

Research Article

Studying the Regional Cyberspace by Exploiting Internet Sequential Information Flows

Biao Jin ^{1,2}, Jin-ming Sha ¹, Jian-wan Ji ¹, Yi-su Liu ^{1,2} and Wu-heng Yang ¹

¹College of Geographical Sciences, Fuzhou, China

²College of Mathematics and Informatics, Fuzhou, China

Correspondence should be addressed to Biao Jin; jinbiao@fjnu.edu.cn and Jin-ming Sha; jmsh64@sina.com

Received 22 June 2018; Accepted 23 September 2018; Published 8 October 2018

Academic Editor: Stanislav Vitek

Copyright © 2018 Biao Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The study of cyberspace is faced with the challenge of the data shortage and model verification. This paper proposed a method to explore the regional cyberspace by employing Internet sequential information flows crawled from social network platforms. Compared with previous studies which only use one type of data sources for analysis, the main contribution of this manuscript is adopting the scheme that uses one kind of Internet information flow to extract cyberspace feature while relevant data collected from the other network platform is used for verification. Moreover, starting from measuring the informatization level of a region, a modified gravity model is designed by adding the value of informatization level to the traditional method. Then, an information association matrix based on the improved gravity model is constructed for analyzing the characteristics of cyberspace. To demonstrate the efficiency, Fuzhou city is considered as an interesting regional sample in this paper. The reasonable results indicate that the proposed approach is practical for regional cyberspace.

1. Introduction

By breaking through the limits of space and time, the relationships and interactive behaviors among humans have extended from the realistic geo-space to cyberspace. Cyberspace is the communication and information space created by computer, which is an abstract concept in the field of philosophy and computer science. It uses information flow as its studied data, while realistic geo-space is based on material flows.

The research of cyberspace has received various attentions. In the world, the research on cyberspace mainly focuses on three aspects: (1) cyberspace security; (2) the access control mechanism and communication protocol in cyberspace; (3) the study of the spatial network pattern based on sequence information in cyberspace. For example, Clark [1–7] designed trustworthy mechanisms and access control models to ensure the security of cyberspace. Chawki [8] tried to find the balance between privacy and security. Iyer [9] focused on “smart grid” and used cryptography and key management techniques to overcome some attacks to cyber security. Wechsler [10] advanced new directions for cyber security using adversarial learning and conformal prediction in order

to enhance network and computing service. Slonim [11] presented a novel sequential clustering algorithm which is motivated by the Information Bottleneck method. Prinzie [12] tried to overcome the inability to capture sequential patterns by modeling sequential independent variables by sequence-analysis methods. Mcculloch [13] and Mishra [14, 15] use the sequential information flow to diagnose the Swiss inflation in real time. Meanwhile, Tijsseling [16] presented a variant of Categorization-and-Learning-Module network, which is capable of categorizing sequential information with feedback. Copeland [17] studied the effect of sequential information arrival on asset prices. Mishra [18] utilized the Sequence and Set Similarity Measure with rough set based similarity upper approximation clustering algorithm to group web users based on their navigational patterns. Lottes [19] proposed a novel crop-weed classification system that relies on a fully convolutional network with an encoder-decoder structure and incorporates spatial information by considering image sequences. In mainland China, the study of cyberspace mainly focuses on two aspects: (1) relationship between cyberspace and realistic geo-space; (2) characteristics of cyberspace in specific areas. For instance, based on the

TABLE 1: Detailed information of the nine indicators.

Indexes	Indexes	Indexes
(1) Per capital GDP	(2) The number of express	(3) Income of express business
(4) Number of Colleges and Universities	(5) Length of optical cable line	(6) Number of IPv4 addresses
(7) Comprehensive coverage rate of TV programs	(8) The number of computers used by every hundred people	(9) The number of students in Colleges and Universities

Internet infrastructure, Wang [20–22] discussed the relationship between the Internet geographical structure and the Internet urban system of China. Bakis [23], Dong [24] and Sun [25] conduct a comprehensive analysis on the hierarchical structure and information flows pattern, as well as, reveal the spatial distributions of the Internet network structure of China. By linking the relationship between micro-blog users and geography, Zhen [26], Wang [27], and Chen [28] studied the centrality of nodes in the networks and the consistency of the whole network. Although Zhang [29–32] explored some methodologies of mining the relationship between geospace and cyberspace by using of information flows, it is difficult to obtain the sequential information of cyberspace in mainland China except cooperating with the data owners (usually government agencies). Therefore, researchers often obtain data from various statistical yearbooks and annual public reports. In this paper, we focus on the use of sequence information crawled from social networks to analyze the pattern and characteristics of cyberspace.

Currently, there are two main shortcomings in the prior researches of cyberspace: (1) many scholars conducted their studies only based on one kind of sequential information flow, which makes their conclusions less convincing; (2) when they explore the linkages among studied regions, they often directly use the classical gravity model which ignores the attributes of the studied region itself. In order to fix the first weakness, this paper adopts the scheme that uses one kind of internet information flow (data of Sina micro-blog) to extract cyberspace feature while relevant data collected from the other network platform (Baidu Index) is used for verification. Aiming at the second shortcoming, a method is firstly proposed to measure the informatization level of a region and then the classical gravity model is improved by introducing some attributes of the studied regions themselves; finally, an information association matrix is built based on the improved gravity model. By inputting the information association matrix into the network analysis tools (e.g., UCINET software) and selecting appropriate evaluation indicator (e.g., the degree centrality of nodes), the important nodes in the network space can be detected.

In order to explore the efficiency, Fuzhou city, the capital of Fujian province, is considered as an interesting region for approach verification. Specifically, firstly, we use a crawler to grab information about Sina micro-blog users, such as their registration addresses and other fundamental information. Then, for those users whose registration addresses are Fuzhou, we also grab the information in their concern lists and concerned lists to analyze their social relationships. Based on the obtained data, the intensity of active connection, passive connection, and total connection are used to study

the spatial pattern of cyberspace in Fuzhou city. To make the conclusions more convincing and more credible, the data from the Baidu Index is used for verification. According to the data collected from Baidu Index, we can get the number of times that one research unit is retrieved by another. Finally, we explore some possible factors which may have impacts on the pattern and characteristics of cyberspace.

2. Measurement Method of Informatization Development Level of Provinces in Mainland China

The regional informatization level is one of the most important factors that may affect the spatial pattern of cyberspace, so we propose a method for measuring the informatization level of a province in this section.

In this paper, we selected some indicators that can well reflect the regional informatization level according to the following steps.

Step 1. Obtain 186 indicators from China National Information Center which can be used for describing the development level of information society.

Step 2. Use world cloud analysis tools for counting the frequency of keywords contained in the 186 indicators. Then, 42 indicators that contain the keywords with higher frequency remain.

Step 3. The correlation coefficient and variation coefficient of each of the 42 indicators are calculated. Then, 17 indicators with poor correlation or high redundancy are eliminated.

Step 4. KMO test and factor analysis are carried out on the remaining 25 indicators. Finally, 9 indicators (shown in Table 1) remain and will be used to evaluate the level of information development in a region.

