Spatial Distribution Patterns of Research Output in Academic Disciplines: A Case Study of Computer Software and Applications (Postprint)

Authors: Ma Chao, Li Gang, Mao Jin, Gu Yansong

Date: 2023-07-26T00:00:00+00:00

Abstract

[Objective/Significance] This study investigates the spatial distribution of research output in disciplinary fields, identifies whether spatial autocorrelation exists in its distribution, and analyzes the influencing factors that contribute to spatial clustering.

[Methods/Procedures] Based on bibliographic data from core journals in the computer science field from 1997-2016, institutional geographic location information was extracted and used as observation points. Their distribution was visualized, and the Gini index and centrality index were calculated to investigate distribution concentration. Global Moran's I and local spatial autocorrelation indicators were calculated to study spatial autocorrelation patterns. Pearson correlation coefficients were calculated to examine correlations with the number of universities in the region, Gross National Product, research and experimental development investment, and number of industry practitioners.

[Results/Conclusions] Research output in core journals of the computer science field exhibits uneven geographical distribution with obvious concentration trends and spatial autocorrelation. It also shows high correlation with the number of high-level universities in the discipline in the region, the number of industry practitioners, and the number of scientific researchers.

Full Text

Preamble

Vol. 63 No. 15, August 2019
ChinaXiv Partner Journal

Research on the Spatial Distribution Patterns of Scientific Research

Output by Discipline: A Case Study of Computer Software and Applications

Ma Chao, Li Gang, Mao Jin, Gu Yansong
Center for Studies of Information Resources, Wuhan University, Wuhan 430072

Abstract

[**Purpose/Significance**] This study investigates the spatial distribution of scientific research output within a discipline, explores whether spatial autocorrelation exists in its distribution patterns, and analyzes the influencing factors that contribute to spatial clustering.

[**Method/Process**] Based on bibliographic data from core journals in the computer science field spanning 1997–2016, we extracted institutional geographic location information as observation points to visualize their distribution patterns. We calculated the Gini index and centrality index to examine distribution concentration, measured global Moran's I and local autocorrelation indicators to study spatial autocorrelation patterns, and computed Pearson correlation coefficients to analyze correlations with regional university counts, gross domestic product, research and experimental development (R&D) investment, and industry practitioner numbers.

[**Result/Conclusion**] Research output in the computer science field exhibits uneven geographic distribution with clear concentration tendencies and spatial autocorrelation. The output shows strong correlations with the number of high-level universities in the region, industry practitioner numbers, and scientific research personnel numbers.

Classification Number: G322
Keywords: spatial knowledge spillover, spatial autocorrelation, urban innovation, exploratory spatial data analysis

Introduction

With the innovation of research methods and continuous development of interdisciplinary studies, research institutions and universities have achieved improving scientific research capabilities, with both the quantity and quality of research outputs rising steadily. However, due to regional development imbalances, significant disparities exist between different regions, prompting numerous studies on the spatial distribution patterns of research output. For instance, L. Bornmann et al. constructed density maps to detect "hot regions" of scientific output in SCI [?]; C.W. Matthiesson et al. analyzed European research output based on article counts in SCI [?]; and P. Zhou et al. examined provincial research output in China using SCI data [?]. Visualizing research output on geographic maps has fully revealed its distribution and concentration patterns. Concurrently, cross-institutional and cross-regional collaboration has been widely adopted, with some scholars visualizing cooperation networks on maps [?]. However, does geographic location merely influence cooperation networks, or does it indirectly affect other factors that shape research output distribution? According to Tobler's First Law of Geography [?], everything is related, but near things are more related than distant things. This proximity effect often manifests as geographically close research institutions engaging in more frequent collaboration, exhibiting spatial clustering. Does this geographic influence only affect cooperation, or does it have broader impacts? This clustering and correlation constitute spatial autocorrelation [?]. Due to the ubiquity of spatial autocorrelation, analyzing research output from a geographic perspective holds significant research value. Moreover, spatial clustering of research output may also be influenced by regional development levels, and such influence is mutual—concentrated research output can affect regional economic development and innovation capacity, a phenomenon often summarized as spatial knowledge spillover [?]. What factors drive these effects and to what extent they influence outcomes are also valuable research questions.

