First, we inputted authors as seeds for the pennant maps seen in Fig. 8, thereby identifying the relevance of co-authorship to seeds. The pennant diagram was proposed based on the Relevance Theory of language pragmatics, and its basic expression is Relevance = Cognitive effects/Processing effort [62]. The pennant diagram in information science utilizes Weight = TF × IDF to represent the importance of a word to the document, the X-axis corresponds to TF (Term Frequency), and the Y-axis corresponds to IDF (Inverse Document Frequency). If the term of seed is an author's name, all authors that co-occur with that author can be listed and sorted by frequency of co-occurrence [20,28].

Since Jinfeng Wang is the first author and corresponding author of the first authoring paper of the Geodetector, we inputted this author as the primary seed to construct the pennant model shown in Fig. 9A. As shown in the figure, the spreading author is on the far right and the bottom of the figure, indicating that the biggest cognitive effect in Wang's collaborative research is produced by reading

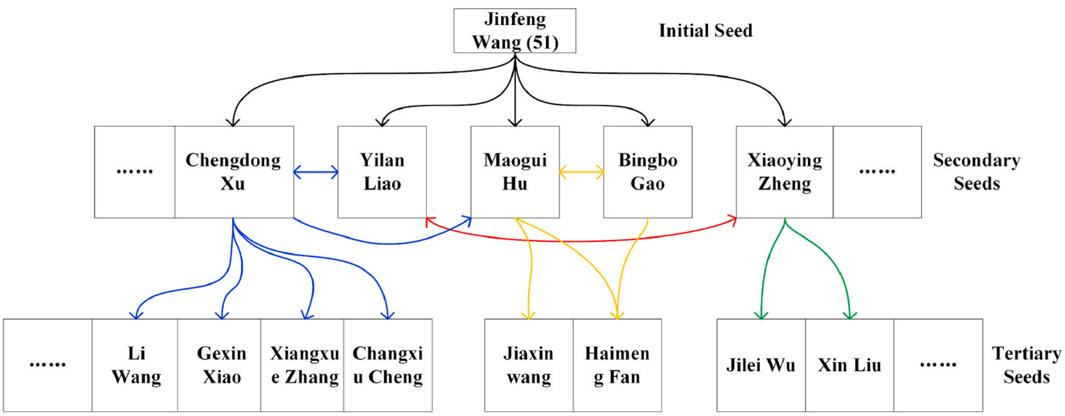

Fig. 10. Schematic diagram of the spreading author's knowledge diffusion chain. The text in the box is the name of the spreading author, and the number in parentheses is the publication count.

Wang's research. As shown in Fig. 9A, area *a* corresponds to some specific fields of the spreading author's research. Area *b* corresponds to some other fields of research in the discipline in which the spreading author majors. Area *c* corresponds to research in other disciplines. In Wang's pennant map, Chengdong Xu is in area *b*, while other authors are in area *a*, indicating that Xu's research direction is close to Wang, and the application cases of Geodetector are distributed in different disciplines. Most of the authors in block *a* are distributed close to the x-axis origin, indicating that they have less relevant studies. These studies are sporadically distributed in different disciplines and have the characteristics of wide diffusion and shallow depth, which reflects the general characteristics of Geodetector being used in interdisciplinary research on spatial stratified heterogeneity.

3.3.2. Knowledge diffusion chain

Then we constructed pennant maps for the 5 second-level spreading authors (see Fig. 9B ~ F), including four authors in area *a* that are significantly far from the origin of the x-axis and close to the origin of the y-axis and only one author in area *b*. Then, the authors most associated with these five authors are selected according to the pennant graphs. So, in Fig. 10, Jinfeng Wang is the primary seed at the top of the diffusion chain, the second layer (the secondary spreading authors) shows the top five authors with the highest research similarity and co-occurrence frequency among Jinfeng Wang's co-authors, and the third layer (the tertiary spreading authors) shows the authors most associated with secondary spreading authors.

Therefore, JF Wang's main collaborators are CD Xu, YL Liao, MG Hu, BB Gao, and XY Zheng. The two-way arrows in Fig. 9 show a strong coupling relationship, and the research contents of the two are also similar. According to the author's cooperation network, there are 160 network links between the original seeds and secondary seeds, accounting for 22.73% of all network links (704). Six spreading authors account for 1.6% of all nodes (365), which means that less than 2% of spreading authors promote more than 20% of scientific cooperation and knowledge diffusion. The knowledge diffusion chain constructed according to the spreading authors plays a key role in analyzing the knowledge diffusion process.