After that, Standard Deviation method, CRITIC (Criteria Importance through Inter Criteria Correlation), and Entropy Weight method are used for calculating the weight of each indicator for each province. The calculation formulas of each method are shown as follows.

(1) *Standard Deviation Method*

$$\delta_j = \sqrt{\frac{\sum_{i=1}^n (v_i - \bar{v})^2}{n}} \quad (1)$$

$$W_{j-SD} = \frac{\delta_j}{\sum_{j=1}^n \delta_j}$$

where W_{j_SD} is the weight of indicator j in indicator system, n stands for the number of research units, v_i represents the specific value of index i , and \bar{v} stands for the arithmetic mean of index j .

(2) CRITIC (Criteria Importance through Inter Criteria Correlation) Method

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}$$

$$C_j = \delta_j \sum_{i=1}^n (1 - r_{ij}) \quad (2)$$

$$W_{j_CRITIC} = \frac{C_j}{\sum_{j=1}^n C_j}$$

where δ_j and n have the same meanings in formula (1) and r_{ij} is the correlation coefficient between indicator j and indicator i .

(3) Entropy Weight Method

$$P_i = \frac{d_{ij}}{\sum_{i=1}^m d_{ij}}$$

$$E_j = -(\ln m)^{-1} \sum_{i=1}^m P_i \ln P_i \quad (3)$$

$$W_{j_EW} = \frac{(1 - E_j)}{(n - \sum_{j=1}^n E_j)}$$

where d_{ij} stands for the value of indicator j being normalized and n have the same meanings in formula (1).

Then, W_j will be used as the final weight of the indicator j . The value of W_j can be calculated using formula (4).

$$W_j = \frac{(W_{j_SD} + W_{j_CRITIC} + W_{j_EW})}{3} \quad (4)$$

Finally, the composite scores of information for each province on these indicators could be calculated according to the following formula.

$$Score_i = \sum_{j=1}^9 (W_j * V_{p_i}^j) \quad (5)$$

where $V_{p_i}^j$ represents the value of the indicator j of province i , and $i = 1, 2, 3, \dots, 31$.

3. Construction of Information Association Matrix Based on Improved Gravity Model

Gravity model is a widely used model for measuring spatial interaction capability; its formula is shown as follows:

$$G_{ij} = k \frac{M_i * M_j}{(d_{ij})^c} \quad (6)$$

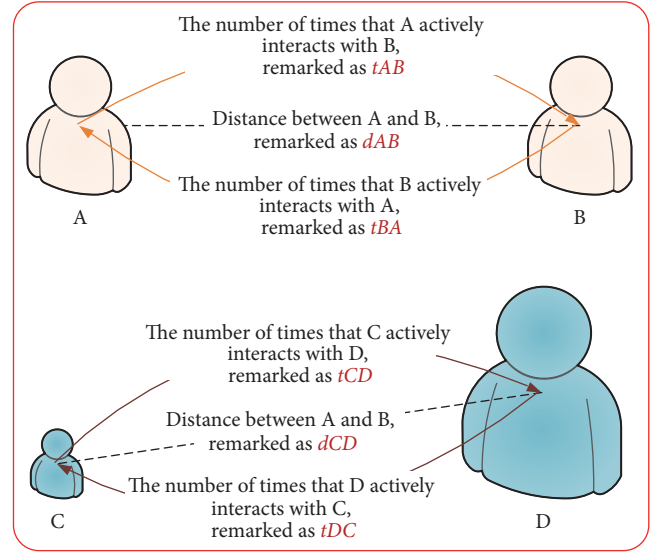


FIGURE 1: The interpretation of relevant variables in the scenario assumption.

where d_{ij} stands for the distance between unit i and unit j and k and c represent the coefficient of gravity and the distance attenuation coefficient, respectively, while the meaning of M_i (M_j) is varying in different applications. For example, if we study the intensity of communication between two regions, the meaning of M_i (M_j) can be the number of calls made by the mobile users in the two regions.

When using the above model, researchers usually only focus on the connection between nodes, but ignore the attributes of nodes.

Assume the following cases (Figure 1).

(1) The interactions between A and B and the interactions between C and D occur at the same time periods.

(2) The number of times that A actively interacts with B is the same as the times that B actively interacts with A, such that $t_{AB} = t_{BA} = 10$.

(3) The number of times that C actively interacts with D is different from the times that D actively interacts with C, such that $t_{CD} \neq t_{DC}$, $t_{CD} = 100$, $t_{DC} = 1$.

(4) The distance between A and B is the same as that between C and D ($d_{AB} = d_{CD}$).

Then, if we use the classical formula of gravity model, we would get the conclusion that the interaction intensity between A and B is the same as that between C and D. It is clearly incorrect, because it ignores the essential attributes of the research objects.

In view of the above analysis, and combined with the actual situation, in this paper, we modify the model as follows:

$$Y_{ij} = g_Y \frac{Score_i * Score_j}{d_Y^r} R_{ij} \quad (7)$$

where Y_{ij} stands for the intensity of information flow between province i and province j , g_Y is the gravity coefficient of information network, in which d_Y^r is the distance attenuation factor of information space, d_Y is the shortest road distance

TABLE 2: The value list of Social App in 2016.

Order Number	App's Name	MAU(million)	Growth Rate	Proportion of the high value users
1	Sina micro-blog	341.18	45.70%	76.30%
2	QQ Zone	92.60	-35.80%	70.20%
3	Momo	70.03	19.50%	68.70%
4	Tantan	15.89	128.90%	76.90%
5	OPPO paradise	11.87	-34.80%	53.50%
6	Her community	7.41	-29.00%	69.10%
7	SNOW	7.06	16692.60%	75.60%
8	Douban	5.18	-42.90%	89.70%
9	Idol	4.65	-4.50%	74.00%
10	Apploving	3.95	25.80%	79.30%

between province i and province j , and R_{ij} represents the intensity of network concern between province i and province j . In this paper, we use the parameter estimation method in Wang [31, 32] and set the values of g_Y and r to 0.85 and 1, while the value of d_Y can be obtained from Baidu map. We take the average number of searches between corresponding provinces from January-1-2016 to January-1-2017 as the value of R_{ij} . Then, for the sake of comparison, we standardize the value of Y_{ij} using formula (8).

$$Y_{ij}' = \frac{Y_{ij} - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (8)$$

Finally, we construct the information association matrix as follows:

$$M_Y = \begin{bmatrix} Y_{11}' & Y_{12}' & \cdots & Y_{1j}' & \cdots & Y_{1n-1}' & Y_{1n}' \\ Y_{21}' & Y_{22}' & \cdots & Y_{2j}' & \cdots & Y_{2n-1}' & Y_{2n}' \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ Y_{i1}' & Y_{i2}' & \cdots & Y_{ij}' & \cdots & Y_{in-1}' & Y_{in}' \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ Y_{n-11}' & Y_{n-12}' & \cdots & Y_{n-1j}' & \cdots & Y_{n-1n-1}' & Y_{n-1n}' \\ Y_{n1}' & Y_{n2}' & \cdots & Y_{nj}' & \cdots & Y_{m-1}' & Y_{nn}' \end{bmatrix} \quad (9)$$

In this matrix, for example, Y_{nj}' stands for the standardized intensity of information flow between province n and province j . Inputting this matrix into network analysis software, we can find out the important nodes in the network.