Previous studies have primarily focused on internal factors of the academic system as main influences. For example, M. Mabe demonstrated a clear positive correlation between researcher numbers and research output [?]. Geographic location is often treated as an objective condition but is less frequently considered as a direct influencing factor. This study examines the spatial distribution of research output in computer software and applications, revealing not only the spatial distribution patterns but also analyzing external causes of uneven distribution, which holds significant meaning for identifying commonalities in regional research development and studying regional research directions. Based on bibliographic data from core journals in computer software and applications from 1997–2016, we employ statistical analysis and exploratory spatial data analysis (ESDA) methods combined with geographic visualization to explore spatiotemporal distribution patterns, spatial autocorrelation, and influencing factors.

1. Related Research

1.1 Spatial Scientometrics

The earliest application of spatial analysis methods to research output studies emerged in the 1970s, but the field did not develop substantially until recent years. In 2009, K. Frenken et al. defined the application of spatial analysis to scientometrics as "spatial scientometrics" [?], categorizing its research into three areas: spatial distribution, spatial biases, and citation impact. Studies on spatial distribution have primarily focused on the national level, with fewer in-depth regional-level studies due to challenges in obtaining geographic location information and precision. Research methods mainly involve visualization and descriptive analysis. Spatial bias research concentrates on the spatial clustering of research output institutions and its underlying causes, which can be explained from at least three perspectives: (1) geographically proximate research entities have higher probabilities of intersection; (2) spatial distance correlates positively with face-to-face communication costs (e.g., transportation expenses, travel time); and (3) spatially close entities share more similar "research styles." Spatial distribution and spatial biases constitute the main research themes in spatial scientometrics.

1.2 Factors Influencing Urban Research Output

While few studies specifically address factors influencing urban research output, research on urban innovation capacity provides a foundation due to the spatial spillover effects of research output. Papers and patents, as important forms of research output, are considered sources of urban innovation and competitiveness [?] and are frequently used as key indicators for evaluating urban innovation [?, ?]. Many factors contribute to differences in urban innovation capacity. For instance, Cao Yong et al. used urban economic development level, foreign exchange capacity, research input level, and achievement transformation capacity as primary indicators to compare innovation capacity differences among China's four municipalities, finding that R&D resource input was the most critical factor [?]. Numerous scholars have analyzed the relationship between research input and output from various perspectives. Shi Xinxiang et al. examined the relationship between R&D resources and research output based on R&D resource census data [?], while Yang Jun and Wu Yang explored the relationship between university research input and output [?, ?]. As universities serve as primary research hubs, affiliated research institutes, laboratories, and centers critically influence urban research output, making university counts important influencing factors.

Previous research has mainly focused on traditional bibliometric methods and analysis of research cooperation networks, with significance lying in revealing internal network patterns and causes. However, spatial statistical analysis from a spatial econometrics perspective, particularly regarding external influencing factors based on knowledge spillover phenomena, remains relatively underexplored. Building on prior research, this study focuses on two aspects: (1) visualizing and analyzing the spatial distribution of research output in a specific discipline to reveal its distribution patterns; and (2) comparing and analyzing the formation causes of concentrated research output regions and related factors.

2. Research Methods

2.1 Data Sources

Our experimental data were obtained from CNKI (China National Knowledge Infrastructure). First, we used the journal search function to select core journals under the computer software and applications category, including 10 journals such as *Journal of Software*. We then exported complete bibliographic information containing institutional affiliations from these journals for the period 1997–2017, obtaining 78,112 records. After removing 1,954 invalid entries (e.g., calls for papers, conference announcements), we retained 76,113 valid records containing title, author, institution, keywords, and abstract information.

2.2 Institution and Address Parsing

A key challenge in data cleaning was inconsistent data format standards. For example, 1997 export files contained only full institution names separated by commas, while 2002 files stored institution information in the format "full institution name, city, postal code" separated by exclamation marks. After 2007, semicolons were used as separators. Additionally, institution field formats varied across journals and years, creating significant obstacles for batch processing. We addressed this using string replacement and extraction combined with regular expression matching to extract complete institution names from raw data.

