

Article

Verification of Geographic Laws Hidden in Textual Space and Analysis of Spatial Interaction Patterns of Information Flow

Lin Liu, Hang Li *, Dongmei Pei and Shuai Liu

College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China

* Correspondence: lihang_11@126.com

Abstract: The rapid development of Internet technology has formed a huge virtual information space. In the information space, information flow has become a link of communication between objects. Information flow is an alternative or supplement to the traditional physical flow for the study of the spatial interaction of geographical entities. The research uses toponym co-occurrence and search index as information flow data, verifies the geographical laws hidden in the information space by spatial autocorrelation analysis and gravity model fitting, and analyzes the spatial interaction patterns of provinces in China in the information space by complex network analysis methods. The results show that: (1) information flow in the information space obeys Tobler's first law of geography and Goodchild's second law of geography. The spatial interaction represented by information flow has a distance decay effect. The best distance decay coefficients for toponym co-occurrence and the search index are 0.189 and 0.186, respectively. (2) The inter-provincial spatial interaction network of China shows a hierarchical pattern of the triangular primary network and diamond secondary network, and the ranking of provinces in the centrality analysis is basically stable, but the network hierarchy is deepening. The gravity center of spatial interaction is located in the east-central region of China. (3) The information flow-based interaction network is of higher asymmetry than the population mobility network, and its spatial structure is also obvious. This research provides a new idea for studying the spatial interaction of geographical entities in the physical world from the perspective of information flow.



Citation: Liu, L.; Li, H.; Pei, D.; Liu, S. Verification of Geographic Laws Hidden in Textual Space and Analysis of Spatial Interaction Patterns of Information Flow. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 217. <https://doi.org/10.3390/ijgi12060217>

Academic Editors: Wolfgang Kainz, Hangbin Wu and Tessio Novack

Received: 15 March 2023

Revised: 13 May 2023

Accepted: 24 May 2023

Published: 26 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: textual space; toponym co-occurrence; geospatial mapping; spatial interaction; complex network

1. Introduction

With the social–economic development and the increasing sophistication of science and technology, the popularizing rate of the Internet has been increasing year by year, and the Internet has become an important way for people to obtain information. Information elements break through the limitations of geographic space and form a huge virtual information space. The information space takes information elements as the main component and information flow as the connection between regions. Information flow is an excellent alternative to the relationship between entities in the real world, and it can reflect the spatial connection strength of geographic entities [1]. Toponym is a general term for natural or humanistic geographic entities that exist in a certain spatial location [2], and it provides valuable information for the research of disciplines such as geography. Toponym is used as geographic reference information in about 70% of the texts [3], which provides a basis for researching geography in the form of text. Toponym co-occurrence means multiple toponyms appear in the same text with a certain frequency, which generates informational ties between geographical entities. The search index is statistical data based on users' search behavior, which can reflect the attention of Internet users in a certain place at a certain time for other places [4]. In the information space, the geographic entity is indicated by the toponym text, and the connection or interaction between geographic entities occurs as the

information flow, such as toponym co-occurrence and search index. Studying the spatial interaction of geographical entities from the perspective of information flow is a convenient research method that can replace the traditional physical flow in the information age.

The connection and interaction between geographic regions are the important research contents of geography, and the urban system has been studied from the aspects of traffic flow [5], social relation [6], enterprise perspective [7], and economic flow [8]. From the perspective of information flow, toponym co-occurrence provides a new idea for researching the spatial interaction pattern of geographic entities. As a medium of information flow, the news contains many aspects, such as politics, economy, culture, and society. It is a real-time, convenient, and informative research way to mine geographic information through toponym co-occurrence from massive web news texts. Toponym co-occurrence can be used for extracting popular tourist destinations [9], extracting core toponyms [10], toponym disambiguation [11], identifying urban hinterland [12], simulating urban growth [13], and city interaction [14,15]. Toponym co-occurrence is also used in the study of spatial interaction in China. Liu et al. [1] first proposed a method to study the relatedness between geographic entities by toponym co-occurrences on web pages and discussed the co-occurrence pattern and spatial organization of provinces in China. The complex network methods are commonly used to analyze the spatial interaction pattern of China based on the toponym co-occurrence. Zhong et al. [16] applied the complex network analysis to toponym co-occurrence to study the characteristics of the toponym co-occurrence network, such as degree distribution, centrality, and small world. Hu et al. [17] used co-occurrence word frequency in news texts to measure the relatedness strength between cities in China and studied the spatial distribution of city influence and the interaction network pattern. In the study of the world city pattern based on the toponym co-occurrence, Zhang et al. [18] conceptualized various mesoscale structures in the world city network and explored the unique structure of the world city network presented by the toponym co-occurrence on web pages. However, these studies are only based on the undirected data of the toponym co-occurrence in the news, which cannot better represent the asymmetric interaction strength between geographical entities and did not focus on both network structure and time-varying characteristics of spatial interaction patterns.

Search index captures the attention of users to specific things through big Internet data and has characteristics, such as large volume and high velocity, which is a high-quality data source of information flow for spatial interaction analysis. The subjective will of users is not limited, so the search index is asymmetric and can show the difference between the two objects that generate the interaction. The Baidu index is an important source of search indexes. Based on the Baidu index, some studies have been conducted in terms of tourism patterns [19,20], information dissemination [21], and network attention [22]. Most of the studies on the spatial interaction of geographical entities in China are to analyze the network pattern of cities. Zong et al. [23] analyzed the current characteristics and evolution rule of the urban network structure of the twin-city economic circle in the Chengdu-Chongqing region, concluding that the agglomeration effect of large cities is obvious and the unevenness of urban networks is high. Guo et al. [24] introduced web search data to quantify the attractiveness of cities that reflects their ability to attract labor, then studied the evolution of Chinese urban systems. Wei and Pan [25] used the Baidu index as the weight of spatial connections between cities in China and used complex network analysis methods to explore the structural resilience of the city network. Dai et al. [4] used the Baidu index to analyze the network characteristics of information flow in cities along the Grand Canal by means of advantage flow analysis and cultural penetration analysis. As the province is the first-level administrative unit in China, it contains many cities and receives more attention than cities [16]. The information carried by provinces can reflect not only province-level relationships but also relationships between regions and cities across provinces. The studies of inter-provincial spatial interaction in China based on the search index are conducted from the perspective of complex network analysis. Yu et al. [26] used the Baidu index as a connection indicator to explore the pattern and structure of the

inter-provincial information connection network. Wang et al. [27] used the Baidu index to construct a provincial connection network, explored the structural characteristics of the connection network in the information space, and identified regional Balkan degrees at different levels. The search index is affected by the regional Internet development level and user psychology and is relatively weak in terms of coverage and reflecting reality. Research based on the single search index will make the results less objective.

The recent research on spatial interaction based on toponym co-occurrence and search index mostly uses a single data source, which has some shortcomings in the presentation of information flow characteristics. In addition, the recent research lacks theoretical verification that information flow can represent spatial interaction in the information space, and the analysis of inter-provincial spatial interaction patterns in China based on information flow needs more discussion. In order to take into account the comprehensive characteristics of information flow, verify the geographical laws implied in the information space, and study the spatial interaction pattern and time-varying network characteristics of provinces in China from the perspective of information flow, we propose a spatial interaction analysis method of geographical entities based on multivariate information flow. In order to reduce the limitations of a single element, we chose two data sources of toponym co-occurrence and search index for the research. Toponym co-occurrence and search index focus on the objectivity and subjectivity of information flow, respectively, and combining these two kinds of data can effectively improve the ability of information flow to represent spatial interaction. Compared with the single data of toponym co-occurrence or search index, it can reflect the spatial interaction of geographic entities more comprehensively. However, there are few studies on the spatial interaction patterns and network characteristics of provincial geographic entities in China based on integrated information flow data of toponym co-occurrence and search index; thus, we have made an attempt to do so in this research. This research innovatively uses the entropy weight method to calculate the toponym co-occurrence and search index and obtains multivariate information flow data with subjective and objective comprehensive characteristics. Through spatial autocorrelation analysis and distance decay effect test, the research verifies the underlying geographical laws of information flow in the virtual information space, and on this basis, uses complex network analysis methods to analyze the temporal changes of inter-provincial spatial interaction pattern and network structure characteristics in China from the perspective of information flow. In addition, we compare the spatial interaction network based on information flow and population flow and discuss the pattern characteristic differences between information flow and physical flow. This research provides a new idea for studying the spatial interaction of geographical entities in the physical world from the perspective of information flow.

The rest of the paper is organized as follows. In Section 2, the toponym co-occurrence dataset and the analysis methods are presented. In Section 3, in order to verify that the information flow follows the laws of geography, its correlation, heterogeneity, and distance decay effect are verified; then, complex networks based on the information flow are constructed to analyze the spatial interaction pattern of geographical entities and the changes of network characteristics, and to compare with the population migration network; center of gravity movement is used to explore the reflection of real events in the information flow. In Section 4, the significance, limitations, and contributions to future research of this study are discussed. Finally, a summary of this study is made.

2. Data and Methods

2.1. Data

In the research, toponym co-occurrence data and search index data are obtained as the source data for information flow, and population mobility data are used for comparison. The key information of the research data is shown in Table 1. The collecting scope of data is the 31 provincial-level administrative regions (hereinafter called “province”) in China, excluding Hong Kong, Macao, and Taiwan.

Table 1. Key information of research data.

	Time Period	Time Accuracy of Acquisition	Time Accuracy of Research	Sample Size per Unit of Time	Description
Toponym co-occurrence	2011 to 2020	Year	Year	31×31	Represented by co-occurrence news volume, it reflects the connection strength between provinces in real-life events.
Search index	2011 to 2020	Day	Year	31×31	Calculated by weighting the search frequency of provinces, it reflects the interprovincial attention generated by internet user behavior.
Migration ratio	2020	Day	Year	31×31	Represented as the proportion of migration, it reflects the proportion of population migration from each province to other provinces.
Migration scale index	2020	Day	Year	31	It reflects the scale of population migration in each province.
Coordinates				31	Represented by the longitude and latitude of each provincial capital, it is used to calculate the gravity center of interaction and the Euclidean distance between provinces.

This research obtains interprovincial co-occurrence data by searching on a mainstream news website. As the search engine for data collection, China News Network (<http://www.chinanews.com>, accessed on 21 January 2021) is a state-owned news website affiliated with China News Service. It provides formal news reports, which can avoid the influence of bad information on the web page and obtain high-quality toponym co-occurrence data. The data collection method is to filter the news from 2011 to 2020 in chinanews.com by using the names of two provinces as keywords (e.g., “Beijing Shandong”; the abbreviation was not considered) and to take the number of co-occurrence news obtained as the co-occurrence value of the two provinces. When the two provinces are more closely connected, they tend to appear in the same news more frequently; that is, the co-occurrence news is more, and the co-occurrence value is larger.