4. Research Object and Experimental Data

In this section, Fuzhou city, the capital of Fujian Province in mainland China, is used as an example of model verification. The sequential information flows which we grab from Sina micro-blog platform and Baidu website are applied to study the pattern and characteristics of the cyberspace in Fuzhou.

4.1. Brief Introduction of the Research Object. In May 6, 2009, the State Council of the People's Republic of China issued

the "opinions on supporting Fujian province to speed up the construction of the Economic Zone on the West Coast of China (EZWCC)." As an important part of China's coastal economic zone, the EZWCC is separated by the Taiwan Strait, north of the Yangtze River Delta and south of the Pearl River Delta. It occupies an important position in the layout of regional economic development. The EZWCC is composed of Fujian, Guangdong, Zhejiang, and Jiangxi provinces. Fujian province is the most closely related to Taiwan, because of their geographical proximities and historical and cultural similarities. With this unique advantage, Fujian province occupies the dominant position in the EZWCC.

As the capital of Fujian province, Fuzhou city is a hometown of overseas Chinese people. There are many overseas Chinese people with Fuzhou descent distributed all around the world. Fuzhou and Taiwan face each other across the Taiwan Strait. People in these two regions have close links with each other. As one of the important city nodes of the EZWCC, the development of Fuzhou has received attention from the local government and also the Chinese government.

In this section, Fuzhou is chosen as the studied region, and then the pattern and characteristics of its cyberspace are analyzed.

4.2. Principle Data Acquisition and Preprocessing. In order to analyze the cyberspace pattern of Fuzhou more accurately, two kinds of actual network information flows were chosen as principal experimental data.

4.2.1. Data about Sina Micro-Blog Users. Micro-blog is a completely open network interaction platform for public participation. The research of cyberspace based on micro-blog reveals the communication characteristic among people more clearly and reflects the influence of information on human relations networks more directly. Geographical attributes of micro-blog users provide the basis for the association of cyberspace and realistic geo-space.

According to the report of QuestMobile [33], the number of monthly active users (MAU) of Sina micro-blog has increased more than 45% and reached up to 341 million by the end of 2016. Table 2 shows that the MAU of Sina micro-blog is top-1. Considering the availability of Sina micro-blog users' location information, Sina micro-blog is chosen finally as the principle data source.

TABLE 3: The proportions of Internet users of all ages.

Age range	The proportions
age≤19	23.4%
20≤age≤29	30.3%
30≤age≤39	23.2%
age≥40	23.1%

In this paper, one thousand users' (OTU) information is grabbed. These users meet the following three conditions: (1) their registration addresses are Fuzhou; (2) they are ordinary users rather than celebrities or big V (the verified users who have more than 500,000 fans and use the microblog mainly for commercial or personal propaganda rather than sociality); (3) they are active users who not only are concerned about one hundred to five hundred fans but also concern one hundred to five hundred other users actively. According to the report of CNNIC [34], the proportions of Internet users of different ages at the end of 2016 are shown in Table 3. Accordingly, in order to make the sample more reasonable, the numbers of sampled users in these ranges are 234, 303, 232, and 231.

For the one thousand users, we not only grab their basic information (such as IDs, nicknames, sexes, registered addresses, character signatures, birthdays, marital status, and home links), but also obtain the registered addresses of the top 100 users in OUT's concern lists and concerned lists. Finally, 92015 users concerned about the OTU and 55449 users that concern the OTU are found. Among these users, there are 10451 pairs of friends.

There are three kinds of relationships between these users: active concern, passive concern, and mutual concern. Unilateral concern or be concerned model is a weak relationship, while the mutual concern model is a strong relationship. If user B is concerned about user A, then the direction of information flows can be described from A to B.

Three indicators are used to evaluate the intensity of the network information flows between Fuzhou city and other regions. These three indicators are the intensity of active connection (the number of concerns, X_a), the intensity of passive connections (the number of be concerned, X_b), and the intensity of total connections (sum of concern and be concerned, $X_a + X_b$). The meanings of the three indicators are shown in Table 4. All of the top 100 users in OUT's concern lists and concerned lists were classified and counted according to their registered addresses and relationships. The results are shown in Table 5.

In this paper, the network information flows between Fuzhou city and other provinces in mainland China are studied. To make X_a , X_b and $(X_a + X_b)$ comparable, maximum value standardization was carried out by using formulas (10).

$$X_a' = \frac{X_a}{\max(X_a)} * 100$$

$$X_b' = \frac{X_b}{\max(X_b)} * 100$$

$$(X_a + X_b)' = \frac{(X_a + X_b)}{\max(X_a + X_b)} * 100 \quad (10)$$

Because users in the concern lists and concerned lists are dominated by Fujian province, Fujian province is analyzed separately. When the values of $\max(X_a)$, $\max(X_b)$, and $\max(X_a + X_b)$ are selected, Fujian province is excluded.

4.2.2. Data about Baidu Index. The Baidu Index (<https://index.baidu.com/>) is one of the most important data sharing and statistical analysis platforms. It records a large amount of Internet users' behavior data. With its help, we can obtain the number of times that one specified keyword was retrieved in different areas and at different times, which can reflect the intensity of the connection between two regions to some extent [25, 35–39]. In this paper, the indexes between Fuzhou city and all the provinces in mainland China are obtained. The results are shown in Table 6.

There may be greater contingency if we use the number of passive retrievals to verify the intensity of passive concern or use the number of active retrievals to verify the intensity of active concern. So, the sum of passive retrieval and active retrieval was used for verifying the final intensity of total connections between two research units.

5. Spatial Pattern Analysis of Cyberspace in Fuzhou

5.1. Spatial Difference Analysis on the Intensity of the Network Information Flows

Cyberspace Breaks through the Limitation of Time and Geographical Space. Table 5 shows that the cyberspace in Fuzhou is very wide; the users either in the concern lists or in the concerned lists are distributed in all of provinces in mainland China. As a famous hometown of overseas Chinese people, there were also network information flows between Fuzhou users and some foreigners.

There Were Grade Differences in Cyberspace. Taking into account the actual situation of the object region and the data selected in this paper, all provinces in mainland China are divided into seven levels by the value of X_a , X_b , and $(X_a + X_b)$, and the provinces are divided into five levels by the value of X_a' , X_b' , and $(X_a + X_b)'$. The results are shown in Figures 2(a), 2(b), and 2(c). It is necessary to explain that the solid line with an arrow in Figures 2(a) and 2(b) shows the direction of concern. For example, “Fujian←Xinjiang” means that Xinjiang takes the initiative to pay attention to Fujian, which also means that Fujian is paid attention to by Xinjiang. The solid line without arrow in Figure 2(c) means the total intensity of attention between two units. In all of these three figures, the thicker the solid line is, the stronger the attention is.

In general, there is an obvious grading phenomenon in the intensity of connections, such that, from the east to the west, the intensities become gradually weaker: (1) the

TABLE 4: Three indicators and their meanings.