Data processing involved two main tasks: (1) extracting top-level institution names, and (2) address parsing. The specific workflow is shown in Figure 1 [Figure 1: see original paper].

Stage 1 used keyword matching combined with string truncation to extract top-level institution names, yielding 162,716 entries.

Stage 2, the critical phase, aimed to obtain institutional geographic coordinates and administrative divisions (county, prefecture, and province levels) for subsequent analysis. First, we used the Amap API's geocoding function to submit requests with full institution names. After parsing returned JSON data, we extracted province, city, longitude, and latitude information, storing it in a database. Successful parsing results were also stored in a Redis in-memory database for future use. Due to unresolvable institution names, we ultimately obtained 159,910 valid data entries.

Two key issues emerged during data processing:

1. **Inconsistent institutional hierarchy**: Institution name fields primarily included universities, institutes, companies, and corporations. For universities, some entries specified schools/colleges while others only provided university names. We first matched "university" keywords to extract top-level names (e.g., Wuhan University), then matched "college" keywords for remaining data to avoid ambiguity with names like "XX University XX College." Finally, we matched remaining keywords like "school" and "company."

2. **Insufficient address parsing capability**: This stemmed from three causes: (a) large time span with inaccurate institution names due to name changes, mergers, or closures; (b) overly granular institution names (e.g., "XX Laboratory," "XX Center") that parsed poorly; and (c) multiple non-standard names for the same institution (e.g., "University of Science and Technology of China" also appears as "USTC," "China UST," etc.) [**?**]. We addressed this through manual calibration, using Baidu Baike and search engines to identify current names and locations for unresolvable entries, then replacing old names with new ones. During address parsing, we submitted both institution name and city as parameters to improve resolution. For multi-name issues, we built a data dictionary for incremental

matching alongside manual calibration to improve efficiency.

2.3 Spatial Analysis Methods for Research Output

We examined the spatiotemporal evolution of research output, focusing on spatial distribution and changes within the same time period and across time periods. Python was used for data processing and basic analysis, while the open-source spatiotemporal analysis tool GeoDa was employed for exploratory spatial analysis.

2.3.1 Concentration Analysis D.B. Audretsch et al. used a location-based Gini Index to measure spatial concentration and calculated a centrality index (CR3) [**?**].

The location-based Gini Index is calculated as:

$$G = \frac{1}{2n^2\bar{Z}} \sum_{i=1}^{n} \sum_{j=1}^{n} |z_i - z_j|$$

where n represents the total number of regions containing research output institutions, $\bar{Z}$ is the overall mean research output across regions, z_i represents research output in region i, and j is similar to i. Different observation points can be selected for regional division at provincial, prefecture, county, or institutional levels.

The centrality index is calculated as:

$$CR3 = \sum_{i=1}^{3} w_i$$

where w_i represents the proportion of research output from the top three regions divided by total output within the same period.

2.3.2 Spatial Correlation Analysis In the exploratory spatial analysis stage, we conducted basic spatial statistics and spatial autocorrelation analysis. Spatial autocorrelation refers to the potential interdependence among observations of a given variable within the same distribution area, including global and local spatial autocorrelation. Global spatial autocorrelation identifies overall patterns, while local spatial autocorrelation reveals specific spatial relationships [**?**].

Moran's Index is commonly used to describe global spatial autocorrelation:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W(i,j)(x_i - \bar{x})(x_j - \bar{x})}{S \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where N is the number of papers produced across all regions at time t, $W(i,j)$ is the spatial weight matrix, $S = \sum_{i=1}^{n}\sum_{j=1}^{n} W(i,j)$ is the sum of all elements in the spatial weight matrix, and x_i and x_j represent paper output in regions i and j, respectively. Moran's Index ranges from -1 to 1, where 1 indicates positive spatial autocorrelation, -1 indicates negative spatial autocorrelation, and 0 indicates no spatial autocorrelation.

