Cities 60 (2017) 180–191

Contents lists available at ScienceDirect

Cities

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

Viewpoint

Delineation of an urban agglomeration boundary based on Sina Weibo microblog 'check-in' data: A case study of the Yangtze River Delta

Feng Zhen [a], Yang Cao [a], Xiao Qin [a,b,*], Bo Wang [c]

[a] Department of Architecture and Urban Planning, Nanjing University, Nanjing, China
[b] Department of Geographic and Oceanographic Sciences, Nanjing University, Nanjing, China
[c] Department of Geography, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 8 December 2015
Received in revised form 18 August 2016
Accepted 27 August 2016
Available online 12 September 2016

Keywords:
people's activity
boundary delineation
Sina Weibo check-in data
Yangtze River Delta urban agglomeration

ABSTRACT

Statistical, morphological, and modelling methods have long been used to delineate spatial boundaries of urban agglomeration. According to the theory of space of flows in the era of globalisation and informatisation, we developed a new approach that uses activity data to analyse activity intensity, closeness, and connection to comprehensively delineate the boundaries of urban agglomeration. We also collected 1,545,449 valid check-in records of the Sina Weibo microblog over two weeks to empirically delineate the boundary of the Yangtze River Delta urban agglomeration (YRDUA) in China. According to our new method, we determined that the space of the YRDUA was much smaller than that based on the administrative division. The socioeconomic index system for the regional planning of the YRDUA (2009–2015) yielded a more accurate result.

1. Introduction

Globalisation and information communication technology (ICT) accelerate the worldwide flow of people, goods, capital, and technical innovation, which is tightly connected to different geographic units (or nodes) and thus markedly changes the world. In this era, urban agglomeration, as opposed to cities, describes the clustering of certain cities and several other closely interrelated cities, has become increasingly crucial in the development of most countries (Boix, Veneri, & Almenar, 2012). According to the new technical–economic paradigm, urban scholars have broadly discussed the new emerging space and the effect of this new technology on physical space. Castells (1989) proposed the concept of 'space of flows' to emphasise the decisive role of information technology in urban or regional development. Accordingly, Castells (1996) extended his concept from a purely virtual space to a pluralistic space in which social elites and their activities are critical. More specifically, as carriers, nearly all flows are accompanied by and reflected by types of daily activity (Castells, 2010; Zhen, Wang, & Wei, 2015). For instance, cross-city activity including business, e-shopping, online searches, and commuting can produce flows of capital, goods, information, and people. Therefore, because of the much higher proportion of current ICT usage, the activity of nearly all residents, rather than that of only social elites, can be used to study urban agglomeration (Taylor, Hoyler, & Verbruggen, 2010; Wang, Zhen, Wei, Guo, & Chen, 2015).

People's activity is continually the focus of time and behavioural geography studies. Static data from travel diary surveys, questionnaires, and interviews are often used to analyse people's activity over one day, one week, or more (Chai, Shen, Xiao, et al., 2012; Kwan & Lee, 2004; Lenntorp, 1977; Miller, 1991; Polak & Jones, 1993). However, static data based on small samples may not accurately reflect dynamic changes to illustrate a more comprehensive understanding of urban agglomeration (Ettema, Timmermans, & van Veghel, 1996; Hagerstrand, 1978; Mateos, 2004; Maxwell, 2012). In recent years, ICT services and devices such as websites, smart phones, GPS navigators, smart cards, and location-based APPs have become widely used for convenient recording, which involves a large amount of people's activity data that often contains geographic information (Lynch, 2008; Mayer-Schönberger & Cukier, 2013; Wang & Zhen, 2014; Wang, Zhen, Wei et al., 2015), which is also useful for analysing urban or regional space (Batty, 2012, 2013; Graham & Shelton, 2013; Hu et al., 2015; Kitchin, 2013; Nielsen, 2012; Wakamiya, Lee, & Sumiya, 2011; Wang & Zhen, 2014; Wang, Zhen, Wei et al., 2015). Although the use of 'check-in' data possess sampling problems, namely age bias (younger groups), gender bias (women), and social class bias (e.g., highly educated groups), it has been most widely used in recent studies (Graham & Shelton, 2013; Hollenstein & Purves, 2015; Krings, Calabrese, Ratti, & Blondel, 2009; Wang & Zhen, 2014; Wang, Zhen, & Zhang, 2015; Wang, Zhen, Wei et al., 2015). With the popularity of check-in applications in online social networks such as Twitter, Flickr, and the Sina Weibo microblog in China (a large social network website in China),

