Applied Geography 87 (2017) 85–96

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Difference of urban development in China from the perspective of passenger transport around Spring Festival

Jun Xu [a,*], Aoyong Li [a,b], Dong Li [c], Yu Liu [d], Yunyan Du [a], Tao Pei [a,e], Ting Ma [a], Chenghu Zhou [a]

[a] *State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China*
[b] *University of Chinese Academy of Sciences, Beijing 100049, China*
[c] *Innovation Center for Technology, Beijing Tsinghua Tongheng Urban Planning & Design Institute, Beijing 100085, China*
[d] *Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China*
[e] *Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China*

A R T I C L E I N F O

Article history:
Received 6 June 2016
Received in revised form
26 July 2017
Accepted 26 July 2017
Available online 8 August 2017

Keywords:
Urban development
Spring Festival transport
Crowdsourcing data
Network analysis
Spatial interaction

A B S T R A C T

Because cities have different attractions for laborers, people migrate from city to city, which may widen the gap between cities. Analyzing the fluxes and directions of migrant flows can help us clarify the regional difference of urban development. Using migration data inferred from passenger travels during 2016 Spring Festival from Tencent Location Big Data, this paper analyzed the unbalanced migration between cities and the spatial difference of urban development. Network analysis methods are employed to evaluate interactions among cities. A community detection method identifies 19 city communities, and the directions of migrant flows in the communities are explored. The PageRank algorithm is employed to evaluate the importance of cities on the migration network and divide the cities into 5 grades, and then the hierarchical structure of the migrant network is illustrated and analyzed. Indices based on migrant populations indicate that the most attractive cities for laborers are along the east coast and that cities in the central region export a significant number of laborers. PageRank and attractiveness values are compare with socio-economic data, and the results indicate that both PageRank and attractiveness are positively correlated with the economic and development level of cities, while PageRank works better. It suggests that Spring Festival travel data in China can be used as migration data, however, it should be facilitated with network methods to disclose the relationship between Spring Festival travel and urban development.

1. Introduction

China has experienced rapid urban development in past four decades. However, the regional difference of urban development is very significant (Wang, 2013). The urbanization indices present significant inter-provincial difference (Qin & Li, 2013; Wang, Fang, Yang, & Li, 2010). The unbalanced development of cities will result in several social-economic problems, not only over-urbanization problems in metropolitan cities, but also urban decay, industrial decline, and left-behind children and parents in undeveloped regions (Hu, Lu, & Huang, 2014; Huang, Lian, & Li, 2016). It is important to manage the urban development and keep sustainable urbanization (Tan, Xu, & Zhang, 2016; Zhang, 2016).

Many people has studied urban development with different kinds of urbanization rate indices (Du & Cai, 2013; Fang & Wang, 2011; Li, Tan, Lu, & Zhang, 2004). These indices are all based on local situation of a region, such as population, economy, and urban quality, without consideration of interaction among cities. However, interactions among cities, such as inter-city migrations, are also important to urban development. Spatial interaction, which is an important characterization of spatial heterogeneity, has long been the subject of geographical research (Fischer, Reismann, & Scherngell, 2010; Getis, 1991; Ullman, 1954). For example, Seto et al (2012) redeemed that the change of a place has effect on other places and thus they integrated interconnection between

* Corresponding author. 11A Datun Road, Beijing 100101, China.
E-mail address: xujun@lreis.ac.cn (J. Xu).

http://dx.doi.org/10.1016/j.apgeog.2017.07.014
0143-6228/

distant places in processes of urban land change. Halás, Klapka, and Tonev (2016) redeemed that spatial interaction is the foundation to delineates the functional regions, and they used migration data to define the functional regions in the Czech Republic. The migrant flows among different cities are usually caused by labor transitions and reflect the employment attraction of cities. More developed cities attract more people, which in turn aggravated the differences between cities. Therefore, including spatial interaction in an urbanization study and identifying the pattern of migration can help us clarify the regional difference of urban development.

Traditionally, human migration studies were conducted based on census data. In China, census is done about every ten years, so it is hard to get real-time information of migration. Recently, the availability of big data from various sources may provide us new approaches to study human-environment interactions (Liu, 2016). Big data have been used in simulating land used change, urban growth, and climate change adaption of vegetation (Pijanowski et al., 2014; Tayyebi & Jenerette, 2016; Tayyebi, Pijanowski, Linderman, & Gratton, 2014). Long (2016) used big data to redefine the city system in China, which showed difference with the city system defined officially from the administrative perspective. There is a kind of big data, crowdsourcing data, which are obtained from a large population, such as trajectories and phone calls. The crowdsourcing data can be used in extracting urban land uses (Liu, Kang, Gong, & Liu, 2016; Pei et al. 2014), discovering urban structure and function (Liu, Gong, Gong, & Liu, 2015a; Yuan, Zheng, & Xie, 2012; Yue, Lan, Yeh, & Li, 2014; Zhong, Arisona, Huang, Batty, & Schmitt, 2014), detecting geographical events and environment (Bakillah, Li, & Liang, 2015; Crooks, Croitoru, Stefanidis, & Radzikowski, 2013), and evaluating importance of special areas (Levin, Lechner, & Brown, 2017). Compared with traditional data, big data can provide dynamic people movement in real time, and have more capacity in revealing socio-economic geographical features through spatial behavior patterns of a large population (Liu et al. 2015b). Here, the crowdsourcing migration data are used to provide us socio-economic development information of cities.

