GeoJournal
https://doi.org/10.1007/s10708-021-10509-z

Using Geodetector to explore the factors affecting evolution of the spatial structure of information flow in the middle reaches of the Yangtze River urban agglomeration

Wensheng Zheng · Nanqiao Du · Qian Zhang · Xiaofang Wang

Accepted: 16 August 2021

Abstract This paper explores the influence of the real-world urban attributes on the information flow among cities. UCINET was used to analyze the evolution of the spatial structure of information flow (SSIF). The Geodetector is applied to identify the spatial stratified correlation between the SSIF and urban attributes. The authors choose indicators such as economy, scale and facilities as background indicators. Furthermore, based on the textual categorization of the Baidu urban demand map, we selected the factors that people search more frequently as the direct factors. We conduct empirical analysis with the middle reaches of the Yangtze River urban agglomeration (MYUA) as a typical case. Our results show that the gap between cities is gradually decreasing in the information flow space. In the initial stage, the dominant factors affecting information flow network are mostly background factors. With the development of infrastructure and the increase in internet users, the density of information network is gradually increasing, and the structure of urban network tends to flatten. Direct factors gradually become the dominant factor affecting the evolution of network. The interactions of multiple factors had a stronger impact on the information network than individual factors.

Keywords Spatial evolution of information flow · Geodetector · Baidu index · Urban agglomeration of the Yangtze River middle reaches

W. Zheng · N. Du · Q. Zhang · X. Wang (✉)
Key Laboratory for Geographical Process Analysis & Simulation of Hubei Province, Central China Normal University, Wuhan 430079, Hubei, China
e-mail: wangxiaofang@mail.ccnu.edu.cn

N. Du
e-mail: dunanqiao@sina.cn

W. Zheng · X. Wang
Academy of Wuhan Metropolitan Area, Hubei Development and Reform Commission & Central China Normal University, Wuhan 430079, Hubei, China

W. Zheng
Hubei Institute of Economic and Social Development, Central China Normal University, Wuhan 430079, Hubei, China
e-mail: zhengwensheng@mail.ccnu.edu.cn

Introduction

Ever since the 1990s, the rise of the internet has created large volumes of complex and instantaneous information flow between cities, urban agglomerations, metropolitan regions, and global cities formed rapidly, with intercity connections becoming more networked. Since then, the concept, space of flows, was proposed by Castells, whereas research in urban system moved in the direction of networking (Castells, 1996, 2010; Taylor et al., 2010). Space of flow has gradually become the main tool for researchers to understand regional space (Batty, 2013; Hu et al.,

Published online: 20 September 2021

2020). Research on the urban network in digital space has gained attention (Devriendt et al., 2011; Wang & Loo, 2019). The study of urban network promotes the understanding of regional space. However, the role of central place still exists, and the attribute differences among cities constitute different hierarchies within the region (Meijers, 2005; Zhen et al., 2019). The central place and the space of flow jointly promote regional development (Miller, 2004). Information flow among cities is shaped by specific social, cultural, and economic factors (Rutherford, 2011; Yu & Shaw, 2008). The attribute of cities in the real world affects the information flow among cities. Exploring the influence of urban attributes on urban connection is helpful for further understanding of regional development.

Thus, to expand knowledge of these issues, the aim of this paper is to explore the influence of urban attributes on the information flow among cities. To be specific, four questions are to be discussed: (1) how to reveal the evolution of the spatial structure of information flow; (2) how to choose the urban attributes that affect the information flow; (3) how to detect the impact of urban attributes on information flow; (4) how urban attributes affect the information flow among cities. To solve these issues, we considering the following aspects: (1) the data of the search engine were used to reveal the evolution of the SSIF; (2) based on the analysis of current studies and actual content from the search engine, the authors established the index system; (3) the Geodetector was introduced to identify the spatial correlation between the SSIF and urban attributes.

The remainder of this paper is organized as follows. Section Literature review concentrates on the literature review. Section Methodology describes the study area and data sources. Section Study area and data sources introduces several network measures and Geodetector and discusses how to build the index system for factors affecting the MYUA's information network. Section Characteristics of the evolution of the MYUA's spatial structure of information flow (SSIF) analyzes the evolution of the spatial structure of information flow. Section Geo-detection of the evolution of MYUA's SSIF reports the Geo-detection of the evolution of the spatial structure of information flow. Section Conclusion and discussion discusses the study and its implications for better understanding of the influence of the real-world urban attributes on the information flow among cities, and Sect. 7 concludes and points to future work.

Literature review

In the information age, studies on information flow have increased (Castells, 1996; Janc, 2015a; Tranos et al., 2013). A stream of research has discussed the representation of cities in digital space, especially the visibility, image and flows among cities in the digital space (Chen et al., 2017; Goodchild, 2007; Graham et al., 2013; Janc, 2015b; Jiao et al., 2018; Yuan et al., 2017; Zook & Graham, 2007). There are two approaches used in the analysis of information flow among cities (Devriendt et al., 2008). One assumes that cyber relations between cities are determined by physical infrastructure. The cyberplace approach analyzes the digital intercity linkages via physical infrastructure. The cyberspace approach analyzes the digital intercity linkages through digital communication and information. It describes the transfer of information and identifies the digital connections among cities based on hyperlinks, e-mail traffic and the structure of web browsers' search results (Janc, 2015b). To reveal the digital intercity linkages, in this article, the cyberspace approach is used. And social network analysis is applied to analyze the spatial structure of information flow.

Through editor/reporters' and internet users' experience, the user-created Web 2.0 digital space is affected by physical space (GRAHAM, 2010; Wang & Loo, 2019). Currently, the analysis of digital information flow involves theoretical discussion (Kinsley, 2014; Leszczynski, 2015), qualitative analysis (AlSayyad & Guvenc, 2015) and quantitative analysis (Graham et al., 2014; Wang & Loo, 2019). Among them, quantitative analysis is more subjective and single to the selection of influence factors. And the regression analysis is widely applied to measure the correlation among variables. To overcome the subjectivity and immeasurability of the constructed index system, in this study, we choose indicators such as economy, scale and facilities as background indicators, which serve as the basis. Furthermore, actual contents from the Baidu were used for in-depth exploration and analysis. In this study, we use the method Geodetector proposed by Wang (Wang et al., 2010) to estimate the association between information flow among cities and the real-world urban attributes. Compared to the regression analysis, the Geodetector is capable of detecting spatial stratified heterogeneity, revealing the driving factors behind it and identifying interaction among influence factors (Wang & Xu, 2017). This method has been widely applied in health risk assessment, housing price studies and earthquake risk assessment (Hu et al., 2011; Huang et al., 2014; Wang et al., 2010, 2017). The employment of Geodetector could provide valuable insights into the association between the real-world urban attributes and information flow among cities.

Baidu, as a widely used search engine in China, has become an important data source for studying cities (Zhang et al., 2017; Zhen et al., 2015). Based on Baidu index, a data sharing platform provided by Baidu, this research contributes via the following aspects: first, the evolution of spatial structure of information flow and its relationship with place-based characteristics of the city are discussed. Furthermore, we discuss the influence of urban attributes on information flow. Second, the Geodetector is applied to identify the spatial stratified correlation between the real-world urban attributes and digital connection between cities. Third, based on the background indicators, such as economy, scale and facilities, through the analysis of the urban demand map, we also selected the factors that people search more frequently as the direct factors. Fourth, current research mainly focuses on analyzing global cities and city systems within a country (Devriendt et al., 2011; Wang & Loo, 2019). The information network of urban agglomeration has been relatively poorly analyzed. Thus, we conduct empirical analysis with the middle reaches of the Yangtze River urban agglomeration (MYUA) as a typical case.

Methodology

This study used the social network analysis to study the spatial evolution of information flow from the perspectives of inflow and outflow. Textual categorization of the urban demand map was also performed to overcome the subjectivity and immeasurability of the constructed index system of influencing factors. Finally, Geodetector was used to analyze the spatial variation of the factors affecting the spatial structure of information flow (SSIF).