Unlike global Moran's Index, local spatial autocorrelation requires a series of indicators collectively known as Local Indicators of Spatial Association (LISA). The most critical factor in exploratory spatial analysis is spatial weight determination, which significantly affects results. GeoDa supports multiple spatial weight matrix $W(i,j)$ definitions based on adjacency relationships and distance. Since Chinese administrative divisions are not based on latitude/longitude and vary greatly in area, distance-based weight definitions produce large errors. Therefore, we selected adjacency-based spatial weight definitions.

2.3.3 Analysis of Research Output Influencing Factors Combining spatial correlation analysis results, we comprehensively discussed regional factors influencing research output and conducted correlation analysis. From a spatial knowledge spillover perspective, main influencing factors include regional university counts, gross domestic product (GDP), population, and R&D investment. For correlation measurement, we primarily calculated Pearson Correlation Coefficients:

$$r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right)$$

where $\bar{x}$ and σ_x are the mean and standard deviation of influencing factors, respectively.

3. Results Analysis

3.1 Geographic Distribution of Research Output

Figure 2 [Figure 2: see original paper] shows the distribution of research output (papers) by province from 1997–2016:

Overall, the Yangtze River Delta region, Beijing area, and Central China formed high-output zones for journal papers in this field, while Northwest China lagged behind. The distribution shows clear imbalance, with research output concentrated in North China (Beijing), the Yangtze River Delta, and Central China. The provinces (including autonomous regions and municipalities) with highest output were Beijing (20.54%), Shaanxi (9.63%), and Jiangsu (9.15%). The lowest were Tibet Autonomous Region, Qinghai Province, and Hainan Province, each accounting for less than 0.05%.

Observations at the provincial level provide an overview, but to further examine internal concentration, we refined the observation granularity. Using latitude and longitude, we constructed a heatmap to observe distribution patterns. Figure 3 [Figure 3: see original paper] shows the heatmap of research output distribution across the 20-year period:

Compared with provincial observations, regional concentration becomes more intuitive. Aggregating by city, Beijing remained the highest-output city. The top 10 cities were all municipalities or provincial capitals: Beijing (20.54%), Xi'an (7.80%), Shanghai (6.32%), Chengdu (4.19%), Nanjing (2.89%), Chongqing (2.84%), Wuhan (2.03%), Tianjin (1.70%), Guangzhou (1.68%), and Harbin (1.67%). From 12th place onward, non-provincial capitals like Dalian (1.49%) and Wuxi (0.59%) appeared, though most non-capitals accounted for less than 1%.

We also measured concentration quantitatively by calculating annual Gini and centrality indices (Figure 4 [Figure 4: see original paper]):

The Gini index remained above 0.7 throughout, with a minimum of 0.744 in 1997 and maximum of 0.866 in 2005. Scholars analyzing reasonable Gini coefficient values suggest 0.33 as theoretically optimal, with values above 0.5 indicating imbalance. The centrality index (CR3) showed similar patterns—large fluctuations in the first five years followed by a declining and relatively stable trend. Both indices demonstrate spatial imbalance in journal paper output.

3.2 Spatial Autocorrelation Analysis of Research Output

Spatial distribution patterns and clustering can be studied through spatial autocorrelation. Global Moran's Index reveals whether overall distribution is dispersed, random, or clustered. For deeper analysis, local spatial autocorrelation measurement is required. Unlike global Moran's Index, local spatial autocorrelation uses LISA indicators [?], which can be visually represented through Moran scatter plots, cluster maps, and significance maps. Using prefecture-level cities as observation points, we measured spatial autocorrelation and integrated results at the municipal level (Figure 5 [Figure 5: see original paper]).

3.2.1 Global Spatial Autocorrelation Results Figure 5 shows that global Moran's Index was greater than 0 from 1997–2016, indicating significant spatial autocorrelation in journal paper output that increased overall over time. Specifically, from 1997–2001, global Moran's Index was approximately -0.00072, very close to 0, indicating weak global correlation. In the late 20th century, limited by transportation and communication, inter-institutional collaboration was not widespread, confined to internal cooperation or independent output, resulting in relatively dispersed distribution. In subsequent periods, global Moran's Index gradually increased, reaching approximately 0.168 in 2012–2016, showing clear autocorrelation. As collaborative research developed, institutional connections became closer, creating obvious spatial clustering effects. Additionally, some data points in the scatter plot were highly influential and dispersed, further illustrating clustering characteristics.