The search index data is obtained from the website of Baidu Index (<https://index.baidu.com>, accessed on 5 April 2021). It sets keywords as statistical objects, then calculates the weighted sum of search frequency for each keyword in the search engine. Using the province names as keywords and setting the time range and user location, the research obtains the search indices of each province for each year from 2011 to 2020. The data provided by the website is daily data, so the daily average value is used as the inter-provincial search index data for each year. The data is directional, and the numerical value reflects the connection strength of one province to another.

The population mobility data is obtained from Baidu Map Huiyan (<https://huiyan.baidu.com>, accessed on 19 May 2022), expressed as the migration scale index. Compared to other migration big data, Baidu Map Huiyan can cover all 31 provinces but only provides historical data for a few months of National Day and Spring Festival travel. In order to ensure the spatial integrity of the data, we chose Baidu Migration Data, ignoring the integrity of time. The research obtains the migration scale index and inter-provincial migration ratio for each province and for the whole country in 2020 (only 158 days are available). The migration scale index reflects the daily population mobility scale of each province, and the inter-provincial migration ratio reflects the proportion of daily migration from one province to another province. For a province, the inter-provincial migration scale in 2020 is obtained by summing the product of its daily migration scale index and inter-provincial migration ratio.

The 31 provinces in China have various shapes and large regional differences, making it difficult to express their spatial distance relationships. Provincial capital cities are often cities with intensive political and economic activities within a province, representing the political and economic centers of a province. For the spatial relationship between the two provinces, we use the geographical coordinates of each provincial capital city to calculate the Euclidean distance between provincial capitals as the inter-provincial distance data.

2.2. Spatial Autocorrelation Analysis

Spatial correlation reveals the spatial interaction relationship between elements by discovering the spatial distribution characteristics of inter-provincial interaction strength [28]. This research constructs the spatial weight matrix by the coordinates of the provincial capital through the endogenous adaptive bandwidth in the Gaussian kernel function, conducts spatial autocorrelation analysis of Moran's I and cold-hot spot for inter-provincial interaction values through Python, and visualizes the results in ArcGIS, so as to discover the spatial characteristics of inter-provincial spatial interaction in China based on information flow.

2.2.1. Moran's I Analysis

Moran's I is a kind of spatial autocorrelation coefficient, including global Moran's I and local Moran's I. It is used to determine whether there is a correlation between spatial entities within a certain range.

1. Global Moran's I

For the interaction pattern of the entity k , the formula of global Moran's I is as follows:

$$M = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}, n \in [1, 31] \text{ and } i, j \neq k \quad (1)$$

where z_i is the deviation between the interaction strength of entity i with entity k and the average interaction strength of the other 30 entities with entity k , that is, $z_i = C_{ik} - \bar{C}_k$. w_{ij} is the spatial weight of entities i and j . S_0 is the sum of all spatial weights, $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$. The value range of global Moran's I is $[-1, 1]$. When Moran's > 0 , the spatial distribution is positively correlated, and the larger the value is, the more obvious the correlation is. When Moran's < 0 , the spatial distribution is negatively correlated, and the smaller the value is, the more obvious the spatial disparity is. When Moran's $= 0$, the spatial distribution is random. Meanwhile, to estimate the spatial correlation, Z-score and p -value are needed to indicate confidence.

2. Local Moran's I

In the global correlation analysis, if the global Moran's I is significant, it can be considered that there is a spatial correlation in this region. In addition, the local Moran's I is needed to explain where the spatial aggregation phenomenon exists. For the spatial interaction relationship generated by entity k , the formula of the local Moran's I is as follows:

$$M_i = \frac{Z_i}{S^2} \sum_{i \neq j}^n w_{ij} Z_j, n \in [1, 31] \text{ and } i, j \neq k \quad (2)$$

where S^2 is used to standardize the formula, $S^2 = \frac{1}{n} \sum_{i=1}^n (Z_{ik} - \bar{Z}_k)^2$. Combined with the significance test, there are four kinds of spatial correlation, as shown in Table 2.

Table 2. The spatial correlation is reflected by the local Moran's I.

Z_i	M_i	Spatial Correlation
>0	>0	The interaction strength of entity i is high, and the strength of its surrounding areas is high (H-H).
>0	<0	The interaction strength of entity i is high, but the strength of its surrounding areas is low (H-L).
<0	>0	The interaction strength of entity i is low, but the strength of its surrounding areas is high (L-H).
<0	<0	The interaction strength of entity i is low, and the strength of its surrounding areas is low (L-L).

According to the local Moran's I and the results of the significance test, a LISA map can be plotted to visualize the spatial interaction relationship generated by entities.

The global Moran's I and local Moran's I can reflect the spatial aggregation degree and distribution of spatial interaction of 31 provinces in China based on the toponym co-occurrence and search index, which contributes to discovering the geographical laws of spatial interaction based on information flow.

2.2.2. Cold-Hot Spot Analysis

The cold-hot spot analysis method can be used to reveal the spatial clustering characteristics of a local area. It is used to evaluate each element in the context of adjacent elements, then compare the local situation with the global situation to identify spatial clusters of high values (hot spots) and low values (cold spots) with statistical significance. The Getis-Ord G_i^* index is a common indicator to describe the cold-hot spot, and is calculated as:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}x_j - \bar{X}\sum_{j=1}^n w_{ij}}{S\sqrt{\frac{[n\sum_{j=1}^n w_{ij}^2 - (\sum_{j=1}^n w_{ij})^2]}{n-1}}}, n \in [1, 31] \text{ and } -i, j \neq k \quad (3)$$

where x_j is the interaction strength, w_{ij} is the spatial weight of entities i and j , respectively, \bar{X} is the mean, S is the standard deviation, and the statistic of G_i^* is the Z-score. A statistically significant positive Z-score indicates a hotspot, and if the element has a high Z-score and a small p -value, it indicates a high-value spatial cluster. If the negative Z-score is low and the p -value is small, it indicates a low-value spatial clustering. The higher (or lower) the Z-score, the greater the degree of clustering. If the Z-score is close to 0, then there is no significant spatial clustering. By the cold-hot spot analysis of inter-provincial toponym co-occurrence and search index, we can explore the spatial aggregation of spatial interaction strength based on the information flow of provinces in China, then explore the underlying geographical laws.

2.3. Distance Decay Effect

If the spatial interaction based on information flow conforms to the first law of geography, it indicates that the spatial interaction has a distance decay effect. The essence of the distance decay effect is that the interaction of two geographic entities is related to the spatial distance between them, and the interaction weakens with increasing distance [29]. In order to explore the first law of geography of spatial interaction based on information flow, the distance attenuation effect is quantitatively expressed. Based on the gravity model, the parameters in the distance decay effect are fitted, and the fitting formula is as follows:

$$I_{ij} = K \frac{C_i^\gamma C_j^\nu}{D_{ij}^\beta} \quad (4)$$

In the formula, I_{ij} is the interaction strength of provinces i and j . C_i and C_j are, respectively, the interaction quality of provinces i and j . D_{ij} is the distance between the two provinces, represented by the Euclidean distance calculated from the longitude and latitude coordinates of each provincial capital. K is the correction coefficient. γ and v can reflect the impact of the interaction quality of provinces i and j on the inter-provincial interaction strength. β is the distance decay coefficient. The distance decay coefficient indicates how fast the attraction strength decays with increasing distance, that is, the strength of the distance decay effect in the spatial interaction strength reflected by toponym co-occurrence and search index. The smaller the β is, the weaker the distance decay effect of the spatial interaction based on information flow is; that is, the less distance impedes the interaction strength of information flow.

In this study, the nonlinear fitting of the gravity model is carried out in SPSS, and the influence coefficients γ and v of provincial interaction quality and the distance decay coefficient β in the spatial interaction based on the information flow are obtained. Similarly, the β of the spatial interaction based on the physical flow of population mobility is calculated and it is compared with the β based on information flow to explore the spatial characteristics of spatial interaction of geographical entities in virtual information space.

2.4. Complex Network Analysis

Complex network analysis is used to study the overall characteristics of a network, for which the centrality measurements are important tools. The spatial interaction network based on information flow constructed in the research is a directed weighted network, the importance of province nodes in the network can be measured by the centrality, such as degree centrality and eigenvector centrality.

PageRank (PR) centrality is a variation of eigenvector centrality, and the classical PR algorithm is used to rank pages through hypertext links. In a complex network, if a node is pointed to by more other nodes, the PR value of the node is larger. Additionally, if a node has a higher PR value itself and it points to another node, the PR value of the pointed node is higher. Different from the classical algorithm, the ranking results of nodes in the weighted network also take into account the weight of inter-provincial spatial interaction [16].

$$PR(i) = \alpha \sum_j^n \frac{W_{ij} \times PR(j)}{\sum_{i=1, i \neq j}^n W_{ij}} + \frac{1 - \alpha}{n} (i \neq j) \quad (5)$$

where $PR(i)$ represents the ranking score of the province i , W_{ij} is the connection weight between provinces i and j , respectively, and n is the number of nodes. α is a stability coefficient, which is used to prevent the algorithm from “sinking node” and is generally set to 0.85. Through iterative calculation, the PR value of each province tends to be stable. Provinces with large PR centrality are of more importance in the spatial interaction network.

Relative degree centrality is the normalization of degree centrality. For a weighted spatial interaction network, the relative degree of centrality is as follows:

$$C_{RD}(i) = \frac{\sum_j^n W_{ij}}{(n - 1) \times W_{max}} (i \neq j) \quad (6)$$

where $C_{RD}(i)$ is the relative degree centrality of province i , $n(1 \leq n \leq 31)$ is the number of nodes, W_{ij} is the connection weight of provinces i and j , respectively, and W_{max} is the maximal weight in the network. For a spatial interaction network, network centralization can denote the extent to which the network is organized around one or some central provinces. The network centralization can be expressed by the following formula:

$$Cent = \frac{\sum_i^n (C_{RDmax} - C_{RD}(i))}{n - 1} \quad (7)$$

where $Cent$ is the network centralization and C_{RDmax} is the maximal relative degree centrality in the network. The greater difference in the degree of centrality of province nodes leads to greater network centralization. It indicates that as the degree distribution of nodes becomes more unbalanced, the provinces with a higher degree centrality have stronger control over other provinces, and the network tends to expand from these core provinces.