Metrics for the spatial structure of information flow

Degree centrality

Degree centrality measures the amount of information flow that a city generates from direct searches or being searched; it comprises the in- and out-degrees (Golbeck, 2013). In-degree refers to the amount of information that a city directly searches for about other cities; the out-degree is the amount of information that other cities directly search for about the particular city. These are expressed using the following formulas:

$$C_{\mathrm{Di}}(\mathrm{in}) = \frac{\mathrm{di}(\mathrm{in})}{k-1} \tag{1}$$

$$C_{\mathrm{Di}}(\mathrm{out}) = \frac{\mathrm{di}(\mathrm{out})}{k-1} \tag{2}$$

where k is the total number of urban nodes; C_{Di}(in) and C_{Di}(out) represent the standardized in-degree and out-degree of degree centrality, respectively; and di(in) and di(out) are the number of nodes directly pointing to the subject node and number of nodes that the subject node is directly pointing to, respectively.

Closeness centrality

Closeness centrality measures whether a city is free from the control of information of other cities and refers to the proximity of a city to all other cities within a network (Golbeck, 2013). The formula is as follows:

$$C_{\mathrm{RPi}}^{-1} = \frac{\sum_{i=1}^{k} d_{\mathrm{ij}}}{k-1} \tag{3}$$

where C_{RPi}^{-1} is the standardized closeness centrality and dij is the shortcut distance between nodes i and j.

Factor determination

SSIF is affected by certain social, cultural and economic factors. Studies have shown that positive correlations exist between urban scale, informatization level, and core strength in the evolution of information flow (Graham et al., 2015; Wang & Loo, 2019; Zhang et al., 2017). Since these three main factors are the basic components of a city, they were regarded as urban background factors in this study.

While the internet is user-created and internet users' focus in terms of information about the various cities is diverse (GRAHAM, 2010; Kwan, 2001), using only background factors can no longer fully explain the evolution of the spatial pattern of information flow within an urban agglomeration. So, in this study, actual contents from the Baidu Index were used for the in-depth exploration and analysis. This methodology led to further construction of the direct factors affecting the MYUA's information flow space.

The aims of the direct factors were to reflect the current situation in the development of the information flow space between cities in the MYUA and understand internet users' focus in terms of information about the various cities. The investigative process involved "crawling" through the top 10 related preceding and succeeding search terms for the urban agglomeration's 31 cities in the Baidu Index platform's demand map. A big data search was performed using Natural Language Processing and Information Retrieval (NLPIR), and sharing platforms were mined to analyze 7440 related search terms. This approach confirmed the types of information pertaining to the various nodes in the MYUA's information network that could be searched for directly.

First, a word frequency statistical tool was used to combine similar search terms and identify terms that occurred frequently. Searches and verifications were then conducted for related terms with low frequencies of occurrence (previously, those terms with no significance were screened out). Next, a keyword extraction tool was used to obtain the representative terms before a textual classification tool was used for preliminary categorizing them. Finally, these categories were adjusted and refined based on phenomena and empirical judgments to arrive at nine main categories of detection factors, which included urban medical and healthcare services, scientific and educational services, and cultural services. Table 1 shows the index system.

Studies on geo-detection of factors

The Geodetector is a method to detect spatial stratified heterogeneity and reveal the driving factors behind it (Wang & Xu, 2017; Wang et al., 2010). The basic assumption behind the Geodetector is that if an independent variable has a significant effect on a dependent variable, the spatial distribution of the independent variable and the dependent variable should be similar (Cao et al., 2013; Wang & Hu, 2012; Wang et al., 2010). Based on the proxy analysis and the spatial consistence analysis, the Geodetector is used to explore the influence of physical factors on SSIF. The Geodetector is composed of factor detector, ecological detector, risk detector, and interaction detector.

Factor detection

This approach probe the extent to which a factor X explains the dependent variable Y. To be specific, this method measures the intensity of impact that the physical factors have on the spatial evolution of information flow within an urban agglomeration. The detection method is given follows:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} \tag{4}$$

where *Nh* is the number of type h samples in factor X, n is the total number of samples in H (the MYUA), L is the classification number of factor X, and σ2 is the discrete variance of the MYUA.

When factor X has a decisive effect on the MYUA's information network structure, the discrete variance of each type will be small; however, the discrete variance between types will be large. The value interval of q is [0,1]: the larger the q value, the greater the impact that factor X has on the spatial distribution of information flow in the MYUA.

Ecological detection

This approach compares the impact of two physical factors on the index distribution of the SSIF and determines whether there is any significant difference. It is measured in terms of the *F* test:

$$F = \frac{N_{X1}(N_{X2} - 1)\mathrm{SSW}_{X1}}{N_{X2}(N_{X1} - 1)\mathrm{SSW}_{X2}} \tag{5}$$

where N_{X1} and N_{X2} represent the sample sizes of factors *X*1 and *X*2, respectively; and SSW_{X1} and SSW_{X2} represent the sums of the intra-layer variances formed by the *X*1 and *X*2 layers, respectively. The null hypothesis is H0: $\mathrm{SSW}_{X1} = \mathrm{SSW}_{X2}$. If H0 is rejected

Table 1 Index system for factors affecting the MYUA's information network

Dimension	Detection factors	Factor symbol	Specific indicators
Background factors	Urban equivalent scale	X1	Per capita GDP
		X2	Resident population
		X3	Administrative level
	Urban level of informatization	X4	Number of international Internet users
		X5	Number of mobile phone users
	Urban core strength	X6	GDP
		X7	Value added by secondary industries
		X8	Value added by tertiary industries
		X9	General budgetary expenditure of the local administration's finances
		X10	Deposit balance of financial institutions
Direct factors	Urban ecological environment	X11	Number of days for which air quality met the standard requirements
		X12	Total area of green spaces
		X13	Green coverage within built-up areas
	Urban geographic proximity	X14	Bordering/neighboring cities
		X15	Time cost of public roads
		X16	Time cost of railways
	Urban network communication level	X17	Media index
		X18	Number of major websites
	Urban economic development opportunities	X19	Average wage of employees
		X20	Urban registered unemployment rate
		X21	Total retail sales of social consumer goods
		X22	Investments in fixed assets
	Urban scientific and educational services	X23	Educational expenditure(s)
		X24	Number of students in general colleges and tertiary institutions
		X25	Number of students in secondary vocational and technical institutes
		X26	Value added from high-tech industries
		X27	Number of patent applications
		X28	Number of patents granted
	Urban cultural services	X29	Number of employees in cultural agencies
		X30	Total book collection in public libraries
	Urban medical and healthcare services	X31	Number of beds in hospitals
		X32	Number of doctors
	Urban housing level	X33	Real estate sales
		X34	Area of real estate sold
		X35	Per capita residential housing area
	Urban tourism appeal	X36	Income from domestic tourism
		X37	Number of scenic spots evaluated as Grade 4A or above