3.2.2 Local Spatial Autocorrelation Results The quadrant distribution of local Moran's Index represents four types of spatial autocorrelation relationships. For observation values like paper counts, first-quadrant points indicate high-value areas surrounded by high-value areas (high-high clustering), while third-quadrant points indicate low-low clustering. Both correlate positively with global Moran's Index. Second and fourth quadrants represent high-low and low-high clustering, respectively, correlating negatively with global Moran's Index.

Using prefecture-level cities as observation points, we identified 66 high-high clustering regions, 25 low-low clustering regions, 69 low-high clustering regions, and 7 high-low clustering regions. Table 1 shows local Moran's Index spatial clustering patterns for these regions:

Table 1. Spatial Clustering Characteristics of Research Output

Spatial Autocorrelation	Regions
High-High Clustering	Beijing, Tianjin, Shanghai, Chongqing, Wuhan, Hangzhou, Changsha, Xi'an, Hefei, Chengdu, Jinan, Shenyang, Guangzhou, Harbin, Changchun, Nanjing, Xiangtan, Dalian, Qingdao, Xuzhou, Yantai, Macao SAR
Low-High Clustering	Chengde, Xianyang, Tai'an, Suzhou, Wuxi, Zhengzhou, Ezhou, Taiyuan, Ningbo, Huainan, Zhenjiang, Zhangjiakou, Yangzhou
High-Low Clustering	Guangyuan, Shantou, Qinzhou, Ganzhou

Spatial Autocorrelation	Regions
Low-Low Clustering	Heihe, Baiyin, Xing'an League, Qiandongnan Miao and Dong Autonomous Prefecture, Xingtai, Hulunbuir, Luoyang, Datong, Liupanshui, Changzhi, Kashgar, Bayingolin Mongol Autonomous Prefecture, Jixi, Huangshan, Tongren, Shannan, Longyan, Handan, Xinxiang, Tacheng, Hegang

The local Moran's Index scatter plot shows each point representing a municipal administrative region. During analysis, we found spatial autocorrelation less obvious at the prefecture level due to large inter-city distances and weak research activity connections, compounded by our adjacency-based spatial weight definition. High-value clustering trends in local spatial autocorrelation align with overall analysis results, with high-output cities showing internal spatial clustering. Geographic proximity (adjacency) between regions creates mutual influence on research output, forming corresponding high-value clusters closely related to inter-institutional collaboration. Do these 66 high-high clustering regions share common characteristics? Why do such clusters form? We analyze this from the perspective of regional influencing factors.

3.3 Analysis of Research Output Influencing Factors

Building on spatial knowledge spillover theory, we know regional research output correlates positively with economic development levels. Therefore, primary considerations include regional economic development level (measured by GDP), technological resource investment [?], and university counts (since institutions of higher education account for over 80% of our data). We also considered regional population, including: (1) information transmission, computer services, and software industry practitioners; and (2) scientific research and technical personnel. Data from the National Bureau of Statistics provide strong reference value.

Using provincial administrative units as observation points, we analyzed correlations between regional GDP, R&D expenditure, university counts, industry practitioner numbers, scientific research personnel numbers, and regional journal paper output (2012–2016). Results are shown in Figure 6 [Figure 6: see original paper]:

All five factors showed significant positive correlations with regional journal paper output. The most influential factor was scientific research personnel numbers, with average correlation coefficients above 0.87, fully reflecting researchers' primary role in paper output. Industry practitioner correlations were second, with average coefficients above 0.85. Although not directly associated with journal paper output, industry patents and copyrights are closely linked to research, demonstrating how research drives regional industry development from a science-to-technology transfer perspective. Regional GDP and R&D expenditure also showed high correlations, reflecting regional economic development levels and development positioning, confirming the interaction between urban innovation investment and research output.