According to the author's evolution tree, JF Wang, BB Gao, and MG Hu belong to academic group 1, CD Xu and YL Liao belong to academic group 2, and XY Zheng belongs to academic group 4. Among them, BB Gao and MG Hu are in the growth stage of research, and the other four spreading authors are in the mature stage. This shows that the publication volume and academic cooperation rate of the spreading authors selected according to the relevance theory are at the middle and upper levels. These authors are related to three of seven of the academic groups. The main collaborators of the primary seed and their collaborators constitute the core of the co-authorship network. According to the authors' institutions, these studies include both intra-institutional and inter-institutional collaborations.

4. Discussion and implications

This study builds upon previous research while addressing existing gaps in the understanding of the Geodetector model's theory evolution and knowledge diffusion [32]. We amalgamate diverse disciplinary backgrounds, offering insights into geography, spatial heterogeneity, and impact factor analysis through the practical implementation of a geographical method in citation analysis and scientometric evaluation. Our efforts and findings bring forth theoretical and practical significance to some fields, such as spatial heterogeneity, evolutionary pattern analysis, literature reviews, and knowledge diffusion.

At the outset, our research stems from the recognition of a research gap in the literature surrounding the Geodetector model [1]. While its widespread application is acknowledged, a comprehensive exploration of its theoretical development history and the mechanisms governing knowledge diffusion was lacking. By bridging this gap, our study lays the groundwork for deeper insights into the Geodetector's impact and significance across various domains. Importantly, this research makes substantial contributions to the advancement of the Geodetector model itself. Through our meticulous literature review, we unveiled the global reach of the Geodetector's influence, as evidenced by the engagement of authors from 48 countries and 3123 authors. This comprehensive mapping of the scholarly landscape provides an invaluable resource for researchers seeking to delve into the Geodetector's theoretical foundations and its evolving trends.

Additionally, our introduction of the Geotree method constitutes a novel approach that contributes to both theory and practice [59]. The Geotree method, which offers a visual representation of knowledge diffusion, serves as a powerful tool for citation analysis and scientometric investigation. This methodological innovation empowers scholars to gain insights into knowledge dissemination trends while also facilitating interdisciplinary collaboration. From a theoretical standpoint, this study enriches the field of evolution trees by shedding light on the diffusion of knowledge within these contexts. The application of the Geotree method serves as a bridge between various disciplines, offering a versatile approach to understanding how ideas spread and impact different areas of knowledge.

Furthermore, our research has practical implications that extend beyond the theoretical realm. The Geotree method's application to knowledge diffusion provides actionable insights for scholars and practitioners seeking to strategically collaborate and leverage scholarly interactions. Our study paves the way for impact factor analysis within the domain of Geodetector research, offering a quantitative assessment of its influence on the academic landscape. By expanding the Geotree method to encompass broader thematic areas, researchers can unravel more generalized patterns of knowledge diffusion [39,40].

In summary, our study stands as a pivotal contribution to the understanding of the Geodetector model's theory application and development trends. By addressing research gaps and introducing the Geotree method, we offer insights that resonate across diverse disciplinary backgrounds. The theoretical and practical implications underscore the importance of interdisciplinary collaboration and the potential for methodological innovation in shaping the trajectory of knowledge dissemination and impact assessment.

5. Conclusions

The Geodetector model's versatility across different fields makes the knowledge it generates applicable in multiple disciplines and globally. Studying its diffusion process through a single bibliometric statistic is challenging. To overcome this obstacle, this paper presents an author evolution tree model based on scientific knowledge graph analysis to understand the diffusion of knowledge generated by the highly adaptable Geodetector model across various fields. The Geotree in multi-dimensional knowledge diffusion analysis illustrates the distribution and evolution pattern of knowledge through multi-level branches. This paper's Geotree of authors is divided into academic groups (first-level branches) and research stages (second-level branches). Scholars are represented as leaves with color depth indicating their research outputs, allowing for the extraction of knowledge diffusion information across different academic groups and stages.

Our study uses bibliometric analysis, relevance theory, and GIS technology to analyze the diffusion of knowledge through a literature review. Compared to other studies (Table 3), which have used different forms of visualization, our study focuses on a deeper analysis of knowledge diffusion, which is lacking in previous studies. Most studies analyze knowledge diffusion through simulations or experiments, which introduce uncertainties, while only a few analyze real-life data. Despite differences, many studies agree on the crucial impact of social space and individual heterogeneity on the diffusion process, providing a solid theoretical foundation for our research.