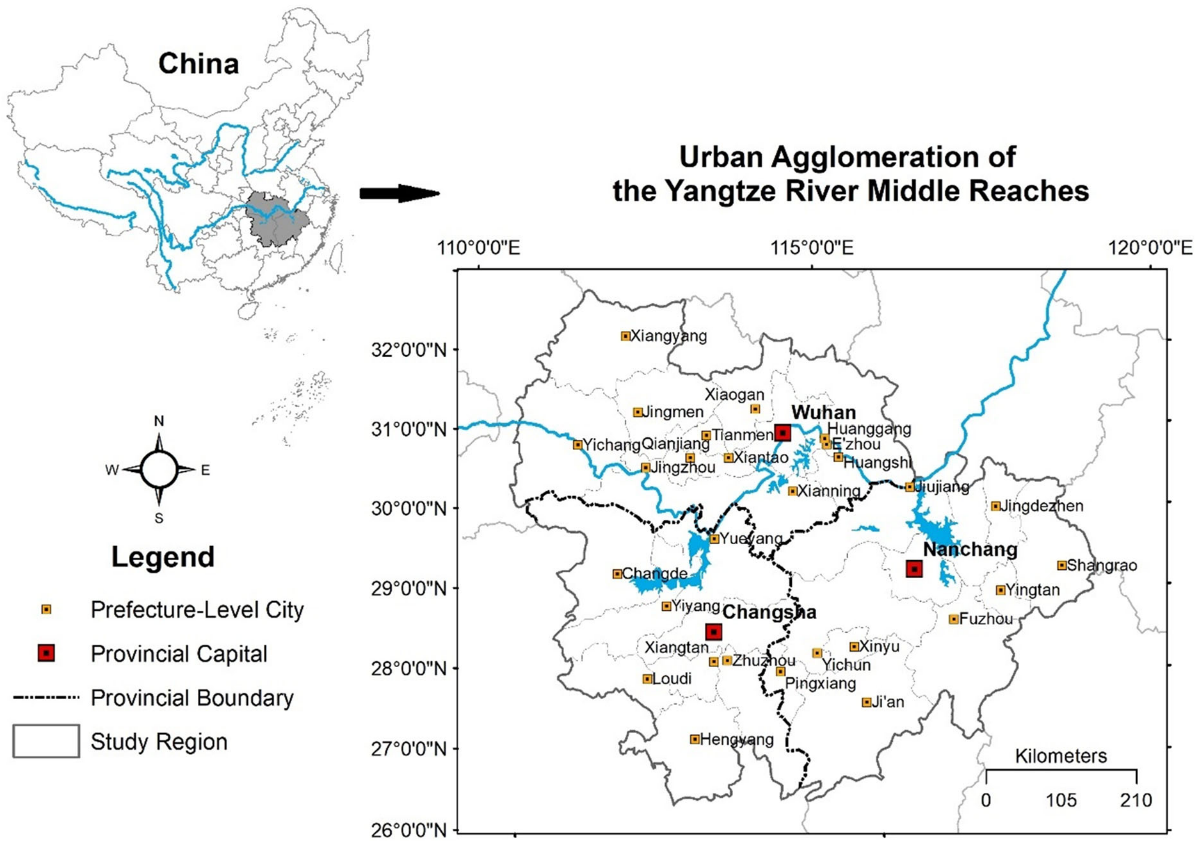


Fig. 1 Study area

at the α significance level, there is a significant difference between the impacts of $X1$ and $X2$ on the spatial distribution of attribute Y.

Risk detection

This method determines whether a significant difference exists between the mean attribute values of two sub-regions. The t test is used as shown as follows:

$$t_{\overline{y}_{h=1}-\overline{y}_{h=2}} = \frac{\overline{Y}_{h=1} - \overline{Y}_{h=2}}{\left[\frac{\mathrm{Var}\left(\overline{Y}_{h=1}\right)}{n_{h=1}} + \frac{\mathrm{Var}\left(\overline{Y}_{h=2}\right)}{n_{h=2}}\right]^{1/2}} \tag{6}$$

where $\overline{Y}_h$ represents the mean attribute value of sub-region h, n_h is the sample size of sub-region h, and Var represents the variance. The null hypothesis is H0: $\overline{Y}_{h=1} = \overline{Y}_{h=2}$. A significant difference exists between the mean attribute values of the two sub-regions if H0 is rejected at the α confidence level.

Interaction detection

Interaction detection is a quantitative characterization of the relationship between the effects of two physical factors on the SSIF within the MYUA. After comparing the impacts of factors in layers A and B versus that in layer C (formed by the superimposition of A and B), it can be determined whether the impact of the two-factor interaction effect and single-factor effect on the network pattern of information flow has strengthened or weakened.

The determining formulas are listed as follows:

- Nonlinear weakening: $q(X1 \cap X2) < \mathrm{Min}(q(X1), q(X2))$
- Single-factor nonlinear weakening: $\mathrm{Min}(q(X1), q(X2)) < q(X1 \cap X2) < \mathrm{Max}(q(X1) + q(X2))$
- Two-factor synergy: $q(X1 \cap X2) > \mathrm{Max}(q(X1), q(X2))$
- Mutual independence: $q(X1 \cap X2) = q(X1) + q(X2)$

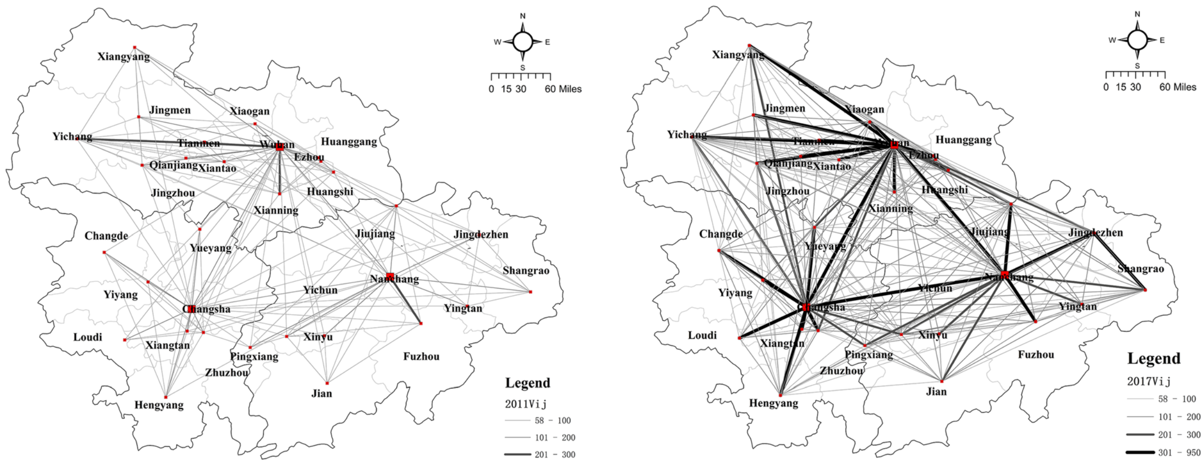


Fig. 2 Structure of the MYUA's network of information linkages in 2011 and 2017

- Nonlinear strengthening: q(X1 ∩ X2) > q(X1) + q(X2)

Study area and data sources

Study area

The study area is the urban agglomeration of the YRMR (26°03′–32°37′N, 110°15′–118°29′E), which contains a total of 31 cities distributed across the provinces of Hubei, Jiangxi, and Hunan, With a total land area of approximately 317,000 square kilometers (Fig. 1). The urban agglomeration encompasses three sub-agglomerations, namely urban clusters around Wuhan, the Changsha-Zhuzhou-Xiangtan city group, and clusters around Poyang Lake. It is also the region in central China with highly concentrated exchanges of socioeconomic information, as well as the focal region for implementing the "Strategy to Promote the Rise of the Central Region" and new forms of urbanization. The spatial stratified heterogeneity and network structure make the YRMR urban agglomeration a useful study area for determining the factors affecting evolution of the spatial structure of information flow.

Data sources

The trend module of the Baidu Index was used, with the time range established as 2011–2017. Based on keyword statistics, MYUA's cities *i* and *j* were respectively regarded as the keyword and region for obtaining city *j*'s search index for city *i*. The index was then used to construct a matrix for the intensity of the MYUA's information linkages. The basic data of the search index were the volume of keyword searches made by Internet users on the Baidu platform, and the frequencies of keyword searches were weighted and averaged.

The Baidu demand map/demand distribution was analyzed to determine the direct factors affecting the MYUA's information flow network. Demand distribution was obtained by comprehensively calculating the degree of correlation between keywords and related terms, as well as the search demand of the related terms. Data of the index system for the influencing factors were obtained from the China Urban Statistical Yearbook 2011 and 2016 and the statistical yearbooks for the various provinces.

Characteristics of the evolution of the MYUA's spatial structure of information flow (SSIF)

The mean of the Baidu Search Index between cities in 2011 and 2017 was used as the threshold value for selecting effective information linkages. The matrix for the intensity of information flow within the MYUA was binarized before the ArcGIS software was used for visualization. This process generated the network structure map of MYUA's information linkages in 2011 and 2017 (Fig. 2).