In this research, we input provincial information flow data over the years in Ucinet, calculate the PR centrality of province nodes and the network degree centrality of each year, and rank the results. PR centrality ranking can reflect the position of each province in the spatial interaction network based on the information flow of the year, and temporal changes in ranking can reflect the changes in its importance in the interaction network. The changing trend of network centralization can reflect the stability changes of spatial interaction network structure.

2.5. Model of Gravity Center

The center of gravity model is an important analytical tool for studying changes in the characteristics of elements in the process of regional development [30]. The coordinates of the gravity center are the indicators describing the spatial distribution of geographic entities, which can clearly and objectively reflect the changes in the spatial and temporal trajectories of their characteristics. In the process of regional development, the movement of the gravity center is a manifestation of the synergistic development of each region. The coordinates of the gravity center of each province are used as the geographical coordinates, and the total interaction quantity of each province is used as a weight indicator to calculate the coordinates of the gravity center for toponym co-occurrence and search index in China. The model of the gravity center can be expressed as follows:

$$X = \frac{\sum_{i=1}^n x_i \times w_i}{\sum_{i=1}^n w_i} \quad (8)$$

$$Y = \frac{\sum_{i=1}^n y_i \times w_i}{\sum_{i=1}^n w_i} \quad (9)$$

where X and Y are the coordinates of the gravity center; x_i and y_i are the latitude and longitude coordinates of province i ; and w_i is the weight, expressed as the cumulative co-occurrence value or search index value of the province i .

2.6. Entropy Weight Method

Toponym co-occurrence and search index are generated from social news and user search behavior, which focus on the objectivity and subjectivity of information flow data, respectively. In order to reflect the characteristics of information flow more comprehensively, the entropy weight method is used to calculate the panel data of toponym co-occurrence and search index to obtain multivariate information flow data that can reflect the comprehensive characteristics. The entropy weight method uses information entropy to calculate the entropy weight of each indicator according to their variation degree, then modifies the weight of each indicator through the entropy weight so as to obtain a more objective index weight. Compared with cross-section data, panel data needs to consider the total information entropy of the data in the overall time. The calculation process of the entropy weight method for panel data is as follows [31]:

(1) Data standardization. The toponym co-occurrence and search index are positive indicators, so the positive standardization formula is adopted.

$$X'_{\theta ij} = \frac{X_{\theta ij} - \min(X_{\theta ij})}{\max(X_{\theta ij}) - \min(X_{\theta ij})} \quad (10)$$

where $X_{\theta ij}$ stands for the value of indicator j of item i in year θ . After standardization, 0 will be generated, so a minimum value needs to be added for data translation. The minimum value is set to 1×10^{-5} .

(2) Proportion calculation. Calculate the proportion of the value j under indicator j of item i in year θ . M is the number of samples.

$$P_{\theta ij} = \frac{X'_{\theta ij}}{\sum_{\theta=1}^d \sum_{i=1}^m X'_{\theta ij}} \quad (11)$$

(3) Information entropy calculation. Calculate the information entropy of the indicator j .

$$E_j = -\frac{1}{\ln(dm)} \sum_{\theta=1}^d \sum_{i=1}^m [P_{\theta ij} \cdot \ln(P_{\theta ij})] \quad (12)$$

(4) Weight calculation. Calculate the discrimination factor of indicator j .

$$G_j = 1 - E_j \quad (13)$$

Calculate the weight of indicator j .

$$W_j = \frac{G_j}{\sum_{j=1}^n G_j} \quad (14)$$

(5) Comprehensive score calculation.

$$Z_{\theta i} = \sum_{j=1}^n (W_j \cdot X'_{\theta ij}) \quad (15)$$

$$Z = X' \cdot W \quad (16)$$

Z is the panel data of multivariate information flow obtained after weighted calculation by toponym co-occurrence and search index. Multivariate information flow data takes into account the attribute and temporal characteristics of toponym co-occurrence and search index, which contributes to more reasonably reflecting the spatial interaction patterns and structural characteristics of the interaction network based on information flow.

3. Results

3.1. Laws of Geography in Information Flow

3.1.1. Discovering Correlation and Heterogeneity

The characteristic of spatial interaction is an important part of geographic analysis [32]. The interaction pattern refers to the behavioral rule of spatial interaction between geographic entities, that is, the distribution of the interaction strength of one entity with other entities. By analyzing the interaction pattern of toponym co-occurrence and search index, the characteristics and rules of information flow in the textual space can be discovered, which contributes to mining the interaction characteristics of the spatial entities implied in information flow.

From the two aspects of toponym co-occurrence and search index, Moran's I is used to measure the spatial correlation of the interaction pattern of each province. The global Moran's I for the co-occurrence strength of each province with the other 30 provinces is calculated by Formula (1), as shown in Table 3.

Table 3. Global Moran’s I of provinces in China, taking the top 10 provinces. The Z-score is the multiple of the standard deviation. The results are all extremely significant, p -value < 0.01, so the p -value is not shown.

Toponym Co-Occurrence			Search Index		
Province	Moran’s I	Z-Score	Province	Moran’s I	Z-Score
Hebei	0.740711	7.642184	Beijing	0.808857	5.998074
Anhui	0.719599	5.449721	Hebei	0.785282	7.149060
Shanxi	0.681523	6.109955	Tianjin	0.765024	6.607754
Jiangxi	0.660554	4.982462	Anhui	0.759295	5.829506
Shandong	0.649321	5.597223	Shanghai	0.688321	5.241796
Jiangsu	0.644998	5.151514	Shandong	0.683801	5.430812
Tianjin	0.643374	6.270135	Guangdong	0.663948	5.055511
Henan	0.628473	5.221496	Shanxi	0.660656	5.696736
Liaoning	0.619655	5.357829	Liaoning	0.660091	5.490013
Fujian	0.615384	4.803019	Zhejiang	0.646384	5.280391

It can be seen that the interaction patterns of provinces in China present different significant positive correlations, and the global Moran’s I of search index is higher than that of toponym co-occurrence. For toponym co-occurrence and search index, there are, respectively, 10 and 16 provinces with Moran’s I greater than 0.6. The provinces with stronger spatial correlation are mainly located in southeastern and eastern China. A few provinces have low spatial correlation, and most of these provinces are located in western China. On the one hand, they tend to interact with neighboring provinces in the west, and on the other hand, they also tend to interact with developed provinces in the east. In this case, the strength of geographic entities weakens the influence of distance, resulting in the random distribution of interaction strength.

Global Moran’s I indicate the existence of spatial aggregation in the interaction pattern. In order to clarify the spatial aggregation distribution in the interaction pattern of each province, local Moran’s I is used to measure it. After the significance test, the maps of local indicators of spatial association (LISA) are plotted. Provinces with a significant positive spatial correlation have a more obvious aggregation distribution. Take Hebei, where the global Moran’s I of toponym co-occurrence and search index are both large; as an example, the Moran scatter maps and LISA maps for its interaction pattern are shown in Figure 1.

It can be found from the figure that for the provinces whose interaction pattern has significant spatial correlation, the H–H cluster is usually distributed around them, and the L–L cluster tends to be distributed farther away. It means that the neighboring provinces have high interaction strength, and the interaction strength decreases with distance. This is consistent with Tobler’s first law of geography (TFL), namely, all things are related, but nearby things are more related than distant things [33].

Goodchild’s second law of geography (GSL) [34] refers that spatial isolation causes differences between objects, forming local spatial heterogeneity. To explore the heterogeneity of spatial interaction, a cold-hot spot analysis for the co-occurrence strength and search index of each province with the others is conducted. For different provinces, there are differences in the degree of spatial autocorrelation. Provinces with high spatial autocorrelation have significant spatial aggregation of cold-hot spots. Taking Hebei with high spatial autocorrelation as an example, the distribution of cold and hot spots is explored, as shown in Figure 2.