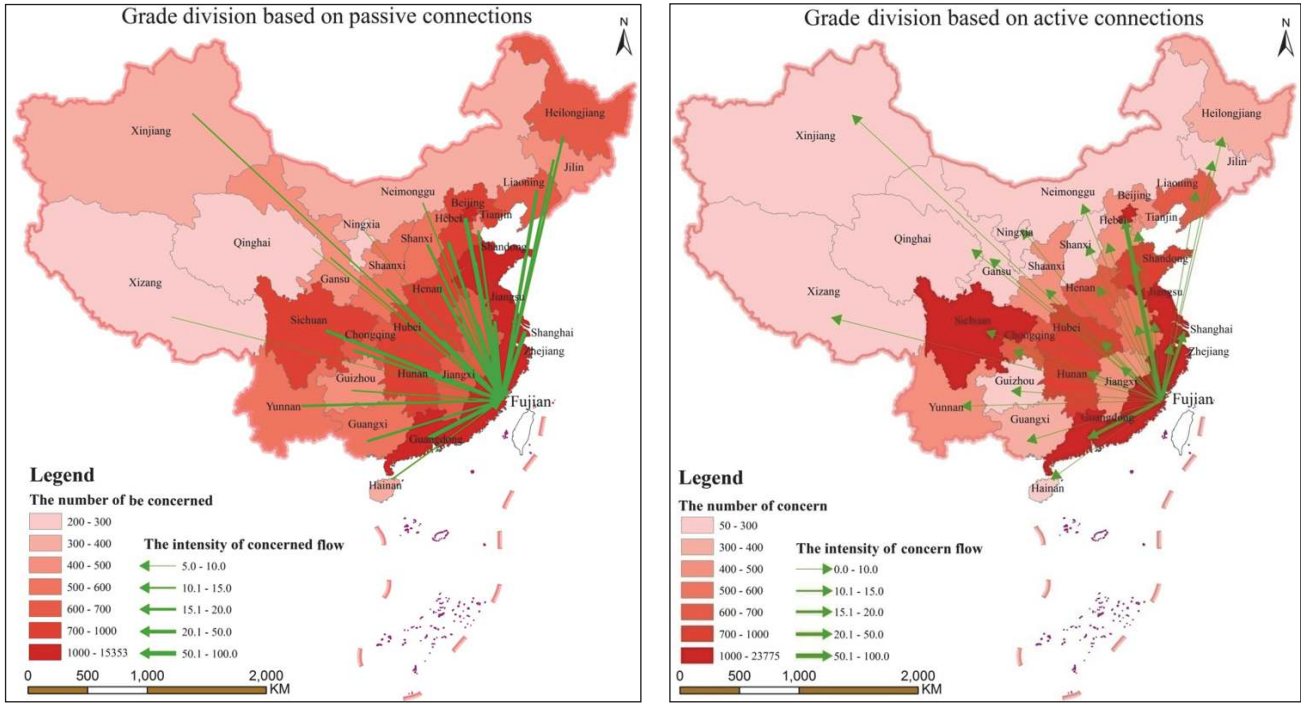
Indicators' names	Meanings of indicators
the intensity of active connection (X_a)	used to measure the active connection between Fuzhou and other research units; the greater the value means the research units have greater impacts on Fuzhou
the intensity of passive connection (X_b)	used to measure the interest of other research units in Fuzhou; the larger the value indicates that the units are more willing to accept the information from Fuzhou
the intensity of total connection ($X_a + X_b$)	used to measure the intensity of the total linkage between Fuzhou and other research units; the larger the value, the greater the intensity of interaction

TABLE 5: Classification statistics of the users in provincial level.

Provinces	Concern (X_a)	Be Concerned (X_b)	Total ($X_a + X_b$)	Provinces	Concern (X_a)	Be Concerned (X_b)	Total ($X_a + X_b$)
Fujian	21093	15353	36446	Jiangxi	406	655	1061
Beijing	23775	2592	26367	Shaanxi	458	597	1055
Foreign	7814	2866	10680	Yunnan	405	568	973
Guangdong	7450	3197	10647	Heilongjiang	324	602	926
Shanghai	6234	1702	7936	Tianjin	427	491	918
Zhejiang	3481	1429	4910	Guangxi	321	573	894
Jiangsu	2083	1361	3444	Shanxi	251	516	767
Hongkong	1740	485	2225	Jilin	255	493	748
Taiwan	1763	353	2116	Guizhou	187	423	610
Sichuan	1261	845	2106	Gansu	179	424	603
Shangdong	998	1064	2062	Neimenggu	182	378	560
Hubei	911	836	1747	Xinjiang	146	399	545
Hunan	786	846	1632	Hainan	154	350	504
Henan	7667	891	1558	Macao	138	225	363
Liaoning	635	693	1328	Ningxia	61	282	343
Chongqing	699	617	1316	Qinghai	61	277	338
Anhui	575	679	1254	Xizang	83	244	327
Hebei	478	721	1199				

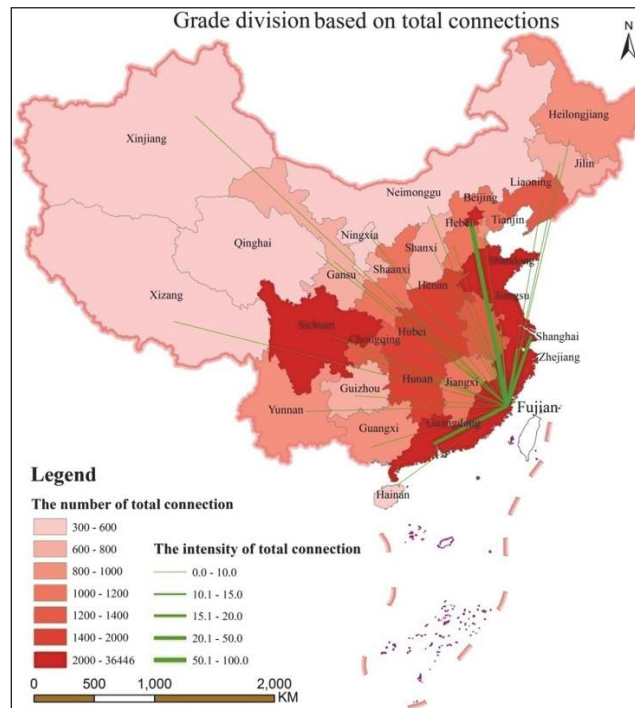
TABLE 6: Baidu indexes between Fuzhou and all the provinces in mainland China.