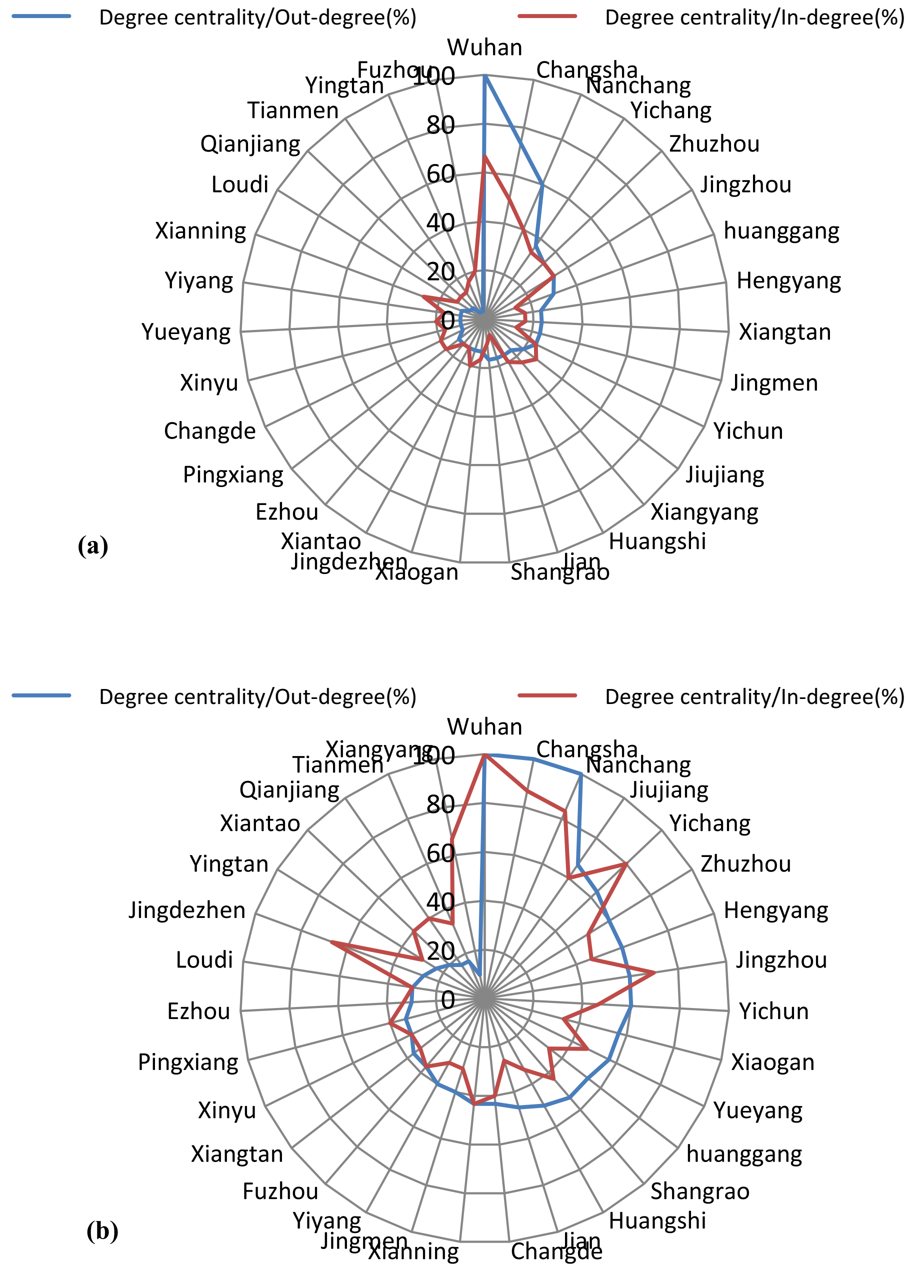


Fig. 3 Degree centrality of the MYUA's urban nodes in **a** 2011 and **b** 2017

The MYUA's spatial structure of information flow (SSIF) transformed from a loose structure of divergence and single centers to the stable structure of a polycentric network. There was a strengthening of direct information linkages between cities, and the inflow and outflow capacities of urban information also increased significantly. The MYUA's outflow-oriented network had transformed into a balanced network of information inflow and outflow.

As shown in Fig. 3, the degree centrality of the various MYUA cities in 2011 was more out- than in-degree. Information flow was highly concentrated in the three provincial capital cities of Wuhan, Changsha, and Nanchang. A loose structure radiating from those three cities as the centers (especially with Wuhan as

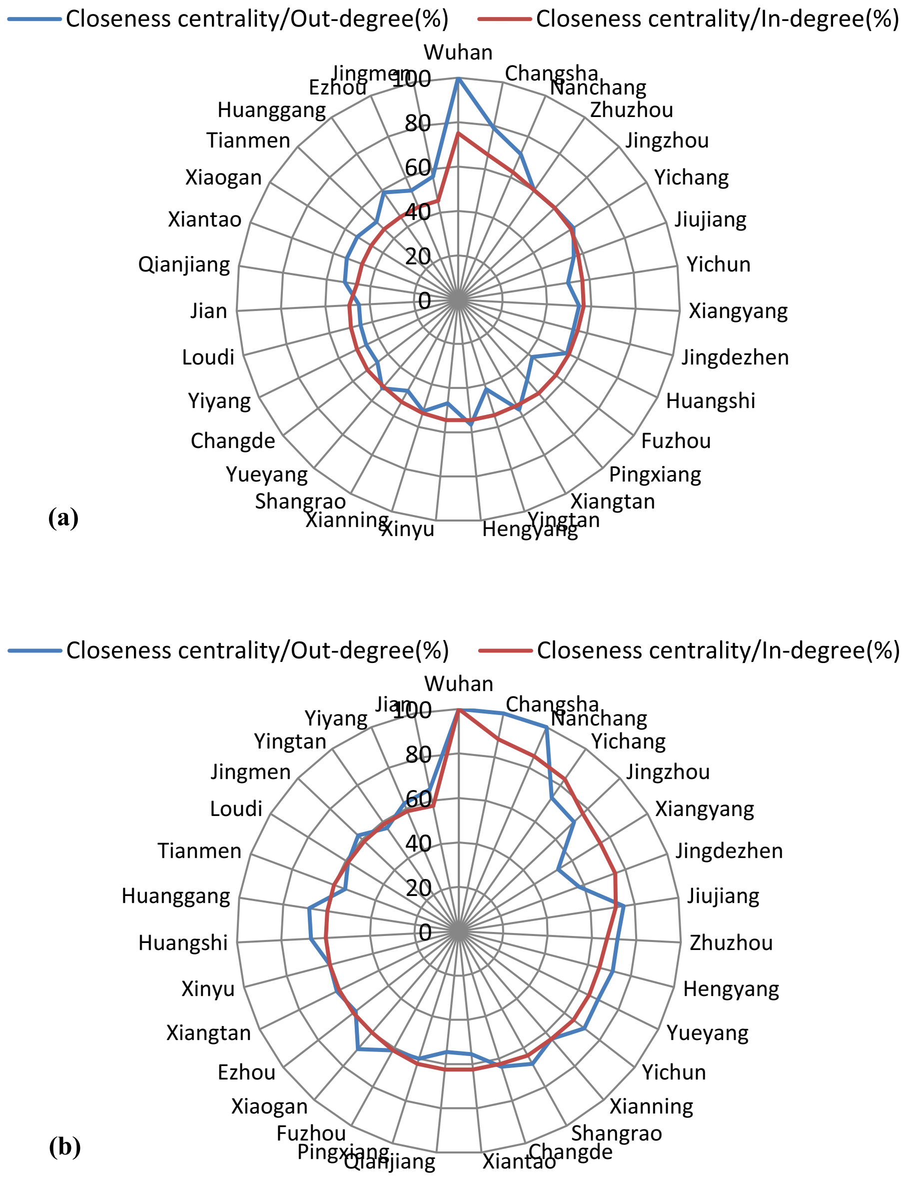


Fig. 4 Closeness centrality of the MYUA's urban nodes in **a** 2011 and **b** 2017

the core) was formed. By 2017, the in- and out-degrees of the MYUA's various cities had improved significantly. Apart from the three major provincial capital cities, the various small- and medium-sized cities gradually became more prominent in the information flow space. At the same time, the in-degree of the various cities increased, together with a significant reduction in the gap between the in- and out-degrees. The spatial consistency of the network was strengthened continuously, basically forming a stable SSIF in the form of a network comprising triangular clusters with net-like extensions.