The results show that Geodetector authors are located in 48 countries with China, the US, Australia, the UK, and others being the core countries. The authors' research covers various fields including geographic information, ecological assessment, life and health, spatial statistics, urban planning, remote sensing inversion, climate change, mathematical modeling, and human settlements. The academic output of authors varies between large and small teams. The relevance theory is used to understand the interdisciplinary application of knowledge, leading to the construction of a knowledge diffusion chain model to show multi-level cooperative relationships between authors.

However, this study only analyzed citations in the core database due to data limitations. To fully reflect the knowledge diffusion process in the citation database, future studies could incorporate a web crawler to obtain all citation records. Our study focuses on the Geodetector case and the academic groups and research stages are representative of a small sample of scholars. In the future, constructing an author evolution tree model based on a macroscopic topic can provide more generalized patterns. The use of multi-level modeling and Machine Learning algorithms can also help predict the academic activity of scholars in the next stage, which is a potential future research direction [46,63]. The findings in this study promote interdisciplinary collaborations to provide a broader perspective and understanding of knowledge diffusion.

Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

Data availability statement

Data will be made available on request.

Table 3
The summary ofiew summaof research on knowledge diffusion and literature review.

References	Research content	Research data	Methods	Visualization
Chen 2006; White 2007; Wei 2015; Lee 2016; Li 2016; Xiao 2017; Zhu 2017; Zhang 2018; Ma 2019	Literature review	Scientific literature	Bibliometric analysis	Network
Duan 2020			Bibliometric analysis and evolution tree methods	Network and evolution tree
White 2010			Relevance theory and bibliometric analysis	Network
Hu 2017; Hu 2018; Yuting 2019	Research hotspots		Bibliometric analysis and GIS technology	Network and map
Cowan 2004	Knowledge diffusion	Simulation data	Network model	Network
Kim 2009		R&D data.		
Pinto 2019		Patent data		
Kiss 2010		Scientific literature	Epidemic model	Curve Chart
Wang 2017; Wang 2019a; Yang 2015; Zheng 2019		Simulation data	Knowledge transmission model	
Our research	Knowledge diffusion and literature review	Scientific literature	Bibliometric analysis, relevance theory, and GIS technology	Network, heatmap, evolution tree, and transmission chain

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by the National Natural Science Foundation of China (Grant No. 42071375), the Innovation Project of State Key Laboratory of Resources and Environmental Information System (Grant No. O88RA205YA), and the China Scholarship Council (Grant No. 202204910278).