Unit A:Unit B	Baidu Indexes	Unit A:Unit B	Baidu Indexes	Total	Unit A:Unit B	Baidu Indexes	Unit A:Unit B	Baidu Indexes	Total
Fujian:Fuzhou	1793	Fuzhou:Fujian	524	2317	Guangdong:Fuzhou	653	Fuzhou:Guangdong	167	820
Shanghai:Fuzhou	417	Fuzhou:Shanghai	362	779	Beijing:Fuzhou	421	Fuzhou:Beijing	304	725
Zhejiang:Fuzhou	534	Fuzhou:Zhejiang	167	701	Jiangsu:Fuzhou	479	Fuzhou:Jiangsu	202	681
Sichuan:Fuzhou	406	Fuzhou:Sichuan	193	599	Shandong:Fuzhou	316	Fuzhou:Shandong	182	498
Anhui:Fuzhou	249	Fuzhou:Anhui	216	465	Hunan:Fuzhou	233	Fuzhou:Hunan	167	400
Shanxi:Fuzhou	240	Fuzhou:Shanxi	150	390	Tianjin:Fuzhou	186	Fuzhou:Tianjin	194	380
Hebei:Fuzhou	248	Fuzhou:Hebei	126	374	Henan:Fuzhou	316	Fuzhou:Henan	57	373
Liaoning:Fuzhou	222	Fuzhou:Liaoning	145	367	Guangxi:Fuzhou	186	Fuzhou:Guangxi	167	353
Jiangxi:Fuzhou	280	Fuzhou:Jiangxi	67	347	Hubei:Fuzhou	279	Fuzhou:Hubei	50	329
Hainan:Fuzhou	127	Fuzhou:Hainan	174	301	Jilin:Fuzhou	167	Fuzhou:Jilin	133	300
Chongqing:Fuzhou	180	Fuzhou:Chongqing	90	270	Ningxia:Fuzhou	87	Fuzhou:Ningxia	164	251
Heilongjiang:Fuzhou	197	Fuzhou:Heilongjiang	45	242	Shaanxi:Fuzhou	191	Fuzhou:Shaanxi	49	240
Qinghai:Fuzhou	75	Fuzhou:Qinghai	161	236	Yunnan:Fuzhou	167	Fuzhou:Yunnan	67	234
Guizhou:Fuzhou	158	Fuzhou:Guizhou	66	224	Xizang:Fuzhou	41	Fuzhou:Xizang	169	210
Neimenggu:Fuzhou	142	Fuzhou:Neimenggu	50	192	Gansu:Fuzhou	140	Fuzhou:Gansu	50	190
Xinjiang:Fuzhou	127	Fuzhou:Xinjiang	56	183					



(a) Passive connections

(b) Active connections



(c) Total connections

FIGURE 2: Grade classification at provincial level.

provinces with the highest intensity are mainly distributed in the southeast coastal area; (2) most of the central provinces are in the middle level; (3) the western provinces are at the lowest level.

Regional Embedding Still Exists in Cyberspace. First, the results of grade division show that although cyberspace breaks through the limit of time and geographical space, the geographic distance factor still has some influence on the

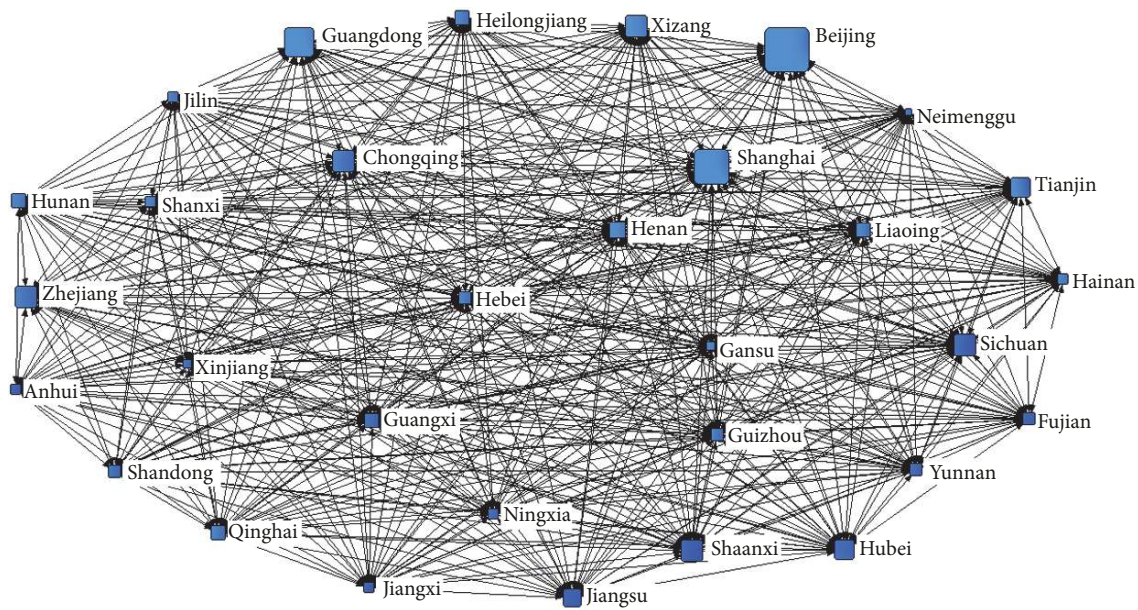


FIGURE 3: The information spatial association network diagram at provincial level in mainland China.

spatial pattern of cyberspace. The distance attenuation phenomenon also exists in cyberspace to some extent. Second, among these research units, Fujian province has the strongest connection with Fuzhou in the intensity of passive connection and total connection. Although information technology has compressed the space-time distance and expanded the scope of social communication, the information connections within the local domain occupied the dominant position in its cyberspace because of their geographical proximity and social cultural similarity.

The Information Flow in Cyberspace Is Asymmetric. “Information potential” refers to the capability to picking up, using, transmitting, formulating, aggregating, and processing information. The difference of “information potential” also leads to the asymmetry of information flow. Regions are more likely to establish contact with other regions with higher “information potential.” For example, Beijing, Shanghai, Zhejiang, and Jiangsu are far less concerned about Fuzhou than Fuzhou’s attention to them.

The Pattern of Cyberspace Can Be Affected by Population Flow (Labor Input and Output). Taking Sichuan province as an example, the distance between Sichuan and Fuzhou is much farther than that between Jiangxi and Fuzhou, but the connection intensity between the former two is stronger than that of the latter. These results are primarily because Sichuan is a big province of labor output in mainland China. Most of the laborers have been outputting to Fuzhou, Xiamen, Quanzhou, and some other cities in Fujian province. The flows of population will inevitably bring about information flows.

The Pattern of Cyberspace Has a High Correlation with Regional Economic Development Pattern. Excluding Fujian province,

strong connections occurred between Fuzhou and some economically developed areas, such as Zhejiang, Shanghai, Guangdong, Jiangsu, and Beijing. It means that the pattern of cyberspace is also affected by the level of economic development. It is primarily because the developed areas have higher “information potential.” The influence of “information potential” sometimes is even greater than that of geographic distance. Taking Beijing as an example, Fuzhou is concerned about Beijing much more actively than Fujian province.

Then, the information association matrix D_{ij}^Y is inputted into the UCINET software. If there is information connection between two provinces, there will be a connection line between them. In this way, the information spatial association network diagram at provincial level in mainland China is generated. In this paper, the degree centrality is used to evaluate the importance of nodes in the information network. If the degree centrality of nodes is higher, then the larger area of the graph is used to describe the node. Finally, we get Figure 3, which shows that economically developed provinces, such as Beijing, Shanghai, Guangdong, and Zhejiang, usually are the key nodes in the cyberspace. Other provinces with relatively sluggish economic development are willing to contact these key nodes more frequently.

5.2. Verifying the Effectiveness of Grade Division. All the analysis results in Section 5.1 are based on the grade division. Therefore, in this section, another kind of data (data from Baidu Index) is used to verify the effectiveness of grade division.