The status of the MYUA's small- and medium-sized cities within the information flow space has increased continuously, and the overall connectivity of that space has become stronger. There has also been a general increasing trend in the closeness centrality of various cities (see Fig. 4). In 2011, the in-degree of urban nodes of an equivalent grade was less than the out-degree. By 2017, the respective internal gaps of

Fig. 5 Detection results for the MYUA's centrality factors in 2011

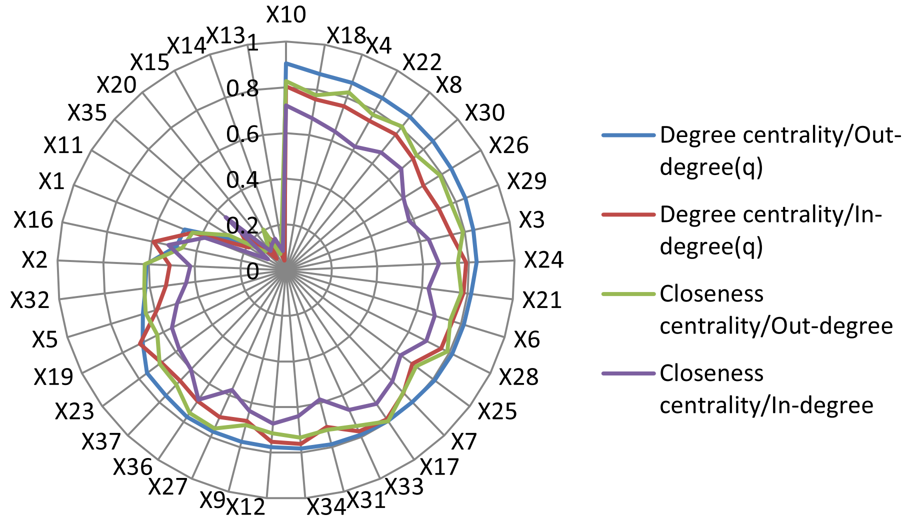
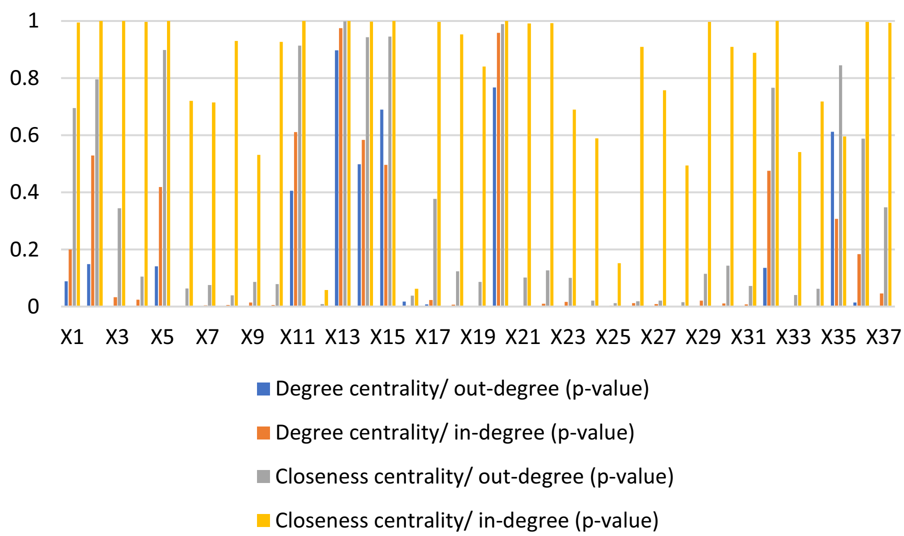


Fig. 6 *P* value of factor detection results for the MYUA's centrality factors in 2011



the in- and out-degrees had diminished, as did the gap between the two degrees. Smaller cities' dependence on larger cities for the information flow space transformed to that of intercity circulation and the sharing of information. Within the region, Wuhan, Changsha, and Nanchang had always reigned supreme in their communication connectivity capabilities, and they continued to grow in strength. The increasing in-degree of other cities such as Yichang, Jingdezhen, and Xiangyang revealed that they had begun paying more attention to other cities in order to obtain more information. The significant increase in the out-degrees of cities such as Yichun, Shangrao, Ji'an, Jiujiang, and Fuzhou indicated that other cities had become familiar with them and that their status in the MYUA's information flow space had risen.

Geo-detection of the evolution of MYUA's SSIF

Factor detection using Geodetector

The natural breaking points method was used to discretize the continuity factor values, whereas the Geodetector was used to quantitatively analyze the contribution of each factor to the spatial evolution of the MYUA's information flow. During the early developmental stage of the information flow space,

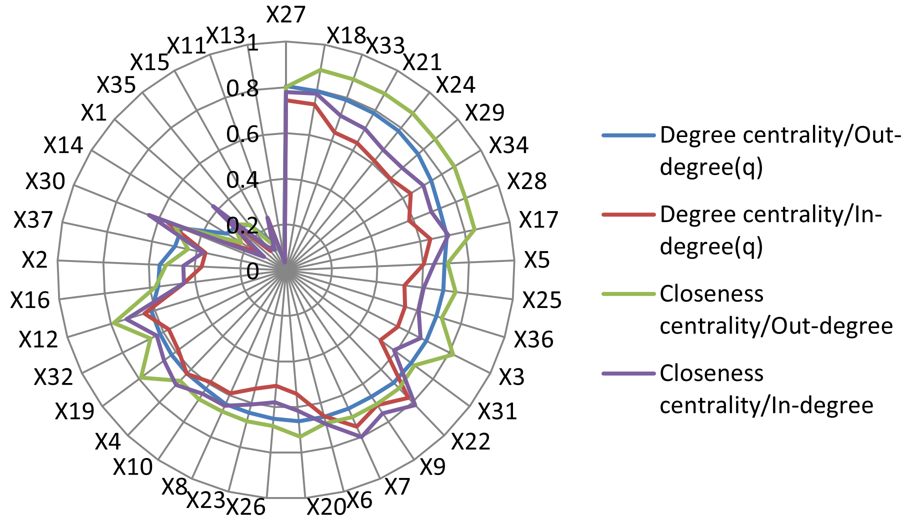


Fig. 7 Detection results for the MYUA's centrality factors in 2017

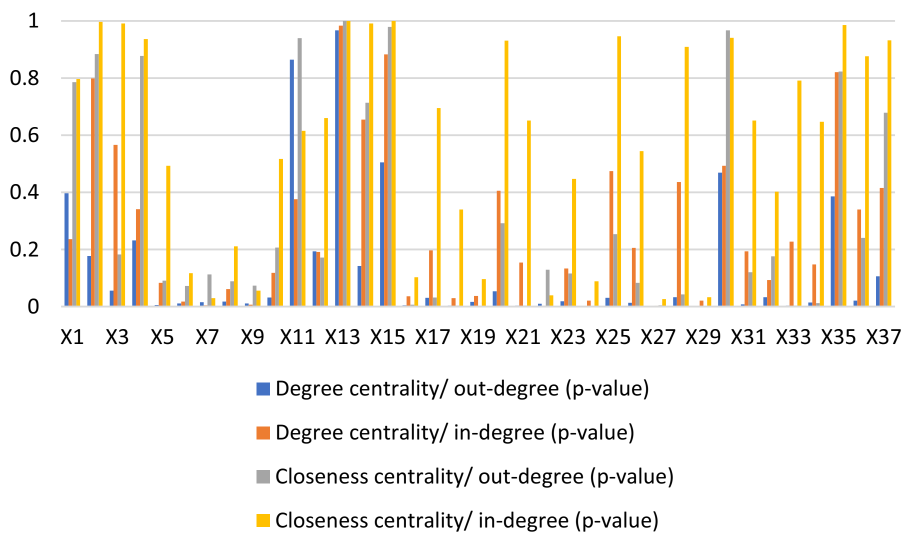


Fig. 8 *P* value of factor detection results for the MYUA's centrality factors in 2017

its spatial formation was predominantly driven by urban background factors, especially urban core strength. However, formation of the MYUA's information inflow and outflow networks was driven by the urban informatization level and urban ability at scientific and educational services, respectively.

In 2011, information searches among the MYUA's various cities were mainly based on their respective urban core strength (see Fig. 5). This factor (especially the financial level) determined a city's status in the information flow space. The level of urban informatization was an important background factor: the higher the level of informatization, the greater the city's ability to disseminate, communicate, and transfer urban information. Urban scientific and educational services were important ways to elevate the quality of Internet users, having positive impacts on their mastery and acquisition of information. The higher the urban scientific and educational levels, the greater the ability of urban Internet users to obtain information, and correspondingly, the greater the demand for information. This situation caused those cities to become more active in the information flow space, thereby facilitating the formation of the urban information inflow network.