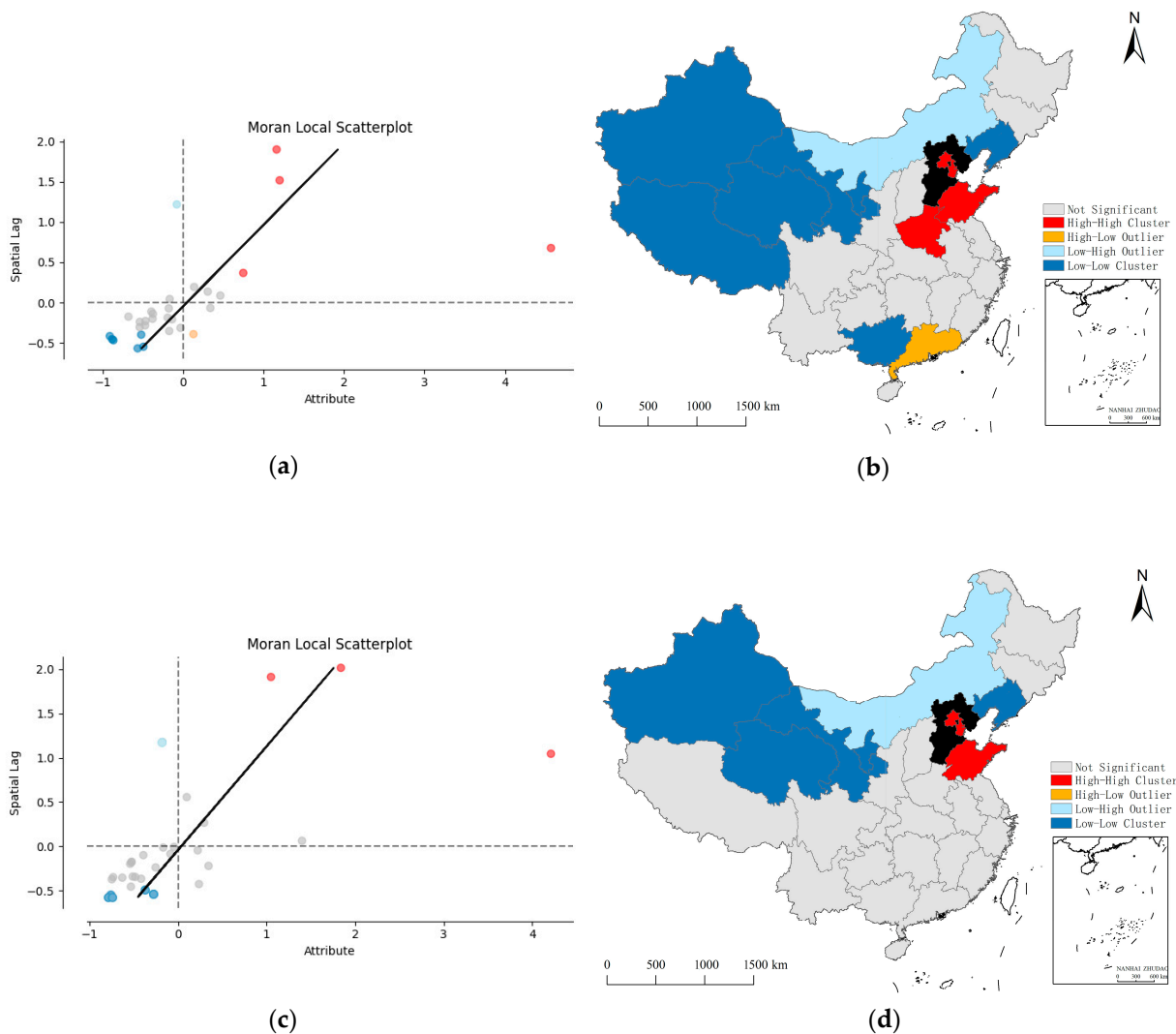


Figure 1. LISA maps and scatter maps for local Moran's I, taking Hebei as an example. (a,b) are for toponym co-occurrence. (c,d) are for the search index. The province in black is the reference one on each map. High–High (H–H) Cluster: it's own and it's neighbors' co-occurrence values are all high. High–Low (H–L) Outlier: its own value is high, but its neighbors' values are low. Low–High (L–H) Outlier: its own value is low, but its neighbors' values are high. Low–Low (L–L) Cluster: its own and its neighbors' values are all low.

Figure 2 shows the cold-hot spot distribution of toponym co-occurrence and search index, both of which show spatial heterogeneity in terms of interaction strength. On the one hand, the differences in interaction strength cause the dissimilarity in the significance of cold spots or hot spots. In the cold-hot spot analysis of the search index, the number of provinces passing the significance test is higher than that of toponym co-occurrence, but the significance is slightly lower. Among them, the number of hot spots is the same, and the number of cold spots in the search index is higher than that of toponym co-occurrence. On the other hand, from the perspective of spatial aggregation, both of them have generated cold spot aggregation and hot spot aggregation, and the hot spot area is relatively close.

The above analysis indicates that the spatial interaction pattern has correlation and heterogeneity by taking toponym co-occurrence and search index as the representation in the textual information space; that is, it conforms to TFL and GSL in the physical space. TFL shows that not only are geographic entities related, but their correlation decreases with increasing distance; that is, the correlation obeys the distance decay effect. In order to

further verify that toponym co-occurrence in the information space obeys TFL in the real world, we then verify the distance decay effect in toponym co-occurrence.

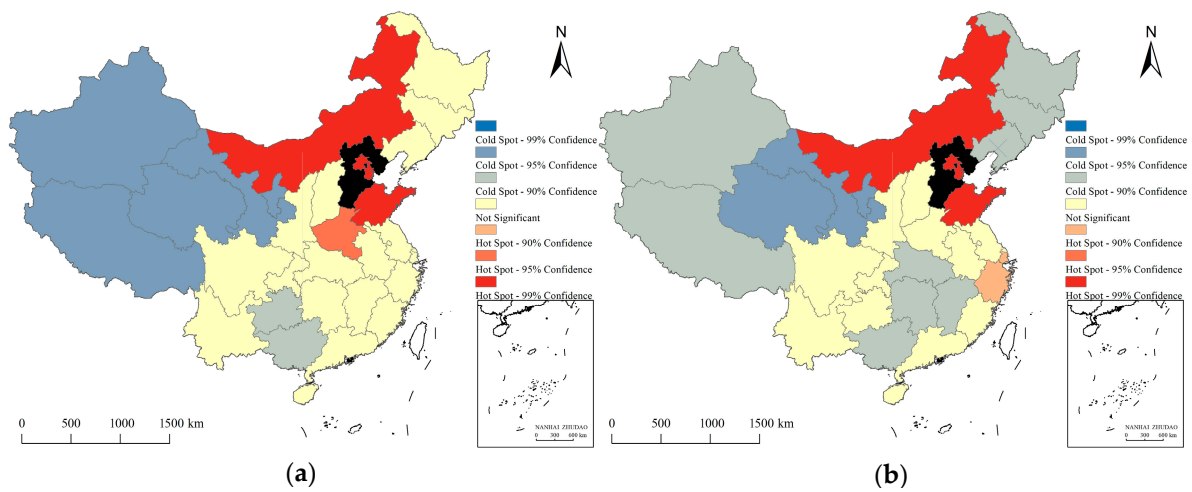


Figure 2. Maps of cold-hot spot analysis for toponym co-occurrence, taking Hebei as an example. (a) Toponym co-occurrence. (b) Search index. The black province is the reference one on each map.

3.1.2. Verifying Distance Decay Effect

In order to quantify the distance decay effect in toponym co-occurrence and search index, we use the gravity model for parameter fitting to obtain the distance decay coefficient β . First, we set different β with a step size of 0.001 to fit other parameters of the gravity model and use R^2 to represent the goodness of fit. When R^2 reaches the maximum value, the values of the optimal β are, respectively, 0.189 and 0.186, and R^2 are, respectively, 0.818 and 0.533. Therefore, the quantitative expression of the distance decay effect of toponym co-occurrence and search index can be obtained, respectively:

$$I_{Tij} = 4.430 \times 10^{-3} \frac{C_{Ti}^{0.682} C_{Tj}^{0.682}}{D_{ij}^{0.189}} \quad (17)$$

$$I_{Sij} = 1.398 \times 10^{-6} \frac{C_{Si}^{1.001} C_{Sj}^{0.857}}{D_{ij}^{0.186}} \quad (18)$$

It can be seen from the two formulas that β of toponym co-occurrence and search index are similar, and β of the search index is slightly smaller than that of toponym co-occurrence. Therefore, the influence of distance on the search index is slightly less than that of toponym co-occurrence; that is, the influence of distance on the spatial interaction generated by users' retrieval behavior is slightly less than that of the spatial interaction generated by objective events. In addition, in Formula (17), the parameters of C_{Ti} and C_{Tj} are the same, which means that provinces i and j have the same contribution to the interaction of toponym co-occurrence. In Formula (18), the parameter of C_{Si} is greater than C_{Sj} , which indicates that province i , as the generator of retrieval behavior, has a greater contribution to the spatial interaction of the search index.

A longer time scale of data allows a more comprehensive analysis of the characteristics of data, but it makes the change of characteristics over time fuzzy. Therefore, we analyze the distance decay effect of toponym co-occurrence and search index data for each year and fit their distance decay coefficients, as shown in Figure 3.

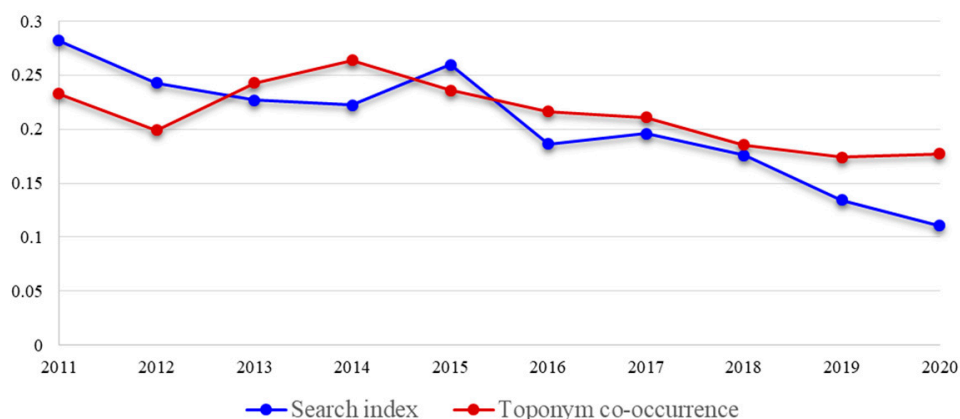


Figure 3. Distance decay coefficients of toponym co-occurrence and search indices for each year.

Overall, the β of toponym co-occurrence and search index show a decreasing trend, and the β of the search index had a greater decrease degree, while that of toponym co-occurrence is relatively stable. The range of β for toponym co-occurrence is 0.264 to 0.174, and the range of β for the search index is 0.282 to 0.111. Their distance decay coefficients for all years also remain lower than those in geographic space. The β of toponym co-occurrence is higher than that of the search index in most years, indicating that the influence of distance on toponym co-occurrence is larger than that on the search index in the long term, which is consistent with the previously obtained β of toponym co-occurrence being slightly larger than that of search index on a 10-year time scale.

Through the above research, it is found that the information space represented by toponym co-occurrence and search index obeys TFL and GDL. Therefore, the information space, specifically the textual space in the research, can be used as a mapping of geographic space in the material world.

3.2. Spatial Interaction Mapping of Geographic Entities

Toponym co-occurrence and search index are weighted by the entropy weight method to obtain the multivariate information flow data. The multivariate information flow data can comprehensively reflect the objectivity and subjectivity of toponym co-occurrence and search index. Compared with some physical flow data, it is less constrained by time and space. Therefore, the interaction of geographic entities can be analyzed based on information flow, which is a data source supplement to the current research methods of city interaction, such as traffic flow and human migration. The complex network constructed based on information flow can be used as the information domain mapping of the interaction network of real geographic entities. Three types of information flow data are used to construct hierarchical complex networks and to analyze the interprovincial spatial interaction pattern. The province is regarded as the network node, and the interaction strength is regarded as the weight of the edge. In the research, there are 465 province pairs and 465 undirected edges or 930 directed edges. The number of classification grades determined by the Goodness of Variance Fit (GVF) method is 4. According to the interaction strength, all provincial pairs are divided into four grades by using the method of natural breaks (Jenks), and the hierarchical spatial interaction networks are drawn.

3.2.1. Spatial Interaction Networks of Toponym Co-Occurrence and Search Index

The toponym co-occurrence network can reflect the interaction of geographic entities in real events, but it is not spatially directed. The search index network can reflect the willingness of geographic entities to generate interactions with other entities. The spatial interaction networks of toponym co-occurrence and search index are constructed, respectively, as shown in Figure 4.