From Tables 5 and 6, the rankings of the number of total connections and total retrievals for each province can be obtained. The detailed rankings are shown in Table 7.

Figure 4 is visualized based on the information in Table 7. And then the correlation between the two discrete

TABLE 7: Total connection rankings and total retrieval rankings.

Provinces	Total connection rankings	Total retrieval rankings	Provinces	Total connection rankings	Total retrieval rankings
Anhui	14	9	Fujian	1	1
Guangdong	3	2	Guizhou	24	27
Hebei	15	13	Heilongjiang	19	23
Hunan	10	10	Jiangsu	6	6
Liaoning	12	15	Ningxia	29	22
Shandong	8	8	Shanxi	17	11
Sichuan	7	7	Xizang	31	28
Yunnan	18	26	Chongqing	13	21
Beijing	2	4	Gansu	25	30
Guangxi	21	16	Hainan	28	19
Henan	11	14	Hubei	9	18
Jilin	23	20	Jiangxi	16	17
Neimenggu	26	29	Qianghai	30	25
Shaanxi	22	24	Shanghai	4	3
Tianjin	20	12	Xinjiang	27	31
Zhejiang	5	5			

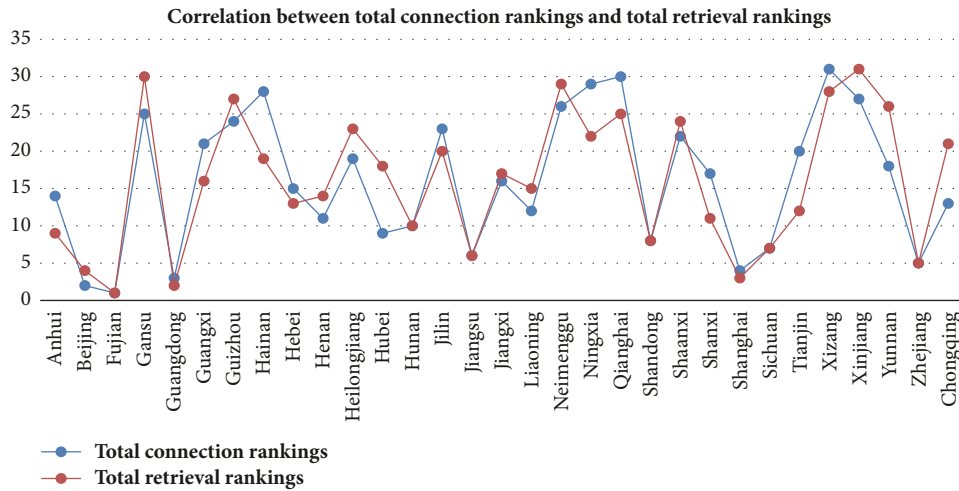


FIGURE 4: Correlation between total connection rankings and total retrieval rankings.

curves is calculated according to formula (11) and 87.10% is achieved.

$$Correl(X, Y) = \frac{\sum_{i=1}^{31} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{31} (x_i - \bar{x})^2 \sum_{i=1}^{31} (y_i - \bar{y})^2}} \quad (11)$$

Although there are little differences in some provinces' rankings by using the two different kinds of data, most of the provinces do not change too much in their rankings. The same conclusions can be achieved by comparing Figure 2(c) with Figure 5. Therefore, the conclusions obtained based on the first set of data in Section 5.1 have high credibility. What needs to be explained is that the provinces or cities in Figure 5 are classified according to the total intensity of mutual retrieval between them and Fuzhou city.

5.3. Analysis of the Possible Influential Factors. In this section, some influential factors which can be quantified and may have impacts on the pattern of cyberspace in Fuzhou are analyzed.

Geographic distance, the "Internet plus" index, and regional informatization development level are considered as the most likely influential factors that can affect the pattern of cyberspace. The "Internet plus" index is an important indicator to reflect the level of regional development. It consists of four subindexes: "Internet plus infrastructure," "Internet plus industry," "Internet plus innovation," and "Internet plus smart city." Further, these subindexes consists of 14 first-class indexes and 135 second-class indexes. Its content covers social, news, video, cloud computing, and the 19 major subindustries of the three industries. It uses the Tencent users' digital economic behavior as basic data and collects data

TABLE 8: Distances between Fuzhou and other provinces.

Research Unit	Distance(km)	Research Unit	Distance(km)	Research Unit	Distance(km)	Research Unit	Distance(km)
Zhejiang	554.7	Jiangxi	626.5	Shanghai	775.9	Guangdong	865.3
Anhui	875.0	Hunan	942.0	Jiangsu	1005.2	Hubei	1167.5
Henan	1295.7	Guangxi	1375.8	Shandong	1422.8	Guizhou	1560.0
Hainan	1587.1	Shanxi	1712.6	Hebei	1758.9	Chongqing	1760.7
Tianjin	1772.2	Shanxi	1812.9	Beijing	1887.1	Sichuan	2044.2
Yunnan	2098.7	Ningxia	2263.8	Liaoning	2406.9	Neimenggu	2468.2
Gansu	2547.2	Jilini	2694.9	Qinghai	2996.2	Heilongjiang	3277.6
Xizang	4097.6	Xinjiang	4375.1				

TABLE 9: “China Internet plus” indexes 2016 for each province.

Rankings	Provinces	“China Internet plus” indexes	Rankings	Provinces	“China Internet plus” indexes	Rankings	Provinces	“China Internet plus” indexes
1	Guangdong	18.072	2	Beijing	11.256	3	Shanghai	6.179
4	Zhejiang	5.512	5	Jiangsu	5.031	6	Fujian	3.967
7	Shandong	3.709	8	Sichuan	3.591	9	Henan	3.202
10	Chongqing	3.021	11	Hunan	2.884	12	Hubei	2.797
13	Hebei	2.579	14	Liaoning	2.239	15	Shanxi	2.153
16	Anhui	2.069	17	Guangxi	1.856	18	Jiangxi	1.764
19	Shanxi	1.702	20	Yunnan	1.586	21	Tainjin	1.491
22	Heilongjiang	1.45	23	Hainan	1.246	24	Jilin	1.23
25	Neimenggu	1.166	26	Guizhou	1.165	27	Xinjiang	0.933
28	Gansu	0.917	29	Ningxia	0.785	30	Qinghai	0.457
31	Xizang	0.35						

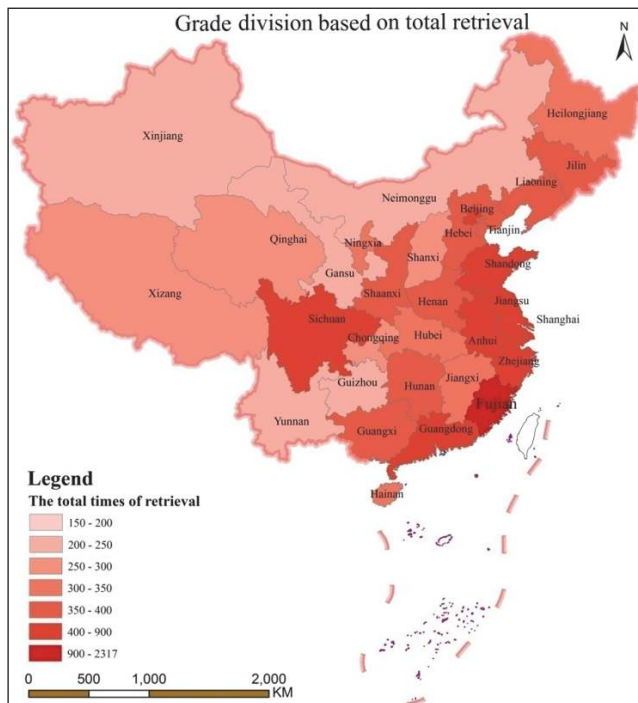


FIGURE 5: Grade division based on total retrieval.

from Didi Taxi, Meituan Dianping, Jingdong Mall, Ctrip, and some other Internet companies. Because the data are very comprehensive, it can reflect the degree of combination of Internet and all walks of life and the ability of the Internet utilization.