References

[1] J. Wang, Xu C. Geodetector, Principle and prospective, Acta Geograph. Sin. 72 (1) (2017) 116–134. PubMed PMID: CSCD:5902670.
[2] National Cancer Center CHCAoMS, China Cancer Atlas 2018, Sinomap press, Beijing, 2019 April, p. 282.
[3] L. Liu, D. Zhao, J. Wei, Q. Zhuang, X. Gao, Y. Zhu, et al., Permafrost sensitivity to global warming of 1.5 °C and 2 °C in the Northern Hemisphere, Environ. Res. Lett. 16 (3) (2021), 034038, https://doi.org/10.1088/1748-9326/abd6a8.
[4] L. Yang, X. Li, Q. Yang, L. Zhang, S. Zhang, S. Wu, et al., Extracting knowledge from legacy maps to delineate eco-geographical regions, Int. J. Geogr. Inf. Sci. (2021), https://doi.org/10.1080/13658816.2020.1806284.
[5] T. Chen, J. Xia, L. Zou, S. Hong, Quantifying the influences of natural factors and human activities on NDVI changes in the hanjiang river basin, China, Rem. Sens. 12 (22) (2020) 3780, https://doi.org/10.3390/rs12223780.
[6] Z. Fan, J. Duan, Y. Lu, W. Zou, W. Lan, A Geographical Detector Study on Factors Influencing Urban Park Use in Nanjing, China, Urban Forestry & Urban Greening, 2021, 126996, https://doi.org/10.1016/j.ufug.2021.126996.
[7] J. Li, A. Lu, C. Xu, Y. Li, M. Chen, Spatial heterogeneity and changes of population on both sides of Hu huanyong line, Acta Geograph. 72 (1) (2017) 148–160.
[8] L. Zhou, C. Zhou, F. Yang, B. Wang, D. Sun, Analysis of spatiotemporal evolution characteristics and driving factors of PM2.5 in China from 2000 to 2011, Acta Geograph. 72 (11) (2017) 2079–2092.
[9] D.C. Marvin, G.P. Asner, D.E. Knapp, Amazonian landscapes and the bias in field studies of forest structure and biomass, Proc. Natl. Acad. Sci. U.S.A. 111 (48) (2014), https://doi.org/10.1073/pnas.1412999111.
[10] D.C. Marvin, G.P. Asner, Spatially explicit analysis of field inventories for national forest carbon monitoring, Carbon Bal. Manag. 11 (1) (2016) 1–12, https://doi.org/10.1186/s13021-016-0050-0.
[11] D. Zhan, M.-P. Kwan, W. Zhang, S. Wang, J. Yu, Spatiotemporal variations and driving factors of air pollution in China, Int. J. Environ. Res. Publ. Health 14 (12) (2017) 1538, https://doi.org/10.3390/ijerph14121538.
[12] B. Chen, Y. Song, M.-P. Kwan, B. Huang, B. Xu, How do people in different places experience different levels of air pollution? Using worldwide Chinese as a lens, Environ. Pollut. 238 (2018) 874–883, https://doi.org/10.1016/j.envpol.2018.03.093.
[13] D. Zhan, M. Kwan, W. Zhang, J. Fan, J. Yu, Y. Dang, Assessment and determinants of satisfaction with urban livability in China(Article), Cities 79 (2018) 92–101, https://doi.org/10.1016/j.cities.2018.02.025.
[14] D. Zhan, M. Kwan, W. Zhang, X. Yu, B. Meng, Q. Liu, The driving factors of air quality index in China, J. Clean. Prod. 197 (2018) 1342–1351, https://doi.org/10.1016/j.jclepro.2018.06.108.
[15] Q. Wang, M.-P. Kwan, J. Fan, K. Zhou, Y.-F. Wang, A study on the spatial distribution of the renewable energy industries in China and their driving factors, Renew. Energy: Int. J. 139 (2019) 161–175, https://doi.org/10.1016/j.renene.2019.02.063.
[16] K. Fan, Q. Zhang, J. Li, D. Chen, C.-Y. Xu, Future surface soil moisture trends in China under global warming, Environ. Res. Lett. 16 (2021), https://doi.org/10.1088/1748-9326/abde5e.
[17] P. Broadbridge, A.D. Kolesnik, N. Leonenko, A. Olenko, D. Omari, Spherically restricted random hyperbolic diffusion, Entropy 22 (2) (2020) 217, https://doi.org/10.3390/e22020217.