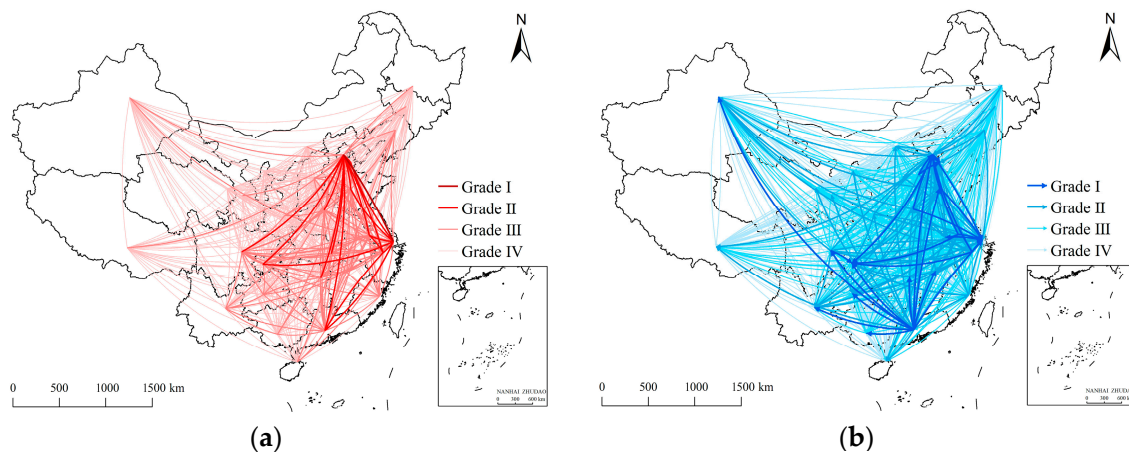


Figure 4. Spatial interaction networks of two types of information flow. (a) Toponym co-occurrence network, which is an undirected network. (b) Search index network, which is a directed network.

In the toponym co-occurrence network, the percentages of the four interaction grades are, respectively, 0.22%, 4.09%, 24.73%, and 70.97%. In particular, the network of grade I contains only the province pair of Beijing–Shanghai, whose interaction strength of toponym co-occurrence is much higher than that of other provincial pairs. Therefore, the initial pattern appears in the network of grade II in the toponym co-occurrence network, i.e., a diamond-shaped pattern with Beijing, Shanghai, Guangdong, and Chongqing as the main nodes. The next-level network is also mainly distributed within this diamond-shaped pattern, with a slight expansion to the southern and northeastern regions and only a few interaction edges in the western region.

In the search index network, the percentages of the four-interaction grade I s are, respectively, 3.12%, 10.22%, 31.83%, and 54.84%. The network of grade I consists of one-way interactions, forming a diamond-shaped pattern similar to toponym co-occurrence. The network of grade II is still mainly one-way interactions, mainly distributed in the eastern and central regions of China and less distributed in the western and northeastern regions. The formation of the unidirectional network indicates that the interaction behaviors sent by two provinces to each other fail to reach the same strength level in the search index; that is, there is an asymmetry in the spatial interaction of geographic entities.

Both networks of toponym co-occurrence and search index present a diamond-shaped spatial interaction pattern which is an important part of the spatial interaction network. In the diamond-shaped pattern, geographic entities have higher spatial interaction strength. However, the search index network is more balanced than the co-occurrence network in terms of hierarchical division, presents a more obvious spatial pattern, and has a more solid network structure. There are long-distance edges in the high interaction grades of the search index network, which reflects that the distance decay effect has less influence on the search index than toponym co-occurrence.

3.2.2. Patterns of Interaction Network Based on Multivariate Information Flow

In order to reflect the spatial interaction pattern presented by the information flow in the text space comprehensively, the interaction network of multivariate information flow is constructed, as shown in Figure 5. The network structures of different interaction grades are compared, and the influencing factors of their formation are speculated for the spatial pattern presented in each grade.

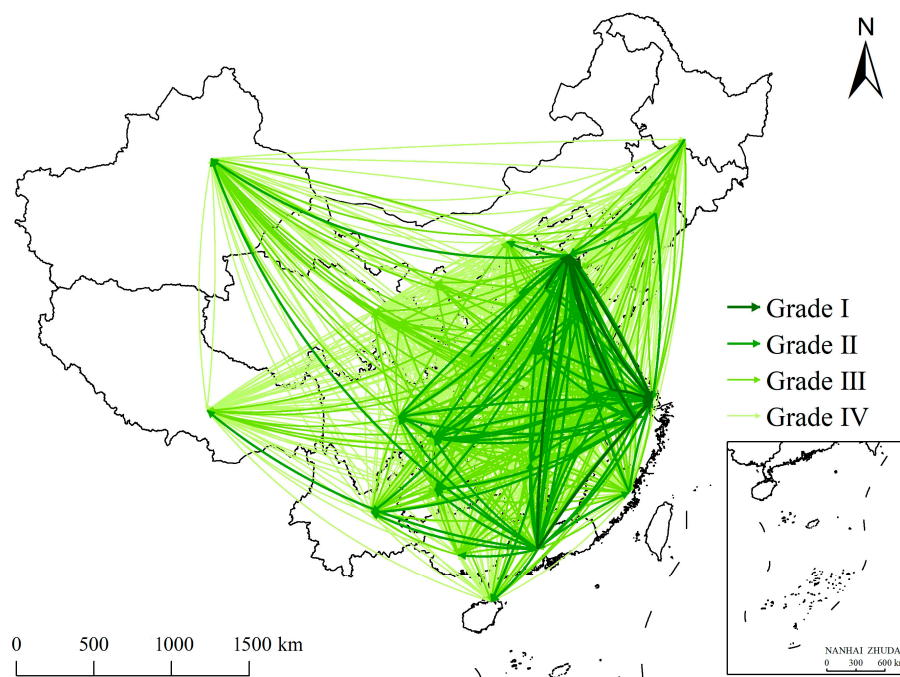


Figure 5. The spatial interaction network is based on multivariate information flow.

- Grade I. There are 11 province pairs, accounting for 1.18% of the total amount of interprovincial interaction. The interaction network in this grade is located in eastern China, forming a triangular primary network with Beijing, Shanghai, and Guangdong as the vertices. The network contains bidirectional interaction between Beijing and Shanghai, which shows the prominence of the interaction between the two cities in the spatial interaction of China. In addition, other edges in the network in this grade are also associated with these two cities. Except for the interaction from Beijing to Tianjin, all other interactions are directed from other provinces to Beijing and Shanghai. Since Beijing and Shanghai are, respectively, the political and economic centers of China, and the interactions are obviously distributed across regions, the influence of political and economic factors in this grade network is much greater than that of the distance factor.
- Grade II. There are 92 province pairs, accounting for 9.89% of the total amount of interprovincial interaction. The interaction network in this grade mainly distributes in the eastern and central regions of China and shows obvious cross-regional interaction. It forms an almost diamond-shaped network of interprovincial spatial interaction. The network includes 27 provinces in China. Beijing, Shanghai, Guangdong, and Chongqing have the closest interaction with other provinces, which is the main part of Grade II. Among them, Guangdong has the most out-edges and the only in-edge. Combined with the interaction in grade I, although Guangdong has higher interaction strength, it fits the role of a generator in spatial interaction more; that is, its executive force is greater than its attraction.
- Grade III. There are 282 province pairs, accounting for 30.32% of the total amount of interprovincial interaction. The interaction network in this grade is concentrated in the eastern and central regions of China, with increased interaction in the western and northeastern regions, which jointly constitute the general network of interprovincial spatial interaction. The network in this grade covers all provinces in China and adds the interaction between adjacent provinces in which the spatial distance performs a major role. With the expansion of the network scale, the influence of the distance factor increases and is similar to the influence of political and economic factors in this grade.
- Grade IV. There are 545 province pairs, accounting for 58.60% of the total amount of interprovincial interaction. In this grade, the network contains the rest of the spatial

interactions of provinces in China. Influenced by the territory of China, the network presents an almost trapezoidal spatial pattern. The network in this grade includes most of the province pairs, which have relatively weak interaction in all aspects of spatial distance, politics, and economy.

As shown in Figure 5, taking the provinces, such as Beijing, Shanghai, Chongqing, and Guangdong, as the core nodes, the triangular primary network of spatial interaction is formed. Taking the southeastern, eastern, and central regions as important regions, the diamond-shaped secondary network is formed. It expands outward grade by grade, and the almost trapezoidal space pattern of the global interprovincial interaction network is finally formed. The influence of political and economic factors is dominated in the triangular primary network and the influence of spatial distance increases in the diamond-shaped secondary network. The overall spatial pattern finally formed is closely related to the geographic distribution of China.

3.2.3. Centrality of Information Flow Network

The constructed toponym co-occurrence network is an undirected fully connected network, where the degree distribution of nodes is characterized by scale-free [16]. The search index network and the multivariate information flow network are directed weighted networks, and the degree distribution of nodes is also characterized by scale-free. Most nodes have a small degree value, while a few nodes that are the core points of the complex network have a large degree value. In order to study the status of provinces in the spatial interaction network and the cohesion of the network, we select the interprovincial toponym co-occurrence, search index, and multivariate information flow data from 2011 to 2020 to construct the interaction network for each year and calculate the PageRank centrality of nodes and the network centralization.

It can be seen from Figure 6 that the importance of provinces in the spatial interaction network is overall in a stable state, and only a few provinces change greatly. The top eight provinces (i.e., Beijing, Shanghai, Guangdong, Zhejiang, Shandong, Jiangsu, Sichuan, and Henan) have a greater degree of centrality than the average per year. Among the centrality of the three networks, the provinces with higher average rankings are in the central position of the spatial interaction network, which is closely related to the above diamond-shaped pattern of the core network for provincial spatial interaction. In particular, Beijing is in the absolute core position with its long-term first place, which shows the outstanding influence of Beijing as the capital of China in the nationwide spatial interaction. Shanghai has the second highest influence.