Surprisingly, while universities serve as research hubs, the correlation between regional university counts and journal paper output was only around 0.5—disappointingly low. The main reason is that examining only macro-level external factors like absolute university numbers provides too coarse a perspective. Universities have multiple specialties with varying strengths and development directions across fields, and their research output capacities differ by nature. Therefore, measuring university impact on journal output solely by quantity is inadequate.

For deeper analysis specific to computer software and applications, we conducted a more comprehensive investigation combining the 2017 national fourth-round discipline evaluation results. We focused on computer science and technology and software engineering disciplines, conducting statistical analysis combined with regional university distribution. To highlight discipline-level effects, we accumulated universities rated "B-" or above and performed correlation analysis. For comparison, we averaged other influencing factors. Final results are shown in Figure 7 [Figure 7: see original paper]:

The results confirm that regional university counts correlate highest with regional research output, but this refers not to absolute numbers but to universities with advantageous disciplines in the specific field.

4. Conclusion and Discussion

This study's main innovation lies in moving beyond traditional quantity-based bibliometric methods to reveal spatial distribution patterns of journal paper output in computer software and applications using spatial statistical analysis, combined with multi-factor correlation analysis.

4.1 Discussion of Results

(1) From 1997–2016, core journals in computer applications showed enormous regional contribution differences. Beijing, as China's capital and cultural center, holds an unshakable historical position in research activities represented by journal papers, accounting for over 20% of the sample. Xi'an, Shaanxi Province, unexpectedly became a high-output region, though it doesn't rank top in university counts, population, or GDP. Institution-level analysis revealed Northwestern Polytechnical University, Xi'an Jiaotong University, and Xidian University as major contributors. We also found military institutions account for approximately 10% of the sample, reflecting the importance of computer applications, particularly graphics and imaging, in national defense technology.

(2) During the study period, significant spatial clustering effects and autocorrelation existed between regions, showing an upward trend with clear high-value clustering in municipalities and some provincial capitals. Two main factors drive this concentration: (a) the regional status of municipalities and capitals ensures leadership in economic development and other aspects; and (b) as research collaboration deepens, cross-institutional and cross-regional co-authorship becomes an effective way to improve research quality. During cross-regional collaboration, geographically proximate institutions remain preferred choices. This is not coincidental—R. Boschma et al. (2005) categorized similarities between collaborating institutions into five dimensions, demonstrating that proximity across these dimensions facilitates broader cooperation, with physical proximity being most important [**?**]. Although communication technology and transportation have reduced cross-regional collaboration costs, and we observed declining concentration in later study periods, geographic advantages still dominate inter-institutional collaboration, influencing spatial clustering of research output. Additionally, different spatial weight definitions affect autocorrelation analysis results.

(3) Regional research output is influenced by multiple factors. We demonstrated correlations with regional GDP, R&D expenditure, population, and university counts, all meeting expectations. Combining discipline evaluation results to measure university development levels in computer software and applications provided deeper insight into how regional university research capacity affects output. While research output efficiency is multifactorial, university counts proved most influential for journal papers, as universities concentrate graduate students and researchers, creating favorable research environments. Subjectively, universities also implement reward and evaluation mechanisms that motivate researchers to utilize resources efficiently and increase output—an important issue in higher education management research [**?**]. R&D expenditure, though crucial, showed lower correlation than university counts in this study, primarily because it mainly affects patent and copyright output, which have broader

sources beyond universities and research institutions. As urban competitiveness and innovation-oriented development progress, high-tech enterprises increasingly treat research output and intellectual property as key competitiveness components.

4.2 Limitations and Future Directions

This study's main limitations involve geographic data timeliness and error control in address parsing. The long time span introduced uncertainties from institutional name changes, reorganizations, and closures, reducing sample quality. Address parsing faced challenges with open map API resolution capabilities, yielding imprecise results. Additionally, this study did not incorporate weighting for cooperation relationships or author priority in paper output, which could be integrated in future research. From a depth perspective, improving precision and adding dimensions would enrich results and yield more valuable insights.