For the MYUA's information flow, the stable spatial structure in the form of a network comprising triangular clusters with net-like extensions was influenced

Table 2 Ecological detection results for the MYUA's information outflow network structure in 2017

Response variable	Dominant factor	Number of differing factors	Differing factors
Degree centrality/ Out-degree	Ability to provide scientific and educational services	20	Equivalent scale, level of informatization, core strength, ecological environment, geographic proximity
	Network communication level	14	
	Urban housing level	18	
	Economic development opportunities	12	
Closeness centrality/ Out-degree	Network communication level	16	
	Urban housing level	25	Equivalent scale, level of informatization, core strength, ecological environment, geographic proximity, cultural services, medical and healthcare services
	Economic development opportunities	17	Equivalent scale, level of informatization, core strength, ecological environment, geographic proximity
	Ability to provide scientific and educational services	19	

Table 3 Ecological detection results for the MYUA's information inflow network structure in 2017

Response variable	Dominant factor	Number of differing factors	Differing factors
Degree centrality/In-degree	Economic development opportunities	14	Equivalent scale, level of informatization, core strength, ecological environment, geographic proximity
	Core strength	6	Equivalent scale, economic development opportunities, network communication level
	Ability to provide scientific and educational services	13	Equivalent scale, ecological environment, geographic proximity
	Network communication level	8	Equivalent scale, ecological environment, geographic proximity
Closeness centrality/In-degree	Economic development opportunities	15	Equivalent scale, level of informatization, core strength, ecological environment, geographic proximity, scientific and educational services
	Core strength	5	Equivalent scale, level of informatization, economic development opportunities, network communication level
	Network communication level	10	Equivalent scale, level of informatization, core strength, ecological environment, geographic proximity
	Ability to provide scientific and educational services	11	Equivalent scale, ecological environment, geographic proximity, economic development opportunities

mostly by urban direct factors. The information outflow network was affected predominantly by levels of urban network communication and urban housing, while the information inflow network was affected by urban economic development opportunities and urban core strength. Urban scientific and educational services had a greater impact on direct information linkages between cities, and urban economic

Table 4 Differential stratified combination number of dominant factors in 2017

Indicator		Risk detection	Partition of dominant factors		
Information outflow network/Out-degree	Degree centrality	High-risk area	Number of major websites 753–239	Real estate sales 10 m–30 m	Number of patent applications 20 K–60 K
		Differential stratified combination number	0.70%	0.80%	0.90%
	Closeness centrality	High-risk area	Number of major websites 753–239	Real estate sales 10 m–30 m	Total retail sales of social consumer goods 20 m–60 m
		Differential stratified combination number	0.70%	0.80%	0.80%
Information inflow network/In-degree	Degree centrality	High-risk area	Investments in fixed assets 60 m–70 m	Value added by secondary industries 40 m–50 m	patent applications 20 K–40 K
		Differential stratified combination number	0.30%	0.50%	0.30%
	Closeness centrality	High-risk area	Investments in fixed assets 60 m–70 m	Value added by secondary industries 40 m–50 m	Number of major websites 753–239
		Differential stratified combination number	0.30%	0.30%	0.20%

development opportunities affected the extent of their information exchanges and linkages.

The MYUA's information network became fully developed by 2017, causing discrepancies in q values of the SSIF's core factors to decrease continuously (see Fig. 6). This result indicated that the spatial development of the MYUA's information flow was the result of interactions between multiple complex elements. The level of urban network communication remained an important dominant factor.

A city establishes its own image and interprets images of other cities through the online platform. The core factor attracting the attention of Internet users has always been media reports, which cities use to disseminate their urban ideals. With further urbanization and an increasing shortage of land resources, urban housing has become a topic of considerable concern for Internet users living in urban agglomerations. The urban housing level has become the core factor affecting the information dissemination network of urban agglomerations. At the same time, this factor also affects the ability of cities to control information resources. With increasing labor mobility and the prosperity of commercial developments, the impact of urban economic development opportunities on the centrality of information networks has strengthened significantly, especially in terms of its core impact on urban information flow and linkage. In contrast, the impact of urban core strength has declined. With the narrowing of the gaps in core strength between the various cities in an urban agglomeration, the status and intermediary ability of this factor to affect the information flow space has diminished. Nevertheless, it still has an important influence on the information inflow networks of urban agglomerations (Figs. 7 and 8).

Table 5 Detection of the MYUA's centrality interaction effects in 2011

Dependent variable	Dominant interacting factors	*q* value	Effects of interaction (strengthening)
Degree centrality/Out-degree	X5 ∩ X11	0.9901	Nonlinear
	X11 ∩ X18	0.9874	Two-factor
Degree centrality/In-degree	X20 ∩ X21	0.9729	Nonlinear
	X9 ∩ X20	0.9706	Nonlinear
Closeness centrality/Out-degree	X11 ∩ X12	0.9765	Two-factor
	X30 ∩ X35	0.9754	Nonlinear
Closeness centrality/In-degree	X34 ∩ X35	0.9480	Two-factor
	X17 ∩ X35	0.9427	Two-factor

Table 6 Detection of the MYUA's centrality interaction effects in 2017

Dependent variable	Dominant interacting factors	*q* value	Effects of interaction (strengthening)
Degree centrality/Out-degree	X16 ∩ X22	0.9917	Two-factor
	X16 ∩ X9	0.9879	Two-factor
Degree centrality/In-degree	X16 ∩ X36	0.9913	Two-factor
	X8 ∩ X27	0.9657	Two-factor
Closeness centrality/Out-degree	X16 ∩ X22	0.9637	Two-factor
	X10 ∩ X16	0.9935	Two-factor
Closeness centrality/In-degree	X16 ∩ X36	0.9950	Two-factor
	X10 ∩ X35	0.9698	Nonlinear

Ecological detection using Geodetector

Ecological detection identifies the difference of impacts between urban attributes (Tables 2 and 3). The result appears to be significant at 95% confidence level. There was no significant difference between the impacts of the various dominant factors on the information outflow network. Such factors determine to a large extent the ability of cities to be searched. Thus, there was a significant difference between the impacts of the dominant and secondary factors on this ability. There was no significant difference between the effects of the dominant factors—the urban ability of scientific and educational services, the level of network communication, the urban housing level, and urban economic development—on the information outflow network, meaning that these factors have similar characteristics in their impacts on the MYUA's outflow network structure. However, there were significant differences between the effects of the dominant factors versus the secondary factors such as urban grade and scale, level of informatization, core strength, ecological environment, and geographical proximity.

In this study, it was discovered that definite differences existed between the dominant factors affecting the structure of the MYUA's information inflow network. Concurrently, there were gaps in the number of divergent factors among the dominant factors, such that the spatial distribution of the dominant factors was different from that of the MYUA's information inflow network. Urban core strength and economic development opportunities had significantly different impacts on the distribution of cities that share direct linkages of information. This meant that fewer cities in the MYUA were simultaneously searching for information related to urban core strength and economic development opportunities. A significant difference also existed between the impacts of urban core strength and level of network communication, with the former factor having an important influence on intercity abilities regarding information linkages.

Risk detection using Geodetector

Risk detection reveals response variable in strata (Table 4). The result appears to be significant at 95% confidence level. There is a significant difference between the information inflow intensive area and the information outflow intensive area in the MYUA's information flow network. For information outflow network, the higher the urban housing level and the urban scientific and educational services, the stronger the information outflow. The relationship between number of major websites and information outflow is inverted U type. When the number of major websites in the city is between 753 and 239, the information outflow of the city is the strongest. For information inflow network, there is a positive correlation between the fixed asset investment and the added value of the secondary industry, and the degree centrality of the city, the closeness centrality. The impact on the degree centrality distribution of the city is more significant. The number of patent applications significantly affects the status of cities in information flow networks.