* Corresponding author at: Department of Architecture and Urban Planning, Nanjing University, Wenke Building 616, 210093, China.
E-mail addresses: zhenfeng@nju.edu.cn (F. Zhen), 834154997cao.yang@gmail.com (Y. Cao), qinxiao1070@126.com (X. Qin), wangbo.nju@gmail.com (B. Wang).

http://dx.doi.org/10.1016/j.cities.2016.08.014
0264-2751/

scholars can more cheaply and easily collect a large amount of geographical information on users' activities compared with mobile phones or other ICT devices (such open data is freely captured from application program interfaces on social network websites). However, young people, the largest group of social media users, typically have much higher mobility and space equalisation of activity than do older people; specifically, young people prefer to reach more places and use more facilities in a city or a region. This relationship can also be observed between women and men, as well as between highly educated groups and less educated groups, particularly in China (Li, 2011; Zhen, 2014; Zhou, 2007). To a certain extent, their activities reflect the functional relationships at the city level (Shen & Karimi, 2016) as well as city connections at the urban agglomeration level.

The word 'check-in' often means that someone confirms their location properties on a social network website when engaging in activities at that location (e.g., a specific restaurant or park) (Todd, Campbell, Meyer, & Horner, 2008). Additionally, it also refers to users choosing to or sometimes automatically sharing their locations when they send messages on such websites. Both habits can provide geographic coordinate records of people's activity over a specific period. In related studies that used check-in data, most researchers focussed on people's activity at the city level. Wakamiya et al. (2011) used check-in data from Twitter to demonstrate crowd activity and further analyse urban area characteristics. Mark and Nick (2011) used nuclear density analysis to determine the distribution of people's activity according to geographic and text information from 9223 tweets in the Leeds metropolitan area. Shen and Karimi (2016) analysed closeness and interactions (connections) between scored places on streets by using social media check-ins to explicitly portray urban structure and related socioeconomic performance in Tianjin, China. Hollenstein and Purves (2015) also adopted kernel density analysis to delineate the boundary of the central business districts of London and Chicago by using 8 million pictures along with geographic information from Flickr. However, few studies have used these datasets to analyse people's activity at regional scales, such as urban agglomeration, which is feasible for using check-in data for analysis by following similar methods of city-level analyses (Graham & Shelton, 2013; Kitchin, 2013). This analysis is particularly crucial for China, where government entities pay more attention to approaches based on administrative divisions and the socioeconomic index system instead of analysing interactions of people's activity among cities within an urban agglomeration, the boundaries of which often limit regional development and equity. For instance, two adjacent cities may form strong connections with each other, but not be placed into the same urban agglomeration.

The Yangtze River Delta urban agglomeration (YRDUA) is the sixth largest urban agglomeration in the world and the most economically developed region in China (Tian, Jiang, Yang, & Zhang, 2011; Wang, Fang, & Wang, 2012). However, the current boundary of the YRDUA was defined through administrative divisions and the socioeconomic index system, which limit spatial connections among cities. According to the concept of people's activity, we first developed a new framework for delineating the boundary of an urban agglomeration. We then further examined Sina Weibo check-in data from the YRDUA to analyse people's activity, namely activity intensity, activity closeness, and activity connection to delineate the boundary of the YRDUA. A conclusion and some limitations are discussed in the final section of this paper.