Geographic distances between Fuzhou city and other research units are measured with the help of Baidu map, the value of “Internet plus” indexes and rankings for each research unit are obtained from T. R. Institute [40], and the informatization development level of each research unit can be calculated with the method that we proposed in Section 2. These factors are shown in Tables 8–10.

SPSS (Statistical Product and Service Solutions) software is used to analyze the correlation between rankings of the three factors and rankings of the connection intensities. The analysis results are shown in Table 11.

Table 11 shows that there are high correlations existing between the connection intensity and the “Internet plus” index, as well as between the connection intensity and information level. The “Internet plus” index has the highest positive correlation with intensities. Results of correlation analysis also illustrate that there is a certain correlation between the distance factor and the connection intensity. However, this factor is no longer the primary factor. Some other factors, such as the government’s economic policies and cultural

TABLE 10: Composite scores and rankings for each province.

Provinces	Composite Scores	Information level rankings	Provinces	Composite Scores	Information level rankings
Guangdong	0.8322	1	Shaanxi	0.2879	17
Jiangsu	0.614	2	Heilongjiang	0.2651	18
Zhejiang	0.5805	3	Jiangxi	0.2471	19
Shanghai	0.5223	4	Shanxi	0.2348	20
Beijing	0.4595	5	Guangxi	0.2314	21
Shandong	0.4173	6	Yunnan	0.2305	22
Sichuan	0.3704	7	Chongqing	0.2253	23
Hubei	0.3584	8	Jilin	0.2162	24
Hunan	0.3245	9	Xinjiang	0.1724	25
Fujian	0.3242	10	Gansu	0.1677	26
Henan	0.3232	11	Qinghai	0.1599	27
Hebei	0.3212	12	Guizhou	0.1527	28
Liaoning	0.3159	13	Ningxia	0.1474	29
Neimenggu	0.3133	14	Xizang	0.0971	30
Tianjin	0.3099	15	Hainan	0.0866	31
Anhui	0.3059	16			

TABLE 11: Pearson correlation in provincial level.

	Information level rankings	Distance rankings	“China Internet plus index” rankings	Connection intensity rankings
Information level rankings	1	.634	.874	.948
Distance rankings	.634	1	.681	.646
‘China Internet plus index’ rankings	.874	.681	1	.972
Connection intensity rankings	.948	.646	.972	1

similarities, can also affect the connections intensity, but they are difficult to quantify.

6. Conclusion

This paper points out the shortcomings of existing works in the field of cyberspace: (1) researchers conduct their studies only based on one kind of sequential information flow, which makes their conclusions less convincing; (2) the study of the cyberspace pattern based on the sequential information flow usually directly employs the classical gravity model but ignores the attributes of the studied objects themselves. To overcome these weaknesses, we hold the idea that it is necessary to use some different kinds of sequential information flows to analyze this problem. We also advocate that we should consider the attributes of our study objects themselves when the gravity model is used. Accordingly, we proposed a method for measuring the informatization level of a region and improved the classical gravity model by adding the score of informatization level to the classical formula. And then, we constructed an information association matrix based on the improved gravity model. Finally, we took Fuzhou city as our study object and focused on its cyberspace characteristics. Experiments in this paper are conducted on two kinds of sequential information flows, data about Sina micro-blog users and data of Baidu Index.

According to our experimental result, the following conclusions can be drawn: *First*, cyberspace breaks through

the limit of geographical distance and has a wider range of communication. *Second*, there is an obvious grade difference in cyberspace. From the east of China to the west, the intensities of the total connections decrease gradually. *Third*, the social communication mode in realistic geospace has also been brought into cyberspace, such that the local domain information still occupied the dominant position in cyberspace. *Fourth*, the economically developed provinces usually are the principle node in network information space. *Fifth*, the information flows in cyberspace are asymmetric. Areas with low information potential are easily attracted by the areas with higher information potential. Economically backward areas are more willing to establish active contacts with economically developed areas.

There are many factors that can affect the pattern of cyberspace. By conducting comprehensive analysis of these factors, the characteristics and future of cyberspace could be understood and grasped more accurately.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Thanks are due to Professor Jin-bo Xiong, Professor Jie Lin, and Professor Xuan Li for their valuable comments. Also thanks are due to Professor Yu-e Jiang and AJE for the English language revision, which greatly improved the original manuscript. This work is supported in part by the EU Erasmus+ Project (NO. 586037-EPP-1-2017-1-HU-EPPKA2-CBHE-JP), the project of Fujian Provincial Department of the Science and Technology (NO. 2018I0005), and the project of Fujian Provincial Department of Education (NO. JAT170149).