References

[1] BORGMANN L, WALTMAN L. The detection of "hot regions" in the geography of science—a visualization approach using density maps[J]. Journal of informetrics, 2011, 5(4): 547-553.
[2] MATTHIESSEN C W, SCHWARZ A W. Scientific centres in Europe: an analysis of research strength and patterns of specialisation based on bibliometric indicators[J]. Urban studies, 1999, 36(3): 453-477.
[3] ZHOU P, THIJS B, GLÄNZEL W. Regional analysis on Chinese scientific output[J]. Scientometrics, 2009, 81(3): 839-857.
[4] MABE M. The growth and number of journals[J]. Serials, 2003, 16(2): 191-197.
[5] Hu Changping, Hu Jiming. Geographic visualization of domain research collaboration networks: a case study of *Journal of Library Science in China*[J]. Information science, 2018, 36(2): 3-8.
[6] TOBLER W R. A computer movie simulating urban growth in the Detroit region[J]. Economic geography, 1970, 46(S1): 234-240.
[7] DORMANN C F, MCPHERSON J M, ARAÚJO M B, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review[J]. Ecography, 2007, 30(5): 609-628.
[8] BOTTAZZI L, PERI G. Innovation and spillovers in regions: evidence from European patent data[J]. European economic review, 2003, 47(4): 687-710.
[9] FRENKEN K, HARDEMANS S, HOEKMAN J. Spatial scientometrics: towards a cumulative research program[J]. Journal of informetrics, 2009, 3(3): 222-232.
[10] FELDMAN M P, AUDRETSCH D B. Innovation in cities: science-based diversity, specialization and localized competition[J]. European economic review, 1999, 43(2): 409-429.
[11] Li Huifen. Construction and empirical study of urban innovation indicator system[J]. Nanjing social sciences, 2010(7): 15-20.

[12] Zhang Xiekui, Wu Siyi. Urban innovation evaluation based on "element-structure-function-environment": a case study of 17 national innovative pilot cities[J]. Science & technology progress and policy, 2015, 32(2): 138-144.
[13] Cao Yong, Cao Xuanzhen, Luo Chuxuan, et al. Comparative study on innovation capacity and influencing factors of China's four municipalities[J]. China soft science, 2013(6): 162-170.
[14] Shi Xinxiang, Feng Li, Liang Tongying. Relationship between existing R&D resources and research output in China: an empirical study based on second national R&D resource census data[J]. Science research management, 2012, 33(10): 1-8.
[15] Yang Jun, Fu Lin, Lin Yiwen, et al. Analysis of research input and output in higher education institutions[J]. Science and technology management research, 2013, 33(16): 102-106.
[16] Wu Yang, He Guangrong, He Jinqiu. Correlation analysis of research input and output in higher education institutions: 1991-2008[J]. Tsinghua journal of education, 2011, 32(4): 104-112.
[17] University of Science and Technology of China. Official website of USTC[EB/OL]. [2018-11-01]. https://www.ustc.edu.cn.
[18] AUDRETSCH D B, FELDMAN M P. R&D spillovers and the geography of innovation and production[J]. The American economic review, 1996, 86(3): 630-640.
[19] WANG Y, HU R, LIU M. The geotemporal demographics of academic journals from 1950 to 2013 according to Ulrich's database[J]. Journal of informetrics, 2017, 11(3): 655-671.
[20] Hu Zuguang. Research on the theoretical optimal value and simple calculation formula of Gini coefficient[J]. Economic research, 2004, 9(1): 60-69.
[21] ANSELIN L. Local indicators of spatial association—LISA[J]. Geographical analysis, 1995, 27(2): 93-115.
[22] Cai Xiang, Cui Xiaolan, Xiong Jing, et al. Regional R&D efficiency and its influencing factors in China: based on the perspective of "research output-achievement transformation"[J]. Soft science, 2013, 27(3): 80-84.
[23] BOSCHMA R. Proximity and innovation: a critical assessment[J]. Regional studies, 2005, 39(1): 61-74.
[24] Li Jiazhe, Hu Yongmei. Review and prospects of research on domestic university research efficiency and productivity[J]. Modern education management, 2018(01): 54-61.

Author Contributions

Ma Chao: Designed research framework, programmed, wrote manuscript;
Li Gang: Proposed research ideas and methods;
Mao Jin: Designed paper structure, revised manuscript;
Gu Yansong: Data preprocessing.

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