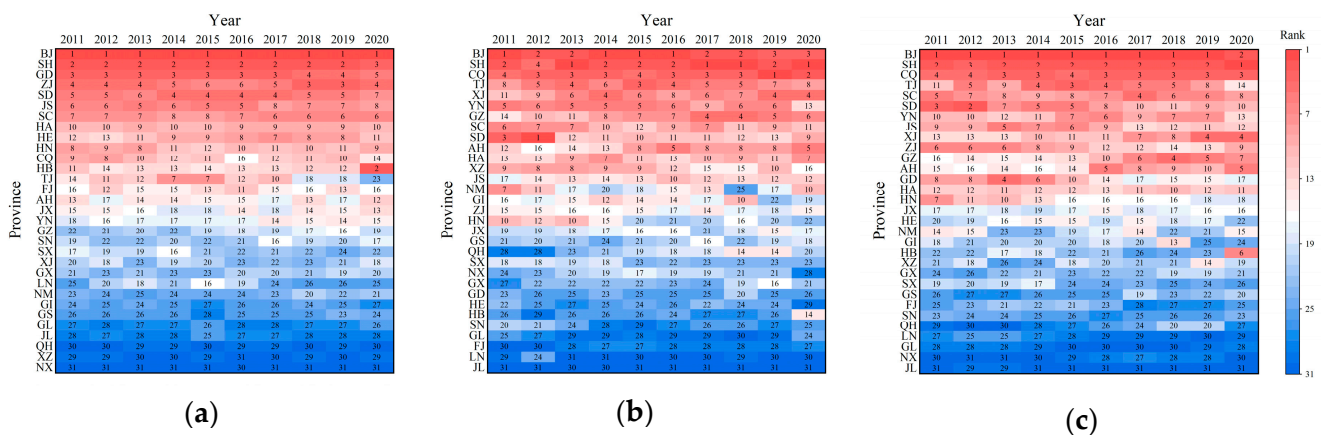


Figure 6. PageRank centrality of the interactive spatial networks from 2011 to 2020. (a) Toponym co-occurrence. (b) Search index. (c) Multivariate information flow. The ordinate is sorted by the average value.

Figure 7 shows the trends of network centralization of toponym co-occurrence, search index, and multivariate information flow. The centralization of toponym co-occurrence shows an increasing trend, indicating that the interaction strength of core provinces in the toponym co-occurrence network is increasing. The in-degree centralization of the search index shows a decreasing trend, while the out-degree centralization basically shows an increasing trend. The unevenness of receiving interactions by provinces moderates year by year, while the behavior of generating interactions increasingly aggregates toward the core provinces. On the whole, the network centralization of toponym co-occurrence is lower than that of the search index, and its fluctuation range is also smaller than that of the search index. The spatial interaction network characterized by toponym co-occurrence is less hierarchical than the search index network.

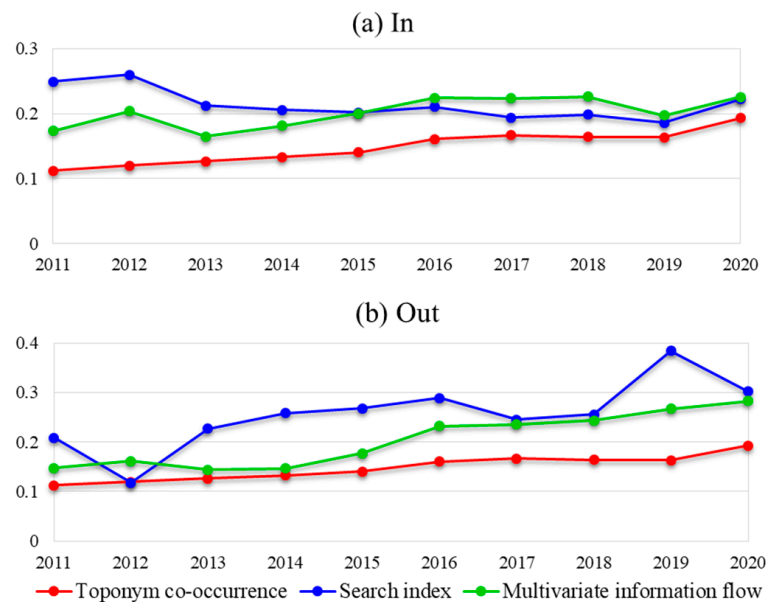


Figure 7. The network centralization changes of the three networks from 2011 to 2020. (a) Centralization of in-degree network. (b) Centralization of out-degree network. As the toponym co-occurrence network is undirected, its centralization of in-degree network is equal to that of out-degree.

3.2.4. Movement of Gravity Center of Spatial Interaction

The movement of the gravity center of spatial interaction can reflect the trend of regional development to a certain extent. The latitude and longitude coordinates of each province and the cumulative interaction strength are used to calculate the coordinates of the gravity center for each year. The location and movement of the gravity center from 2011 to 2020 are shown in Figure 8.

The gravity centers of toponym co-occurrence, search index, and multivariate information flow are all distributed in east-central China during the decade, with an accumulated movement of 180.13 km, 201.19 km, and 175.14 km, respectively. The moving diameters are, respectively, 77.85 km, 96.63 km, and 94.25 km, which are less than two percent of the length of Chinese territory, so the movement is relatively small. It suggests that the change in the importance of Chinese provinces reflected by information flow is relatively steady, with eastern regions being more important than western regions. For toponym co-occurrence, the gravity center moved to the northwest continually from 2013 to 2019 because the “Belt and Road” initiative promotes the development of the western region and alleviates the east–west gap [35,36]. For the search index, the gravity center moved to the northeast in 2011 and 2012 due to Shandong having an extremely high index. It moved in a much smaller range with a relatively stable position from 2013 to 2020. Combining toponym co-occurrence with the search index, the gravity center of multivariate information flow fluctuates from the northeast to the southwest as a whole, which has a high similarity with the regional economic development of China [37].

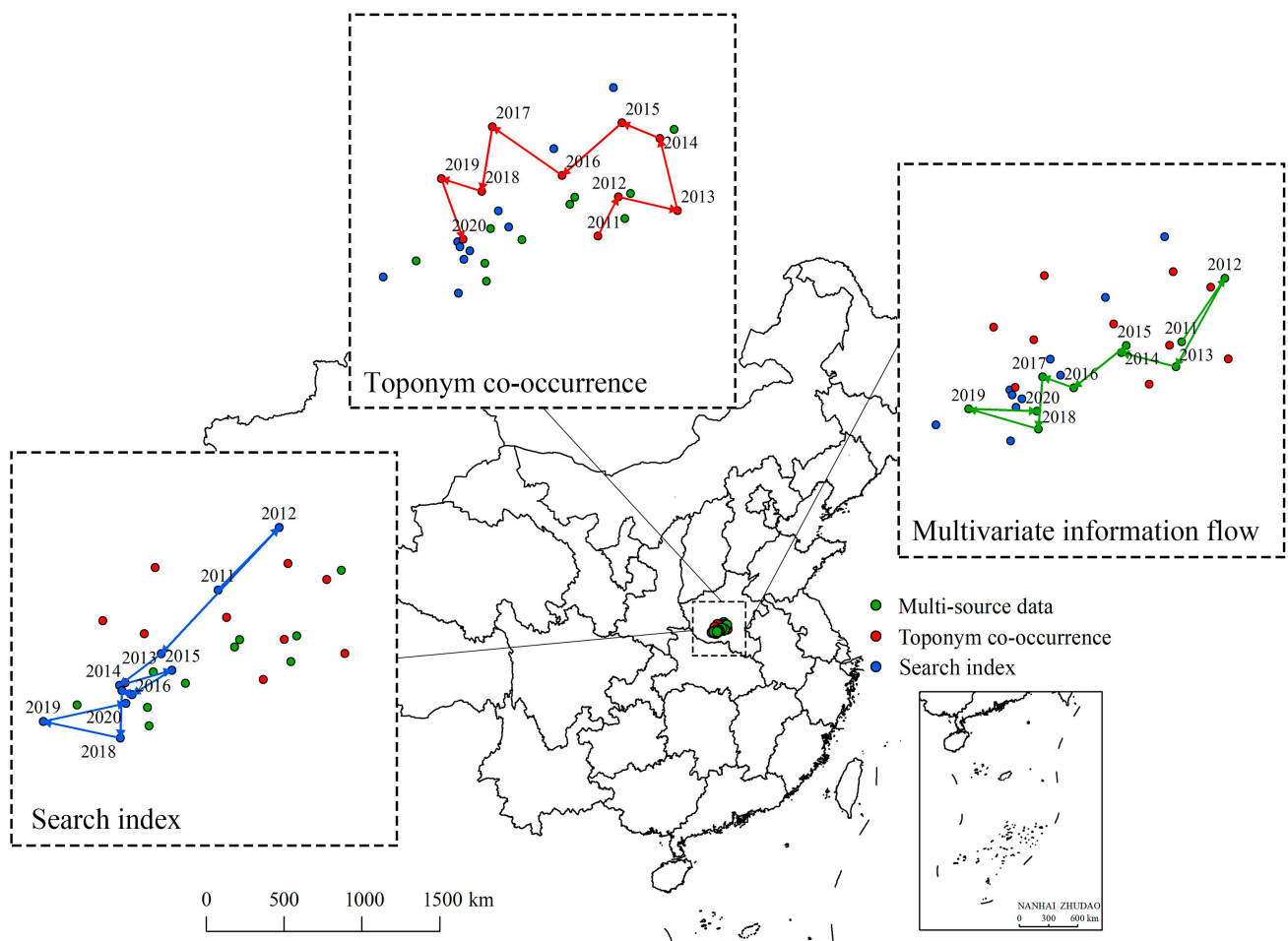


Figure 8. Movement of the gravity center of spatial interaction from 2011 to 2020.

In 2020, the gravity center of toponym co-occurrence, search index, and multivariate information flow all moved eastward because Hubei became the focus due to COVID-19. Thus, it can be seen that whether it is toponym co-occurrence or search index, the information flow can reflect the impact of major social events and policies on the whole country and is their mapping in textual space.

3.3. Comparison between Networks of Information Flow and Population Mobility

Population mobility is the most active factor in the economic and social system, and its role in reshaping the urban network is fundamental [38]. While the information flow covers more information sources, the resulting spatial interaction network reflects the result of more social factors compared with the population mobility network. In order to clarify the differences between the two network structures, the population mobility network is constructed to compare with the multivariate information flow network, as shown in Figure 9.

The central nodes of the population mobility network have high consistency with the multivariate information flow network, such as Beijing, Shanghai, Jiangsu, Guangdong, and Chengdu. Provinces, such as Guangdong and Jiangsu, have a more important position in the population mobility network, and the dominance of Beijing and Shanghai as political and financial centers is weakened. The difference is that the information flow network forms a diamond-shaped major pattern with these nodes, presenting a cross-regional distribution of spatial interactions in China. In contrast, in the population mobility network, these nodes only interact at high strength with neighboring provinces, and they are mostly the developed provinces along the eastern coast. The distance decay coefficient β for

the population mobility in 2020 is 0.976, which is much larger than the β of 0.162 for the multivariate information flow in 2020. The distance decay coefficient of the physical flow is significantly greater than that of the information flow because the friction of the distance to the information flow is greatly reduced in the virtual information space. Therefore, the influence of geographical distance cannot be ignored in causing the difference in the spatial pattern of the two networks. In addition, due to the round-trip characteristic of population mobility, the population mobility network is mostly two-way interaction at the same level, so the asymmetry of interaction is weaker than that of the information flow network.