Cities such as Wuhan, Changsha, Nanchang, Yichang, Jingzhou, Zhuzhou, and Jiujiang, which had more effective scientific and educational services, produced more information. Furthermore, they were in central locations of the MYUA's network space for information flow. There was a positive correlation between the level of urban network communication and intercity information linkages. The three major provincial capital cities and their surrounding cities have many websites and a high level of information network transmission. Correspondingly, they have more information linkages with other cities.

Interaction detection using Geodetector

Interaction detection identifies the factors affecting the spatial evolution of the MYUA's information flow (Tables 5 and 6). These were found to have nonlinear and two-factor strengthening effects, meaning that the interactions of multiple factors had greater impact than individual factors on the MYUA's SSIF.

A good ecological environment and high levels of informatization and network communication at the initial stages of the development of an urban agglomeration can enhance information dissemination and the intermediary communication abilities of cities in the information flow space. Cities rely on the ecological environment, especially ones with excellent air quality, to develop leisure tourism. In the process, these cities attract the attention of Internet users and become bridges for information linkages within the urban agglomeration. A strong urban core strength will certainly bring about more economic development opportunities, which in turn allow cities to acquire more information. The housing level of urban residents is of great concern to Internet users in an urban agglomeration; a comfortable housing level and promotion via the media network can narrow the online distance between cities and other regions.

In the developmental stage of an urban agglomeration's network of information flow space, an excellent transportation infrastructure (especially the development of high-speed rail) facilitates closer economic and cultural exchanges among its cities, thereby enhancing intercity understanding. In real world, cities with convenient transportation will have more intense and in-depth contacts with other cities in the urban agglomeration, which in turn affects the former's status in the network of information. The MYUA's information dissemination network in 2017 was formed mainly by the interactive effects between the urban railway transportation system, urban center strength, and economic development opportunities. Convenient urban railway transportation enhances a city's attractiveness in terms of urban leisure tourism. The mutual promotion and joint effects of transportation and tourism have made MYUA's network of information flow more complete.

Conclusion and discussion

Conclusion

The research subject of this study was the digital intercity linkages of the MYUA in 2011 and 2017. The social network analysis method was used to analyze the evolution of the MYUA's SSIF. Then, the index system of influencing factors was constructed through a textual analysis of the urban demand map. At the same time, the analysis of the formation mechanism of the MYUA's network of information space was made using the geo-detection method. The conclusions obtained were as follows:

With the continuous strengthening of information linkages between the MYUA's cities, networking of the SSIF became a trend. From 2011 to 2017, the MYUA's information flow space developed from an information outflow network toward a balance of information inflow and outflow. In the information flow space, the gap between large, medium and small cities is gradually decreasing. This situation basically led to the formation of a stable SSIF in the form of a network comprising triangular clusters with net-like extensions.

According to the factor detection, in the initial developmental stage of the MYUA's information flow space, the dominant factors affecting its information inflow and outflow networks were mostly background factors. Urban core strength and the level of urban informatization were the dominant factors affecting the MYUA's level of information inflow and outflow. With the increasing complexity of MYUA's information flow space, direct factors played predominant roles in the formation of the polycentric network's stable structure. However, there were major differences in the dominant factors that affected the characteristics of the information flow space. For the MYUA's information outflow network, the number of direct searches that its various cities had was mostly affected by their abilities to provide urban scientific and educational services and the level of network communications. The degree of circulation of the MYUA's information generation network was affected by levels of urban network communication and urban housing. For the MYUA's information inflow network, urban economic development opportunities and urban core strength were the dominant factors.

According to the ecological detection, there was no significant difference between the dominant factors affecting MYUA's SSIF. These factors determined to a large extent the ability of cities to be searched. There were definite differences between the dominant factors affecting the structure of the MYUA's information inflow network. There were also differences in the spatial distribution of the various dominant factors, as well as that of the MYUA's information inflow network.

According to the risk detection, there was a significant difference between the information inflow intensive area and the information outflow intensive area in the MYUA's information flow network. For information outflow network, the urban housing level and the urban scientific and educational services were positively related to the information outflow. The relationship between number of major websites and information outflow is inverted U type. For information inflow network, there is a positive correlation between the fixed asset investment and the added value of the secondary industry, and the degree centrality of the city, the closeness centrality.

According to the interaction detection, the interactive coupling between the various influencing factors determined the evolution of the MYUA's information flow space. Interactions between the various factors affecting the MYUA's SSIF led to nonlinear and two-factor strengthening effects, indicating that the interactions of multiple factors had a stronger impact on the MYUA's information flow space than individual factors.

Discussion

Our findings reveal the influence of urban attributes on information flow at different developmental stage and contribute to a more physical based understanding of information flow. In the initial developmental stage of information network, the dominant factors affecting information flow network are mostly background factors. In other words, wealthy and well-connected places tend to have more locally produced user-generated information; thus, the production of information has been spatially clustered and uneven (Ballatore et al., 2017; Graham et al., 2014, 2015). With the development of infrastructure (both internet infrastructure and transport infrastructure) and the increase in internet users, the interaction among cities and the density of information network are gradually increasing. The networked development of urban agglomeration promotes the coordinated development of region. In this context, people from different cities are further integrated into the information network. Individual cyber-spatial cognition and personal behavior in cyberspace influence the evolution of information network (Kwan, 2001), and direct factors that the internet users concern gradually become the dominant factor affecting the evolution of network. Such as, since urban housing has become a topic of considerable concern for people in China, the urban housing level has become an important factor affecting the information outflow network, and the fixed asset investment has become an important factor affecting the information inflow network. And the interactions of multiple factors had a stronger impact on the information network than individual factors, the mutual promotion and joint effects of transportation and tourism have made the network of information flow more complete. Information flow is highly related to the urban attributes. However, the spatial distribution of the dominant factors is different from that of the information inflow network, which shows that the evolution of information network is not a simple reflection of the urban attributes. The interaction between SSIF and urban attributes needs further discussion.

Baidu Search Index data used in this study were reflective of public information exchanges by urban Internet users, and the data had temporal comparability and a real-time updating feature. In terms of research contents, the information flow space was examined using directed network analysis, with the foci being the two aspects of urban information inflow and outflow. Attention was also paid to discussing the factors affecting the evolution of the information flow space. For the selection of factors affecting the evolutionary mechanism of the information flow space, current studies had not advanced beyond qualitative theoretical explanations or the simple regression analysis. In this study, Internet users' different content requirements were identified through their Baidu searches; then, the NLPIR platform was used for textual classification of the related terms. Construction of the index system for the influencing factors served as a specific reference for the quantitative analysis of the formation mechanism behind the information flow space.

Regarding the research methodology, Geodetector was used to identify the influencing factors affecting a city's centrality in the information flow space. Geodetector could process large amounts of data on the influencing factors quickly and easily without having to consider the collinearity of those factors. The level of influence that the software ascribed to each factor also facilitated the intuitive determination of the dominant factors. Geodetector could quantitatively express the results of mutual interactions between multiple factors, having better explanatory power for the information flow network of a complex urban agglomeration. Concurrently, the software's ecological detection could establish the relationship between the influencing factors, and its risk area detection could screen and select sub-regions with significant differences. The use of Geodetector to determine the evolutionary mechanism of the information flow network can lead to a clearer understanding of intra-regional spatial distributions, thereby deepening the understanding of the information social network hidden within an urban agglomeration.

Admittedly, the impact of the urban attributes on the formation of SSIF is examined in this paper without including a detailed discussion on the interactions between information flow and urban attributes. Since relational data between cities were difficult to acquire, only selected data from the Baidu Search Index that had already been processed by Baidu's official platform were used in this study. Different types of data from various channels and search engines could be used to better reflect the information linkages in an urban agglomeration. When constructing the factors affecting the information flow space, no distinction was made between the related preceding and succeeding search terms or the correlation between those related search terms. Finally, Geodetector was used to examine only the overall spatial structure of MYUA's information flow without any in-depth discussion of the structural evolution internal to the information flow.

Funding This study was funded by National Natural Science Foundation of China (Grant Number 41971167).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent The research has not involved human participants or animals.