There is a special phenomenon in China each year around the Spring Festival, which is called Spring Festival transport, or spring transport. Before the Spring Festival, a lot of people return to their hometown, where they originally came from, from cities where they currently reside for family reunion, and then after the festival they go back to the cities where they live. This event causes a large scale of travel rush which lasts for more than one month. Even though the spring transport is large population flows in a short period of time, it depicts the picture of long term migrant population. Through the passenger flows around Spring Festival, the migrant pattern in China can be inferred. Li, Ye, Deng, Liu, and Liu (2016) compared three types of crowdsourcing travel data during 2015 Spring Festival transport and found that the data from location-based services, such as Tencent and Baidu, are more accurate than that from Qihoo, a ticked booking platform. Therefore, the crowdsourcing data of Spring Festival travel can be used as migration data in China.

In this paper, the aggregated passenger travel data released from Tencent during 2016 Spring Festival are employed to characterize the migrant pattern in China and find the relationship between migrant and urban development. Network analysis methods are applied to detect the communities of cities, and PageRank is used to evaluate the importance of cities therefore the hierarchical structure of cities and migrant flows can be disclosed. An index based on Spring Festival travel data are defined to evaluate the attractiveness of cities. Then the network PageRank and the index are compared with the socio-economic data to disclose the relation between migration and urban development.

2. Data

Passenger travel data from the website of Tencent Location Big Data are employed. This website publishes the volume of passengers that travel among Chinese cities each day. The Spring Festival of 2016 was held on February 8, 2016. The spring transport began on January 24 and ended on March 3. We collected passenger travel data during the first three months of the year (Fig. 1). Two peaks of passenger transport were observed before and after February 8, as shown in Fig. 1.

2.1. Data validation

The data encompasses 369 cities, including 293 prefecture-level cities, four municipalities, some county-level cities, special administrative regions in Hong Kong and Macao, and the Taiwan Province, which comprises the majority of China.

Passenger travel data from Tencent Location Big Data provide the volume of daily passengers from the top ten incoming and outgoing routes in each city. The volume of passengers for the top 10 outgoing routes of the cities from January 24 to March 3 is 2117 million, which accounts for 73% of the 2910 million travelers in the same period, as published by the Ministry of Transport. If only the passengers for the top 10 inbound routes of the cities are counted, the number decreases to 2073 million, which is 71% of the total number of passengers. Although only top 10 outgoing routes of a city were recorded in the dataset, additional outgoing routes can be supplemented by the incoming routes of other cities which come out form this city but are not in top 10, and vice versa for the inbound routes. After merging the incoming and outgoing routes, a migrant network with a maximum incoming degree of 272 and a maximum outgoing degree of 270 was constructed. The total number of travelers in this network during the Spring Transport is 2865 million, which is 98.45% of the official number of travelers. The dataset is valid for representing the real situation of migration among Chinese cities.

2.2. Data processing

The red curve in Fig. 1 shows the daily passenger volume during the three-month study period, and the red dashed line shows the daily average volume, which is approximately 63 million. The passenger volume gradually increased at the beginning of the Spring Transport and rapidly decreased a few days prior to March 3, with a trough during a short period near February 8. To select days with typical Spring Transport volumes, we defined an indicator of the difference ratio to compare the incoming and outgoing populations of cities, that is

$$r = \left(\sum_{city=1}^{N} \left| \frac{in\_mum_{city}}{out\_num_{city}} - 1 \right| \right) \Big/ N \tag{1}$$

where in_num_{city} and out_num_{city} are the incoming and outgoing populations of a city and N is the number of cities. Generally, the incoming passengers and outgoing passengers of a city should be balanced; then the ratio of in_num_{city} to out_num_{city} should be near 1, and the value of r should be near zero. When the incoming passengers and outgoing passengers of many cities are not balanced, r is significantly greater than zero. The blue line in Fig. 1 shows the difference ratio of the incoming passengers and the outgoing passengers of all the cities in each day, and the blue dashed line shows the daily average of this ratio. We chose days when both the daily passengers and the difference ratio are greater than the average values as the typical Spring Festival transport