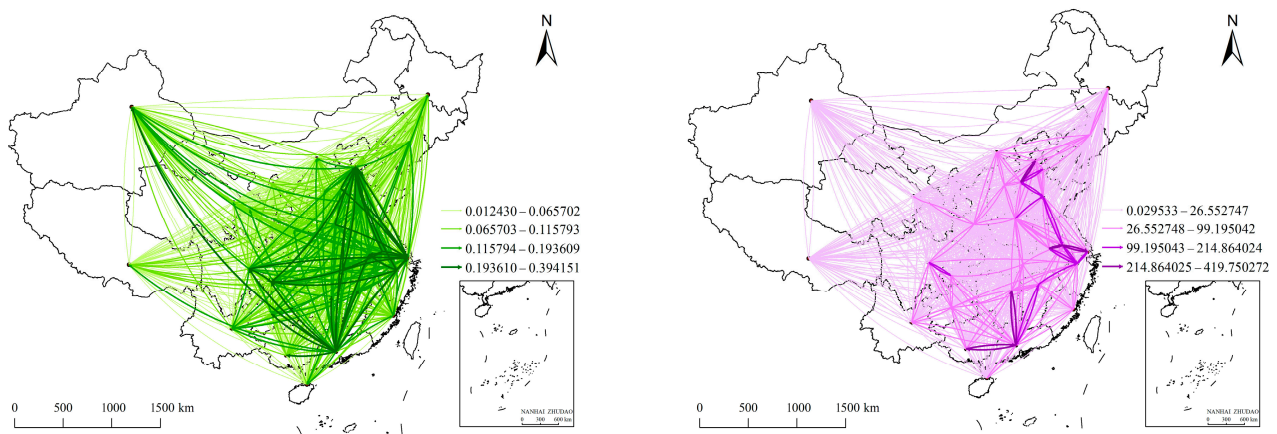


Figure 9. Networks of multivariate information flow and population mobility in 2020.

4. Discussion

4.1. The Laws of Geography in the Information Space

Before conducting spatial autocorrelation analysis and distance decay effect test, it is first necessary to define the spatial relationships of provinces. Previous research used topological distance to define spatial relationships [1], which can reduce distance ambiguity caused by area differences, but its practical significance is weakened. This research defines neighbors by calculating Euclidean distance using the coordinates of the provincial capital. Due to the significant differences in the area of Chinese provinces and the uneven distribution of their capitals, using reciprocal distance to define weights can lead to significant differences in sample size. Thus, we use Gaussian kernel endogenous adaptive bandwidth to construct the spatial weight matrix. The endogenous adaptive bandwidth varies depending on the position of the sample; that is, different analysis scales are taken at different positions. The adjacency matrix constructed by this method has both strong practical significance and high sample statistical significance.

The spatial autocorrelation analysis of toponym co-occurrence and search index shows that, in the information space, the spatial interaction pattern by taking information flow as the representation has correlation and heterogeneity and conforms to TFL and GSL in the physical space. In the spatial autocorrelation analysis, there are outliers inconsistent with the laws of geography, such as Guangdong in southern China in Figure 1b and Jiangsu in southeastern China in Figure 2b. These two provinces belong to economically developed provinces in China, whose GDP consistently ranks among the top two in China. Frequent economic exchanges have significantly increased the interaction between these two provinces compared to the surrounding areas, making it an abnormally high value in spatial autocorrelation analysis. In the local Moran's I and cold-hot spot analysis that reflects spatial clustering, in addition to the differences in significance, most provinces match cold spots with low-value clusters and hot spots with high-value clusters. However, Inner Mongolia exhibits an L–H outlier in the local Moran's I in Figure 1, while it is a hot spot in the cold-hot spot analysis in Figure 2, which is not in line with expectations. In fact, due to TFL, provinces around Inner Mongolia that are not significantly clustered still have high interaction values, some even higher than that of Inner Mongolia. Thus, Inner

Mongolia and surrounding provinces have formed an L–H outlier. In the cold-hot spot analysis, due to the local situation is compared with the global situation, Inner Mongolia has become a significant hotspot.

In the distance decay effect test, the distance decay coefficients obtained by toponym co-occurrence and search index are smaller than those in the actual geographic space, generally 0.85 to 1.97 [39–41]. It shows that compared with the geographic space, the distance has less resistance to the interaction relationship for information flow, and the interaction strength decreases slowly with increasing distance. The reason is that the influence of the distance factor in the virtual information space is smaller than that in the actual geographic space. Therefore, the optimal distance decay coefficients obtained by toponym co-occurrence and search index are smaller than that in the geographic space, and the distance decay effect is relatively weak.

4.2. Spatial Interaction Pattern Based on Information Flow

The spatial pattern presented by the complex network based on multivariate information flow shows the regional interaction strength of China hierarchically and is highly consistent with the actual regional development pattern [42]. The previous conclusion of this research can also be proved from the perspective of positive research: the information space represented by information flow can be used as a mapping of geographic space in the physical world; the complex network based on information flow can be used as an information domain mapping of the interaction network of geographic entities in the real world. In comparison, the population mobility network is more influenced by individual demand factors, such as economic income and living comfort [43,44], while the information flow network is more influenced by social factors, such as social events and comprehensive development. The information flow network will reflect the social pattern more systematically and comprehensively.

In the centrality analysis of the interaction network, the ranking of each province is basically stable. Among the provinces with large changes in rank, it is worth noting that Hubei ranked significantly higher in 2020. The number of news mentioning Hubei in 2020 increased by 254.83% compared with 2019, of which news related to COVID-19 accounted for 47.24% of the total. The annual average value of the Baidu index for Hubei in 2020 increased by 55.55% compared to 2019, reaching a daily peak of 86,200 on January 25, which is 13.17 times the annual average value. On the same day, the Baidu index for “pneumonia” also reached a historical peak of 760,460. Figure 10 shows the subject words of co-occurrence news of Hubei and the top five provinces in co-occurrence intensity in 2020, among which words related to COVID-19, such as “epidemic”, “prevention”, “control”, “Wuhan” and “hospital”, are particularly prominent. COVID-19 was a major public health emergency in 2020, and related news reported the source and destination of cases in detail. COVID-19 in China broke out in Wuhan, Hubei, and spread to all provinces across the country, which made Hubei establish a close relationship with the other provinces. Therefore, the importance of Hubei in the spatial interaction network increased significantly this year.

In the past decade, network centralization has shown an upward trend on the whole, and the hierarchy of the spatial interaction network has become increasingly obvious. It indicates that network cohesion continually converges towards the provinces in the central position, and the degree of interprovincial interaction shows a strong hierarchical characteristic. On the one hand, the provinces in the central position of the network have a stronger driving force than the marginal provinces. On the other hand, the network structure needs to be further optimized, and the provinces in the non-central position need more investment to help them develop to improve their interaction capabilities so as to alleviate the increasingly uneven spatial interaction.

the traffic flow has been greatly reduced, even in a low-value state for a long time. During the closure of the city, the population movement was unusual, but the information flow was unimpeded. In this case, the traffic flow data could not match the outbreak of the epidemic in real-time, but the news reports could. The information flow was more reflective of the epidemic development than the population movement. Moreover, since this major public health event was the focus of the whole society, related news reports were generated in large quantities and rapidly to report the epidemic situation to the public in time. A large amount of news provided richer information, which provides stronger support for analyzing the spread and provided of the epidemic through toponym co-occurrence and search index. Therefore, information flow can replace the traffic flow in the physical space. Based on the information flow network, we can carry out epidemic spread analysis and make epidemic prevention and control decisions, which is the significance and contribution of this research to reality.

From both theoretical and empirical perspectives, it can be concluded that the information space can be regarded as a mapping for the geographic space of the physical world, and the information flow network can be regarded as an information domain mapping for the interaction network of real geographic entities. On this basis, the interaction analysis of real-world geographic entities can be carried out as a supplement to the existing research methods for city interaction based on traffic flow, human migration, etc. In future research, the driving factors of spatial interaction can be further clarified by analyzing the information content, and the spatial interaction of different topics can also be analyzed by restricting different subject terms.

5. Conclusions

This research proposes a spatial interaction analysis method of geographical entities based on multivariate information flow. The research conducts a spatial autocorrelation analysis on the spatial interaction strength of 31 provincial-level administrative regions of China by using the toponym co-occurrence and search index data to verify the geographical law of information flow in text space. Furthermore, the research combines the toponym co-occurrence and search index to generate multivariate information flow data with subjective and objective comprehensive characteristics and analyzes the spatial pattern of the provincial geographic entities in China based on the complex network of multivariate information flow. The results show that the toponym co-occurrence and search index in the textual space have correlation and heterogeneity, conforming to TFL and GSL. The toponym co-occurrence and search index have a distance decay effect. The best distance decay coefficients are 0.189 and 0.186, respectively, which are significantly lower than that of physical flow. Furthermore, the research combines the toponym co-occurrence and search index to generate multivariate information flow data. The research analyzes the spatial pattern of the provincial geographic entities in China based on the complex network of multivariate information flow. The multivariate information flow network has a triangular primary network and a diamond-shaped secondary network, which has a high consistency with the actual regional development pattern. Additionally, the movement of the gravity center of multivariate information flow can reflect the spatial influence of real events. In the time series evolution of the interaction network, Beijing has always been in the leading position of the network. The importance of each province is in a stable state as a whole, and the network is becoming more hierarchical. Compared with the physical flow of population mobility, the distance decay effect of information flow is weaker. The diamond-shaped pattern of the information flow network consists of long-distance and cross-regional interaction edges, and the asymmetry of the interaction is stronger than that of population mobility.

Author Contributions: Conceptualization, Lin Liu; methodology, Lin Liu and Hang Li; validation, Lin Liu and Shuai Liu; data curation, Dongmei Pei; writing—original draft preparation, Hang Li and Dongmei Pei; writing—review and editing, Hang Li and Lin Liu; visualization, Hang Li; supervision, Lin Liu; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of Shandong Province, China (grant number: ZR2019MD034).

Data Availability Statement: The data presented in this study are openly available in FigShare at <https://doi.org/10.6084/m9.figshare.20937943>, accessed on 5 September 2022.