References

- [1] D. D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden, "Tussle in cyberspace: Defining tomorrow's internet," *IEEE/ACM Transactions on Networking*, vol. 13, no. 3, pp. 462–475, 2005.
- [2] Q. Feng, Y. L. Sun, L. Liu, Y. Yang, and Y. Dai, "Voting systems with trust mechanisms in cyberspace: Vulnerabilities and defenses," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 12, pp. 1766–1780, 2010.
- [3] R. Deibert, J. Palfrey, R. Rohozinski et al., "Access Contested: Security, Identity, and Resistance in Asian Cyberspace," *Journalism & Mass Communication Quarterly*, vol. 90, no. 4, pp. 802–803, 2013.
- [4] J. Holliman, M. Zhivich, R. Khazan, A. Swiston, and B. Telfer, "Building low-power trustworthy systems: Cyber-security considerations for Real-Time Physiological Status Monitoring," in *Proceedings of the 35th IEEE Military Communications Conference, MILCOM 2016*, pp. 1083–1089, USA, November 2016.
- [5] F. Li, Z. Li, W. Han, T. Wu, L. Chen, and Y. Guo, "Cyberspace-Oriented Access Control: Model and Policies," in *Proceedings of the 2nd IEEE International Conference on Data Science in Cyberspace, DSC 2017*, pp. 261–266, China, June 2017.
- [6] L.-X. Yang, P. Li, X. Yang, and Y. Y. Tang, "Security evaluation of the cyber networks under advanced persistent threats," *IEEE Access*, vol. 5, pp. 20111–20123, 2017.
- [7] D. Wu, S. Si, S. Wu, and R. Wang, "Dynamic Trust Relationships Aware Data Privacy Protection in Mobile Crowd-Sensing," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2958–2970, 2018.
- [8] M. Chawki, "Anonymity in cyberspace: finding the balance between privacy and security," *International Journal of Technology Transfer & Commercialisation*, vol. 9, no. 9, pp. 183–199, 2010.
- [9] S. Iyer, "Cyber Security for Smart Grid, Cryptography, and Privacy," *International Journal of Digital Multimedia Broadcasting*, vol. 2011, Article ID 372020, 8 pages, 2011.
- [10] H. Wechsler, "Cyberspace security: How to develop a security strategy," *Intelligent Information Management*, vol. 07, no. 4, pp. 195–222, 2015.
- [11] N. Slonim, N. Friedman, and N. Tishby, "Unsupervised document classification using sequential information maximization," in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '02*, pp. 129–136, 2002.
- [12] A. Prinzie and D. Van den Poel, "Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM," *Decision Support Systems*, vol. 42, no. 2, pp. 508–526, 2006.
- [13] A. McCulloch, G. Glover, and T. John S, "Sequential Information Flow and Real-Time Diagnosis of Swiss Inflation: Intra-Monthly DCF Estimates for a Low-Inflation Environment," *Working Papers*, vol. 91, no. 91, pp. 1626–1634, 2004.
- [14] R. Mishra, P. Kumar, and B. Bhasker, *Design of Recommender system Incorporating sequential information using Web Usage Data*, Imrdc, Bangalore, India, 2011.
- [15] R. Mishra, P. Kumar, and B. Bhasker, "A design framework for recommender system by incorporating sequential information," in *Proceedings of the International Conference on Information Integration and Web-Based Applications and Services*, pp. 387–390, ACM, Ho Chi Minh City, Vietnam, December 2011.
- [16] A. G. Tijsseling, "Sequential information processing using time-delay connections in ontogenic CALM networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 16, no. 1, pp. 145–159, 2005.
- [17] T. E. Copeland and D. Friedman, "The Effect of Sequential Information Arrival on Asset Prices: An Experimental Study," *The Journal of Finance*, vol. 42, no. 3, pp. 763–797, 1987.
- [18] R. Mishra, P. Kumar, and B. Bhasker, "An alternative approach for clustering web user sessions considering sequential information," *Intelligent Data Analysis*, vol. 18, no. 2, pp. 137–156, 2014.
- [19] P. Lottes, J. Behley, A. Milioto et al., "Fully Convolutional Networks with Sequential Information for Robust Crop and Weed Detection in Precision Farming," *IEEE Robotics & Automation Letters*, 2018.
- [20] W. Mingfeng and N. Yuemin, "The urban geography of cyberspace: review and prospect," *American Behavioral Scientist*, vol. 17, no. 6, pp. 855–863, 2002 (Chinese).
- [21] M. Wang and Y. Ning, "The internet and the rise of information network cities in China," *Acta Geographica Sinica*, vol. 59, no. 3, pp. 446–454, 2004 (Chinese).
- [22] W. Mingfeng and N. Yuemin, "The network advantage of cities: an analysis of spatial structure and node accessibility of internet backbones in china," *Geographical Research*, vol. 25, no. 2, pp. 193–203, 2006 (Chinese).
- [23] H. Bakis and Z. Lu, "The change from the geographical space to geocyberspace -review on the western scholars on regional effects by telecommunication," *Dili Xuebao/Acta Geographica Sinica*, vol. 55, no. 1, pp. 104–111, 2000 (Chinese).
- [24] D. Zhiliang, L. Zi, and B. Cuiling, "The special character of Chinese networks frameworktake cernet as an example," *Geo-information Science*, vol. 7, no. 3, pp. 5–11, 2005 (Chinese).
- [25] Y. Sun, Z. Honglei, and L. Peixue, "Forecast of tourism flow volume of tourist attraction based on degree of tourist attention of travel network: A case study of Baidu index of different clients," *Human Geography*, vol. 32, no. 3, pp. 152–160, 2017 (Chinese).
- [26] F. Zhen, B. Wang, and Y. Chen, "China's city network characteristics based on social network space: an empirical analysis of Sina micro-blog," *Dili Xuebao/Acta Geographica Sinica*, vol. 67, no. 8, pp. 1031–1043, 2012 (Chinese).
- [27] B. Wang, Z. Feng, and X. Guangliang, "A study of cyber-geography based on micro- blog users relationship with a case of sina micro-blog," *Geographical Research*, vol. 32, no. 2, pp. 380–391, 2013 (Chinese).
- [28] C. Yingxue, Z. Feng, and B. Wang, "Chinese city network structure in the cyberspace based on social network analysis," *Economic Geography*, vol. 33, no. 4, pp. 56–63, 2013.

- [29] Z. Nannan and G. Chaolin, "From geographical space to composite space the urban space under the influence of information networks," *Human Geography*, vol. 17, no. 4, pp. 20–24, 2002 (Chinese).
- [30] Z. Jie, G. Chaolin, and D. Jinkang, "Geographical approach to cyberspace—review and prospect," *Scientia Geographica Sinica*, vol. 20, no. 4, pp. 368–374, 2000 (Chinese).
- [31] W. Ningning, R. Chen, and Y. Zhao, "Analysis of the provincial information space network based on the internet information flow," *Geographical Research*, vol. 35, no. 1, pp. 137–147, 2016 (Chinese).
- [32] W. Ningning, R. Chen, Y. Zhao et al., "Grey Associative Analysis of Regional Population Space and Information Space Coupling in China," *Scientia Geographica Sinica*, vol. 36, no. 10, pp. 1474–1484, 2016 (Chinese).
- [33] QuestMobile, "App value list in 2016," Tech. Rep., QuestMobile, Beijing, China, 2017.
- [34] China Internet Network Information Center(CNNIC), "China statistical report on internet development," Tech. Rep., China Internet Network Information Center, Beijing, China, 2017.
- [35] L. Shan, Q. Rongxu, and C. Ling, "Cyberspace attention of tourist attractions based on baidu index:temporal distribution and precursor effect," *Geography and Geo-information Science*, vol. 24, no. 6, pp. 102–107, 2008.
- [36] L. Zhang, "Baidu-index based analysis of regional network attention: Taking the case of zhenjiang," *Library and Information Studies*, vol. 5, no. 1, pp. 40–47, 2012 (Chinese).
- [37] X. Lifang, Z. Feng, and B. Wang, "The research of the Yangtze river delta core areas city network characteristics based on baidu index," *Economic Geography*, vol. 33, no. 7, pp. 67–73, 2013 (Chinese).
- [38] J. Daliang, Y. Sun, and R. Hang, "Analyses on the city network characteristics of middle Yangtze urban agglomeration based on baidu index," *Resources and Environment in the Yangtze Basin*, vol. 24, no. 10, pp. 1654–1664, 2015 (Chinese).
- [39] K. Liu, T. Wang, Z. Yang et al., "Using Baidu Search Index to Predict Dengue Outbreak in China," *Scientific Reports*, vol. 6, 2016.
- [40] T. R. Institute, "China "internet plus" index 2016," Tech. Rep., Ten- cent Research Institute, Beijing, China, 2016.