[18] L. Fattorini, M. Marcheselli, C. Pisani, L. Pratelli, Design-based consistency of the Horvitz-Thompson estimator under spatial sampling with applications to environmental surveys, Spat Stat 35 (2020), 100404, https://doi.org/10.1016/j.spasta.2019.100404.
[19] C.M. Chen, CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature, J. Am. Soc. Inf. Sci. Technol. 57 (3) (2006) 359–377, https://doi.org/10.1002/asi.20317. PubMed PMID: WOS:000234932600008.
[20] H.D. White, Combining bibliometrics, information retrieval, and relevance theory, Part 1: first examples of a synthesis, J. Am. Soc. Inf. Sci. Technol. 58 (4) (2007) 536–559.
[21] F. Wei, T.H. Grubesic, B.W. Bishop, Exploring the GIS knowledge domain using CiteSpace, Prof. Geogr. 67 (3) (2015) 374–384, https://doi.org/10.1080/00330124.2014.983588.
[22] Y.-C. Lee, C. Chen, X.-T. Tsai, Visualizing the knowledge domain of nanoparticle drug delivery technologies: a scientometric review, Applied Sciences-Basel. 6 (1) (2016), https://doi.org/10.3390/app6010011. PubMed PMID: WOS:000372148500006.
[23] J. Li, C.M. Chen, CiteSpace Technology Text Mining and Visualization, Capital University of Economics and Business Press, Beijing, 2016.
[24] F. Xiao, C. Li, J. Sun, L. Zhang, Knowledge domain and emerging trends in organic photovoltaic technology: a scientometric review based on CiteSpace analysis, Front. Chem. 5 (2017), https://doi.org/10.3389/fchem.2017.00067. PubMed PMID: WOS:000411517100002.
[25] J. Zhu, W. Hua, Visualizing the knowledge domain of sustainable development research between 1987 and 2015: a bibliometric analysis, Scientometrics 110 (2) (2017) 893–914, https://doi.org/10.1007/s11192-016-2187-8.
[26] Q. Zhang, H. Xue, H. Tang, Knowledge domain and emerging trends in vulnerability assessment in the context of climate change: a bibliometric analysis (1991-2017), Knowl. Organ. 45 (6) (2018) 467–483, https://doi.org/10.5771/0943-7444-2018-6-467.
[27] X.2 Ma, L. Zhang, J. Wang, Y. Luo, Knowledge domain and emerging trends on echinococcosis research: a scientometric analysis, Int. J. Environ. Res. Publ. Health 16 (5) (2019) 842, https://doi.org/10.3390/ijerph16050842.
[28] H. White, Some new tests of relevance theory in information science [expanded version], Scientometrics 83 (2010) 653–667, https://doi.org/10.1007/s11192-009-0138-3.
[29] Y. Hu, Y. Han, Y. Zhang, Z. Yuan, Extraction and dynamic spatial-temporal changes of grassland deterioration research hot regions in China, Journal of Resources and Ecology 8 (4) (2017) 352–358.
[30] Y. Hu, Y. Han, Y. Zhang, Information extraction and spatial distribution of research hot regions on rocky desertification in China, Appl. Sci. 8 (11) (2018) 2075, https://doi.org/10.3390/app8112075. PubMed PMID:.
[31] L. Yuting, H. Yunfeng, H. Yueqi, Spatial distribution and dynamic changes in research hotspots for desertification in China based on big data from CNKI, Journal of Resources and Ecology 10 (6) (2019) 692–703, 12.
[32] C. Duan, S. Zuo, Z. Wu, Y. Qiu, J. Wang, Y. Lei, et al., A review of research hotspots and trends in biogenic volatile organic compounds (BVOCs) emissions combining bibliometrics with evolution tree methods, Environ. Res. Lett. 16 (1) (2020), 013003, https://doi.org/10.1088/1748-9326/abcee9.
[33] R. Cowan, Robin, Network Models of Innovation and Knowledge Diffusion, 2004.
[34] S. Breschi, F. Lissoni, Mobility and Social Networks: Localised Knowledge Spillovers Revisited, 2003.