2. Methodology

In previous studies, scholars have usually analysed urban agglomeration from three perspectives. One perspective is hierarchical structure, which is calculated through factor analysis, analytic hierarchy processes, or other statistical formulas according to indices such as GDP, population, the scale of the developed area, and infrastructure (Fang, Song, Zhang, & Li, 2005; Kresl & Singh, 1999; Listengurt, 1975; Mei, Xu, & Ouyang, 2012; Ni, 2008; Rozenblat, 2010; Wang, Deng, & Niu, 2013). The second perspective is spatial structure, or morphology, which is identified through exploratory spatial data analysis (Celebioglu & Dall'erba, 2010; Manzato & Da Silva, 2010), kernel density estimation (KDE) (Borruso, 2003), fractal modelling (Che, Duan, Guo, Wang, & Cao, 2011; Tannier, Thomas, Vuidel, & Frankhauser, 2011), remote sensor identification (Abed & Kaysi, 2003; Kumar, Pandey, Hoda, & Jeyaseelan, 2011), and the compounded night light index (Gao, Huang, He, & Ma, 2015; Zhuo, Shi, & Chen, 2003). The third perspective is network structure, which is analysed through two methods, namely using a gravity model (Chen & Song, 2010; Huff & Lutz, 1989; Kharitonov, 1968; Krings et al., 2009) and a field model (Liang, 2009) as well as breaking point (Wu, Li, & Sun, 2012) or Voronoi analysis (Mei et al., 2012; Ottaviano & Pinelli, 2006) to simulate the economic connections among cities. Additionally, in recent years, social network analysis (SNA), a method for analysing element flows, has been employed to determine connections related to traffic, population, information, and business (Veneri, 2013; Wang-Ming, Zhou-Yue, & JIANG, 2011; Xiong, Zhen, Wang, & Xi, 2013; Zhen, Wang, & Chen, 2012) among cities. Regarding the delineation of urban agglomeration boundaries, many scholars have supported comprehensively analysing the three aforementioned structures in a specific urbanised region (Henderson, Yeh, Gong, et al., 2003; Imhoff, Lawrence, Stutzer, et al., 1997; Wang et al., 2013; Zhang, Hu, Liu, et al., 2012), which has been ignored in previous studies because of the limitations data collection and application.

Based on the previous discussion People's activity largely reflects the actual state of urban agglomeration in the globalisation and informatisation era, indicating that this variable can be used to delineate the boundaries of urban agglomeration. Therefore, combined with the three aforementioned perspectives on analysing urban agglomeration, we developed a new and comprehensive approach to accurately delineate an urban agglomeration boundary to extend the results of past studies (Fig. 1). First, we collected check-in data from social media and performed data cleansing. Second, in accordance with previous studies, three indices were developed to analyse people's activity among multiple cities or within a single city. In particular, because large, competitive cities typically produce or attract more people's activity, the index of activity intensity was calculated to reflect the hierarchical structure of urban agglomeration; urban agglomeration is composed of several core cities and surrounding areas; therefore, its spatial structure or morphology can be calculated according to the index of activity closeness (Shen & Karimi, 2016); the index of activity connection comprehensively reflects to true flow of different elements in a region, which can be used to identify the network structure of an urban agglomeration. Third, an information-gain approach was employed to enhance decision making on the weights for every index of people's activity analysis, and a comprehensive evaluation model was used to delineate the final boundary of urban agglomeration.

The detailed calculations of the three indices of people's activity and boundary delineation are as follows:

2.1. Activity intensity

The index of activity intensity represents the city activity size used to determine the city hierarchy in the urban agglomeration by measuring the check-in number per unit of developed area in the sampling area (Formula (1)).

$$V_{\text{Activityintensity}} = \frac{V_{\text{check-in activities}}}{V_{\text{build-up area}}} \tag{1}$$

The $V_{\text{check-in activity}}$ represents the total check-in activity quantity in each basic unit area, and the $V_{\text{build-up area}}$ represents the urban developed area.