References

AlSayyad, N., & Guvenc, M. (2015). Virtual uprisings: On the interaction of new social media, traditional media coverage and urban space during the 'Arab Spring.' *Urban Studies, 52*(11), 2018–2034. https://doi.org/10.1177/0042098013505881

Ballatore, A., Graham, M., & Sen, S. (2017). Digital hegemonies: The localness of search engine results. *Annals of* *the American Association of Geographers, 107*(5), 1194–1215. https://doi.org/10.1080/24694452.2017.1308240

Batty, M. (2013). *The new science of cities*. The MIT Press. https://doi.org/10.2307/j.ctt9qf7m6

Cao, F., Ge, Y., & Wang, J. F. (2013). Optimal discretization for geographical detectors-based risk assessment [Article]. *Giscience & Remote Sensing, 50*(1), 78–92. https://doi.org/10.1080/15481603.2013.778562

Castells, M. (1996). *Rise of the network society: The information age: Economy, society and culture*. Blackwell Publishers.

Castells, M. (2010). Globalisation, networking, urbanisation: Reflections on the spatial dynamics of the information age. *Urban Studies, 47*(13), 2737–2745. https://doi.org/10.1177/0042098010377365

Chen, Y., Yan, F., & Zhang, Y. (2017). Local name, global fame: The international visibility of Chinese cities in modern times. *Urban Studies, 54*(11), 2652–2668. https://doi.org/10.1177/0042098016646674

Devriendt, L., Boulton, A., Brunn, S., Derudder, B., & Witlox, F. (2011). Searching for cyberspace: The position of major cities in the information age. *Journal of Urban Technology, 18*(1), 73–92.

Devriendt, L., Derudder, B., & Witlox, F. (2008). Cyberplace and cyberspace: Two approaches to analyzing digital intercity linkages. *Journal of Urban Technology, 15*(2), 5–32. https://doi.org/10.1080/10630730802401926

Golbeck, J. (2013). Chapter 3-network structure and measures. In J. Golbeck (Ed.), *Analyzing the social web* (pp. 25–44). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-405531-5.00003-1

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography [journal article]. *GeoJournal, 69*(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y

Graham, M. (2010). Neogeography and the palimpsests of place: Web 2.0 and the construction of a virtual earth. *Tijdschrift voor economische en sociale geografie, 101*(4), 422–436. https://doi.org/10.1111/j.1467-9663.2009.00563.x

Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers, 104*(4), 746–764. https://doi.org/10.1080/00045608.2014.910087

Graham, M., Stephens, M., & Hale, S. (2013). Featured graphic. Mapping the geoweb: A geography of twitter. *Environment and Planning A: Economy and Space, 45*(1), 100–102.

Graham, M., Straumann, R. K., & Hogan, B. (2015). Digital divisions of labor and informational magnetism: Mapping participation in wikipedia. *Annals of the Association of American Geographers, 105*(6), 1158–1178. https://doi.org/10.1080/00045608.2015.1072791

Hu, X., Wang, C., Wu, J., & Stanley, H. E. (2020). Understanding interurban networks from a multiplexity perspective. *Cities*. https://doi.org/10.1016/j.cities.2020.102625

Hu, Y., Wang, J., Li, X., Ren, D., & Zhu, J. (2011). Geographical detector-based risk assessment of the under-five mortality in the 2008 Wenchuan earthquake China. *PLoS ONE, 6*(6), e21427–e21427. https://doi.org/10.1371/journal.pone.0021427

Huang, J., Wang, J., Bo, Y., Xu, C., Hu, M., & Huang, D. (2014). Identification of health risks of hand, foot and mouth disease in China using the geographical detector technique. *International Journal of Environmental Research and Public Health, 11*(3), 3407–3423.

Janc, K. (2015a). Geography of hyperlinks—spatial dimensions of local government websites. *European Planning Studies, 23*(5), 1019–1037. https://doi.org/10.1080/09654313.2014.889090

Janc, K. (2015b). Visibility and connections among cities in digital space. *Journal of Urban Technology, 22*(4), 3–21. https://doi.org/10.1080/10630732.2015.1073899

Jiao, J., Holmes, M., & Griffin, G. P. (2018). Revisiting image of the city in cyberspace: Analysis of spatial twitter messages during a special event. *Journal of Urban Technology, 25*(3), 65–82. https://doi.org/10.1080/10630732.2017.1348881

Kinsley, S. (2014). The matter of 'virtual' geographies. *Progress in Human Geography, 38*(3), 364–384. https://doi.org/10.1177/0309132513506270

Kwan, M.-P. (2001). Cyberspatial cognition and individual access to information: The behavioral foundation of cybergeography. *Environment and Planning B: Planning and Design, 28*(1), 21–37. https://doi.org/10.1068/b2560

Leszczynski, A. (2015). Spatial media/tion. *Progress in Human Geography, 39*(6), 729–751. https://doi.org/10.1177/0309132514558443

Meijers, E. (2005). Polycentric urban regions and the quest for synergy: Is a network of cities more than the sum of the parts? *Urban Studies, 42*(4), 765–781. https://doi.org/10.1080/00420980500060384

Miller, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers, 94*(2), 284–289. https://doi.org/10.1111/j.1467-8306.2004.09402005.x

Rutherford, J. (2011). Rethinking the relational socio-technical materialities of cities and ICTs. *Journal of Urban Technology, 18*(1), 21–33. https://doi.org/10.1080/10630732.2011.578407

Taylor, P. J., Hoyler, M., & Verbruggen, R. (2010). External urban relational process: Introducing central flow theory to complement central place theory. *Urban Studies, 47*(13), 2803–2818. https://doi.org/10.1177/0042098010377367

Tranos, E., Reggiani, A., & Nijkamp, P. (2013). Accessibility of cities in the digital economy. *Cities, 30*, 59–67. https://doi.org/10.1016/j.cities.2012.03.001

Wang, B., & Loo, B. P. Y. (2019). The hierarchy of cities in Internet news media and Internet search: Some insights from China. *Cities, 84*, 121–133. https://doi.org/10.1016/j.cities.2018.07.013

Wang, J.-F., & Hu, Y. (2012). Environmental health risk detection with GeogDetector. *Environmental Modelling & Software, 33*, 114–115. https://doi.org/10.1016/j.envsoft.2012.01.015

Wang, J. F., Li, X. H., Christakos, G., Liao, Y. L., Zhang, T., Gu, X., & Zheng, X. Y. (2010). Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. *International*

Journal of Geographical Information Science, 24(1), 107–127. https://doi.org/10.1080/13658810802443457

Wang, J. F., & Xu, C.-D. (2017). Geodetector: Principle and prospective. *Acta Geographica Sinica, 72*(1), 116–134.

Wang, Y., Wang, S., Li, G., Zhang, H., Jin, L., Su, Y., & Wu, K. (2017). Identifying the determinants of housing prices in China using spatial regression and the geographical detector technique. *Applied Geography, 79*, 26–36. https://doi.org/10.1016/j.apgeog.2016.12.003

Yu, H., & Shaw, S.-L. (2008). Exploring potential human activities in physical and virtual spaces: A spatio-temporal GIS approach. *International Journal of Geographical Information Science, 22*(4), 409–430. https://doi.org/10.1080/13658810701427569

Yuan, Y., Liu, Y., & Wei, G. (2017). Exploring inter-country connection in mass media: A case study of China. *Computers, Environment and Urban Systems, 62*, 86–96. https://doi.org/10.1016/j.compenvurbsys.2016.10.012

Zhang, L., Du, H., Zhao, Y., Wu, R., & Zhang, X. (2017). Urban networks among Chinese cities along "the Belt and Road": A case of web search activity in cyberspace. *PLoS ONE, 12*(12), e0188868. https://doi.org/10.1371/journal.pone.0188868

Zhen, F., Qin, X., Ye, X., Sun, H., & Luosang, Z. (2019). Analyzing urban development patterns based on the flow analysis method. *Cities, 86*, 178–197. https://doi.org/10.1016/j.cities.2018.09.015

Zhen, F., Wang, B., & Wei, Z. (2015). The rise of the internet city in China: Production and consumption of internet information. *Urban Studies, 52*(13), 2313–2329. https://doi.org/10.1177/0042098014547369

Zook, M. A., & Graham, M. (2007). Mapping DigiPlace: Geocoded internet data and the representation of place. *Environment and Planning B: Planning and Design, 34*(3), 466–482. https://doi.org/10.1068/b3311

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.