[35] B. Allen, Some stochastic processes of interdependent demand and technological diffusion of an innovation exhibiting externalities among adopters, Int. Econ. Rev. (1982) 595–608.
[36] R. Cowan, N. Jonard, Network structure and the diffusion of knowledge, J. Econ. Dynam. Control 28 (8) (2004) 1557–1575.
[37] H. Kim, Y. Park, Structural effects of R&D collaboration network on knowledge diffusion performance, Expert Syst. Appl. 36 (5) (2009) 8986–8992.
[38] P.E. Pinto, A. Vallone, G. Honores, The structure of collaboration networks: findings from three decades of co-invention patents in Chile, Journal of Informetrics 13 (4) (2019), 100984, https://doi.org/10.1016/j.joi.2019.100984.
[39] G.-Y. Yang, Z.-L. Hu, J.-G. Liu, Knowledge diffusion in the collaboration hypernetwork, Phys. Stat. Mech. Appl. 419 (2015) 429–436, https://doi.org/10.1016/j.physa.2014.10.012.
[40] P.-A. Balland, D. Rigby, The geography of complex knowledge, Econ. Geogr. 93 (1) (2017) 1–23.
[41] I.Z. Kiss, M. Broom, P.G. Craze, I. Rafols, Can epidemic models describe the diffusion of topics across disciplines? Journal of Informetrics 4 (1) (2010) 74–82, https://doi.org/10.1016/j.joi.2009.08.002.
[42] S.-G. Liao, S.-P. Yi, Modeling and dynamic analysis of knowledge transmission process: a model considering individual perception of knowledge value, Commun. Nonlinear Sci. Numer. Simulat. 95 (2021), 105598.
[43] H. Wang, J. Wang, L. Ding, W. Wei, Knowledge transmission model with consideration of self-learning mechanism in complex networks, Appl. Math. Comput. 304 (2017) 83–92, https://doi.org/10.1016/j.amc.2017.01.020.
[44] W. Zheng, H. Pan, C. Sun, A friendship-based altruistic incentive knowledge diffusion model in social networks, Inf. Sci. 491 (2019) 138–150.
[45] H. Wang, J. Wang, M. Small, J.M. Moore, Review mechanism promotes knowledge transmission in complex networks, Appl. Math. Comput. 340 (2019) 113–125.
[46] J.F. Wang, X.H. Liu, L. Peng, H.Y. Chen, L. Driskell, X.Y. Zheng, Cities evolution tree and applications to predicting urban growth, Popul. Environ. 33 (2–3) (2012) 186–201, https://doi.org/10.1007/s11111-011-0142-4. PubMed PMID: WOS:000304167400005.
[47] Y. Wang, F. Yuan, Comprehensive evaluation of urban compactness in the Yangtze River Delta based on evolutionary tree model, Yangtze River Basin Resources and Environment 23 (6) (2014) 741–750.
[48] Y.-L. Liao, J.-F. Wang, G. Chen, W. Du, X.-M. Song, X. Yun, et al., Clustering of disability caused by unintentional injury among 15-to 60-year-old: a challenge in rapidly developing countries, Geospatial Health 8 (1) (2013) 13–22, https://doi.org/10.4081/gh.2013.50. PubMed PMID: WOS:000330210500002.
[49] W. Zhang, X. Duan, W. Zhang, Measurement and evolution analysis of sustainable development in the yangtze river delta region, Yangtze River Basin Resources and Environment 22 (10) (2013) 1243–1249.
[50] C. Wang, H. Liu, Regional differentiation and evolution characteristics of foreign direct investment in Shandong Province, Regional research and development 35 (1) (2016) 70–75.
[51] C. He, L. Li, H. Zhou, P. Huang, Z. Yang, Spatial-temporal evolution and driving factor analysis of urbanization in the Dadu River Basin, Journal of Southwest Normal University (Natural Science Edition) 41 (10) (2016) 54–60.
[52] Y. Hu, L. Peng, X. Li, X.J. Yao, H. Lin, T.H. Chi, A novel evolution tree for analyzing the global energy consumption structure, Energy 147 (2018) 1177–1187, https://doi.org/10.1016/j.energy.2018.01.093. PubMed PMID: WOS:000429391100089.
[53] Y. Wang, J. Wang, Modelling and prediction of global non-communicable diseases, BMC Publ. Health 20 (1) (2020) 822, https://doi.org/10.1186/s12889-020-08890-4.
[54] J.F. Wang, X.H. Li, G. Christakos, Y.L. Liao, T. Zhang, X. Gu, et al., Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China, Int. J. Geogr. Inf. Sci. 24 (1) (2010) 107–127, https://doi.org/10.1080/13658810802443457.
[55] J.F. Wang, T.L. Zhang, B.J. Fu, A measure of spatial stratified heterogeneity, Ecol. Indicat. 67 (2016) 250–256, https://doi.org/10.1016/j.ecolind.2016.02.052. PubMed PMID: WOS:000388785300025.
[56] C.M. Chen, Searching for intellectual turning points: progressive knowledge domain visualization, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 5303–5310, https://doi.org/10.1073/pnas.0307513100. PubMed PMID: WOS:000220823000020.
[57] L.C. Freeman, Centrality in social networks conceptual clarification, Soc. Network. 1 (3) (1978) 215–239.
[58] X. Liu, Simulation Analysis of the Driving Force of Land Use Change in China, 2005.
[59] J. Wang, Y. Liao, X. Liu, Spatial Data Analysis Tutorial, second ed., Science Press, Beijing, 2019.
[60] W. Luo, J. Jasiewicz, T. Stepinski, J. Wang, C. Xu, X. Cang, Spatial association between dissection density and environmental factors over the entire conterminous United States, Geophys. Res. Lett. 43 (2) (2016) 692–700.
[61] H. Ju, Z. Zhang, L. Zuo, J. Wang, S. Zhang, X. Wang, et al., Driving forces and their interactions of built-up land expansion based on the geographical detector–a case study of Beijing, China, Int. J. Geogr. Inf. Sci. 30 (11) (2016) 2188–2207.
[62] D. Sperber, D. Wilson, Relevance: Communication and Cognition, Citeseer, 1986.
[63] M. Ishfaque, Q. Dai, Haq Nu, K. Jadoon, S.M. Shahzad, H.T. Janjuhah, Use of recurrent neural network with long short-term memory for seepage prediction at tarbela dam, KP, Pakistan, Energies 15 (9) (2022) 3123, https://doi.org/10.3390/en15093123. PubMed PMID:.