Acknowledgments: The map data is obtained from National Catalogue Service for Geographic Information (<https://www.webmap.cn/main.do?method=index>, accessed on 9 October 2020).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Liu, Y.; Wang, F.; Kang, C.; Gao, Y.; Lu, Y. Analyzing relatedness by toponym co-occurrences on web pages. *Trans. GIS* **2014**, *18*, 89–107. [CrossRef]
- Chen, Y.; Zhang, X. Research on the address names census and database building. *Bull. Surv. Mapp.* **2015**, *53*, 103–107. [CrossRef]
- Hill, L.L. *Georeferencing: The Geographic Associations of Information*; MIT Press: Cambridge, MA, USA, 2006.
- Dai, J.; Xie, F.; Na, K. Research on Network Characteristics of Information Flow of Cities Along the Grand Canal Based on Baidu Index. *Urban Dev. Stud.* **2022**, *29*, 7–13.
- Liu, Q.; Zhan, Q.; Liu, W.; Yang, C. Research on the characteristics of urban network association and spatial organization structure based on railway passenger flow in Hubei province. *J. Geo-Inf. Sci.* **2020**, *22*, 1008–1022. [CrossRef]
- Ye, X.; Gong, J.; Li, S. Analyzing asymmetric city connectivity by toponym on social media in China. *Chin. Geogr. Sci.* **2021**, *31*, 14–26. [CrossRef]
- Sheng, K.; Wang, Y.; Fan, J. Dynamics and mechanisms of the spatial structure of urban network in China: A study based on the corporate networks of top 500 public companies. *Econ. Geogr.* **2019**, *39*, 84–93. [CrossRef]
- Tu, J.; Luo, Y.; Zhang, Q.; Tang, S.; Wu, Y. Evolution of spatial pattern of economic linkages between cities since the 40th anniversary of reform and opening up. *Econ. Geogr.* **2019**, *39*, 1–11. [CrossRef]
- Zhi, L.; Li, R.; Fu, X.; Guo, F. Data mining method of hot-toponym and its co-occurrence in crowdsourcing text written by tourists. *Sci. Surv. Mapp.* **2016**, *41*, 144–151. [CrossRef]
- Zhong, X.; Gao, Y.; Wu, L. Extract core toponyms from web page text based on link analysis. *J. Geo-Inf. Sci.* **2016**, *18*, 435–442.
- Wang, X.; Zhang, R.; Zhang, Y. Toponym resolution based on geo-relevance and D-S theory. *Acta Sci. Nat. Univ. Pekin.* **2017**, *53*, 344–352. [CrossRef]
- Wu, J.; Feng, Z.; Zhang, X.; Xu, Y.; Peng, J. Delineating urban hinterland boundaries in the Pearl River Delta: An approach integrating toponym co-occurrence with field strength model. *Cities* **2020**, *96*, 102457. [CrossRef]
- Lin, J.; Li, X. Simulating urban growth in a metropolitan area based on weighted urban flows by using web search engine. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 1721–1736. [CrossRef]
- Hu, Y.; Ye, X.; Shaw, S.-L. Extracting and analyzing semantic relatedness between cities using news articles. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2427–2451. [CrossRef]
- Meijers, E.; Peris, A. Using toponym co-occurrences to measure relationships between places: Review, application and evaluation. *Int. J. Urban Sci.* **2019**, *23*, 246–268. [CrossRef]
- Zhong, X.; Liu, J.; Gao, Y.; Wu, L. Analysis of co-occurrence toponyms in web pages based on complex networks. *Phys. Stat. Mech. Its Appl.* **2017**, *466*, 462–475. [CrossRef]
- Hu, D.; Li, R.; Meng, Y.; Wu, H. China's urban network from the perspective of toponym co-occurrences in the news. *Geomat. Inf. Sci. Wuhan Univ.* **2020**, *45*, 281–288. [CrossRef]
- Zhang, W.; Thill, J.-C. Mesoscale Structures in World City Networks. *Ann. Am. Assoc. Geogr.* **2019**, *109*, 887–908. [CrossRef]
- Han, J.; Ming, Q.; Shi, P.; Luo, D. The structural characteristics and influencing factors of tourism information flow network in China based on Baidu index. *J. Shaanxi Norm. Univ. Sci. Ed.* **2021**, *49*, 43–53. [CrossRef]
- Liu, Y.; Liao, W. Spatial Characteristics of the Tourism Flows in China: A Study Based on the Baidu Index. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 378. [CrossRef]
- Huang, X.; Sun, B.; Zhang, T. The influence of geographical distance on the dissemination of internet information in the internet. *Acta Geogr. Sin.* **2020**, *75*, 722–735.
- Xu, Y.; Lu, L.; Zhao, H. Dynamic Evolution and Spatial Differences of Network Attention in Wuzhen Scenic Area. *Econ. Geogr.* **2020**, *40*, 200–210. [CrossRef]
- Zong, H.; Hao, L.; Dai, J. Study of Urban Network Structure in Chengdu-Chongqing Economic Circle Based on Baidu Index. *J. Southwest Univ. Sci. Ed.* **2022**, *44*, 36–45. [CrossRef]
- Guo, H.; Zhang, W.; Du, H.; Kang, C.; Liu, Y. Understanding China's urban system evolution from web search index data. *EPJ Data Sci.* **2022**, *11*, 20. [CrossRef]

25. Wei, S.; Pan, J. Resilience of Urban Network Structure in China: The Perspective of Disruption. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 796. [[CrossRef](#)]
26. Yu, Y.; Song, Z.; Shi, K. Network Pattern of Inter-Provincial Information Connection and Its Dynamic Mechanism in China: Based on Baidu Index. *Econ. Geogr.* **2019**, *39*, 147–155. [[CrossRef](#)]
27. Wang, N.; Chen, R.; Zhao, Y.; Zhong, S. The information space and the Balkans phenomenon of the Chinese provinces. *Econ. Geogr.* **2016**, *36*, 17–24. [[CrossRef](#)]
28. Chen, Y.; Li, K.; Zhou, Q.; Zhang, Y. Can Population Mobility Make Cities More Resilient? Evidence from the Analysis of Baidu Migration Big Data in China. *Int. J. Environ. Res. Public Health* **2022**, *20*, 36. [[CrossRef](#)]
29. Liu, Y.; Gong, L.; Tong, Q. Quantifying the Distance Effect in Spatial Interactions. *Acta Sci. Nat. Univ. Pekin.* **2014**, *50*, 526–534. [[CrossRef](#)]
30. Wang, H.; Zhang, B.; Liu, Y.; Liu, Y.; Xu, S.; Zhao, Y.; Chen, Y.; Hong, S. Urban expansion patterns and their driving forces based on the center of gravity-GTWR model: A case study of the Beijing-Tianjin-Hebei urban agglomeration. *J. Geogr. Sci.* **2020**, *30*, 297–318. [[CrossRef](#)]
31. Zhou, R.; Liu, G.; Zhang, Y. Sustainability evaluation and spatial heterogeneity of urban agglomerations: A China case study. *Discov. Sustain.* **2021**, *2*, 1. [[CrossRef](#)]
32. Liu, Y.; Yao, X.; Gong, Y.; Kang, C.; Shi, X.; Wang, F.; Wang, J.; Zhang, Y.; Zhao, P.; Zhu, D.; et al. Analytical methods and applications of spatial interactions in the era of big data. *Acta Geogr. Sin.* **2020**, *75*, 1523–1538.
33. Tobler, W.R. A computer movie simulating urban growth in the detroit region. *Econ. Geogr.* **1970**, *46*, 234–240. [[CrossRef](#)]
34. Goodchild, M.F. GIScience, Geography, Form, and Process. *Ann. Assoc. Am. Geogr.* **2004**, *94*, 709–714.
35. Chen, M.; Liu, W.; Yeerken, W.; Gong, Y. The impact of the Belt and Road Initiative on the pattern of the development of urbanization in China. *Mt. Res.* **2016**, *34*, 637–644. [[CrossRef](#)]
36. Zhao, Z. How the west area of China is not far away: To accelerate the process of China's westward economic development. *West Forum* **2019**, *29*, 64–70. [[CrossRef](#)]
37. Liang, L.; Xian, Y.; Chen, M. Evolution Trend and Influencing Factors of Regional Population and Economy Gravity Center in China Since the Reform and Opening-up. *Econ. Geogr.* **2022**, *42*, 93–103. [[CrossRef](#)]
38. Wang, L.; Liu, H.; Liu, Q. China's city network based on Tencent's migration big data. *Acta Geogr. Sin.* **2021**, *76*, 853–869.
39. Li, B.; Gao, S.; Liang, Y.; Kang, Y.; Prestby, T.; Gao, Y.; Xiao, R. Estimation of regional economic development indicator from transportation network analytics. *Sci. Rep.* **2020**, *10*, 2647. [[CrossRef](#)]
40. Peng, H.; Du, Y.; Liu, Z.; Yi, J.; Kang, Y.; Fei, T. Uncovering patterns of ties among regions within metropolitan areas using data from mobile phones and online mass media. *GeoJournal* **2019**, *84*, 685–701. [[CrossRef](#)]
41. Zhao, Z.; Wei, Y.; Yang, R.; Wang, S.; Zhu, Y. Gravity model coefficient calibration and error estimation: Based on Chinese interprovincial population flow. *Acta Geogr. Sin.* **2019**, *74*, 203–221. [[CrossRef](#)]
42. Fan, J.; Wang, Y.; Liang, B. The evolution process and regulation of China's regional development pattern. *Acta Geogr. Sin.* **2019**, *74*, 2437–2454. [[CrossRef](#)]
43. Sheng, Y.; Yang, X. Spatial patterns and mechanisms of the floating population agglomeration among top three city clusters in China. *Popul. Econ.* **2021**, 88–107.
44. Zhang, W.; Yan, J.; Nie, G. Evolution of the pattern of China's urban population flows and its proximate determinants. *Chin. J. Popul. Sci.* **2021**, *35*, 76–87.
45. Wang, J.; Xu, J.; Xia, J. Study on the spatial correlation structure of China's tourism economic and its effect: Based on social network analysis. *Tour. Trib.* **2017**, *32*, 15–26. [[CrossRef](#)]
46. Zhao, Y.; Zhu, L.; Ma, B.; Xu, Y.; Jiang, B. Characteristics of inter-provincial network connection based on railway freight flow in China, 1998-2016. *Sci. Geogr. Sin.* **2020**, *40*, 1671–1678. [[CrossRef](#)]
47. Barabási, A.-L. *Network Science*; Cambridge University Press: Cambridge, UK, 2016.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.