Remote Sensing of Environment xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Using multi-source geospatial big data to identify the structure of polycentric cities

Jixuan Cai [a], Bo Huang [a,b,*], Yimeng Song [a,*]

[a] Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong
[b] Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 19 July 2016
Received in revised form 11 June 2017
Accepted 28 June 2017
Available online xxxx

Keywords:
Nighttime light image
Social media
Image segmentation
Spatial statistics
Polycentric structure
Subcenter

ABSTRACT

Identifying the structure of a polycentric city is vital to various studies, such as urban sprawl and population movement dynamics. This paper presents an efficient and reliable method that uses multi-source geospatial big data, including nighttime light imagery and social media check-in maps, to locate the main center and subcenters of a polycentric city. Unlike traditional methods that rely on statistical data categorized by administrative units, the proposed method can effectively identify the boundaries of urban centers, and the data source guarantees a timely monitoring and update. Four main procedures are involved: 1) a new observation unit is developed using object-oriented segmentation; 2) main centers are located using cluster analysis (Local Moran's I); 3) subcenter candidates are selected using significant positive residuals from geographically weighted regression (GWR); and 4) final centers are filtered using global natural breaks classification (NBC). These steps can be reproduced in different regions. To evaluate the effectiveness, the method was applied to three rapidly developing Chinese cities: Beijing, Shanghai, and Chongqing with different natural and economic characteristics. The performance of the proposed method has been carefully evaluated with qualitative and quantitative analyses. Comparative experiments were also conducted across different datasets to prove the benefits of combining a social media check-in map with remotely sensed imagery in a human environment study.

1. Introduction

Urbanization has led to an increase in the number of urban dwellers and significant changes to urban structures (Bai et al., 2014). In recent decades, more polycentric cities have emerged, resulting from the previously close-by but independent urban settlements that become a larger and more integrated city-system (Liu and Wang, 2016). The studies of polycentric city structure have been undertaken at different geographical scales, including the inter-city scale and the intra-city scale (Yang et al., 2015). A polycentric city at the inter-city scale usually covers more than one urban areas, as well as satellite cities, towns and intervening rural areas that are socio-economically tied to the urban core (Liu and Wang, 2016).

The components of the intra-city polycentric structure usually include the main center and the subcenters (McMillen and McDonald, 1997). The main center is the core of a city and generally covers the central business district (CBD). Subcenters are areas with greater densities of human activity than nearby locations within a city, which include edge cities and satellite towns. Such areas enjoy the benefits of agglomeration, but offer lower commuting costs for citizens and cheaper land costs to corporations than the urban downtown (McMillen, 2001). Accurately delineating the polycentric structure of a city is important for a better understanding of urban expansion and provides the public and city managers with information needed to evaluate the effectiveness of planning layouts. However, the distribution of urban centers is influenced by a variety of topographic and socio-economic circumstances, which are rarely parameterized into circles or ellipses (Redfearn, 2007).

Our knowledge of the urban structure of cities is highly restricted by the availability of data. Previous researchers looking at this study have mainly relied on statistical sources, like population census and economic data. A rigorous and sophisticated method using this kind of data for defining the polycentric structure was developed by McMillen (2001). He significantly advanced this field by adopting non-parametric techniques such as locally weighted regression (LWR) and semiparametric employment density functions to define subcenters as areas with significantly higher human densities than the expected density based on their distance from a CBD. This procedure has been widely used in subsequent work by McMillen (2003, 2004) and other researchers (Garcia-López, 2010; Riguelle et al., 2007). However, this method needs to subjectively select the CBD location, prior to the other processes. Identifying that location is very difficult for users who do not have

* Corresponding authors.
E-mail addresses: bohuang@cuhk.edu.hk (B. Huang), yimengsong@link.cuhk.edu.hk (Y. Song).

http://dx.doi.org/10.1016/j.rse.2017.06.039

detailed knowledge of the study area; furthermore, main centers rarely have sharp region boundaries and the number of CBDs can change as a city grows. Those problems become more intractable when the study areas are huge and developing rapidly.

Besides, datasets like population censuses have a high level of accuracy and representativeness but a low update frequency, usually being renewed once every five or ten years. In addition, when such spatial statistics are used to determine the number of subcenters, the size of the observation units limits the adoption of the center definition method. For example, large units may produce fewer subcenter sites than more disaggregated data. Statistical data aggregated into administrative boundaries are unable to reveal the accurate distribution of human density below the administrative division level. When an administrative region is large, dense population sites may be ignored due to the large amounts of unused land within the same region.

Remote sensing data, like nighttime light satellite imagery, could also provide various features of urban landscape and infrastructure, adding a new potential source to study urban structure. For a long time, researchers have applied data from Defense Meteorological Satellite Program–Operational Linescan System (DMSP-OLS) to detect urban settlements (Elvidge et al., 1999, 2007; Ma et al., 2012; Sutton, 2003). Yu et al. (2014) developed an object-based method to characterize urban spatial patterns from nighttime light satellite images. As the new released data from Visible Infrared Imaging Radiometer Suite (VIIRS), more detailed inner-city structure monitoring became possible (Elvidge et al., 2013). Some studies have been carried out to estimate the socioeconomic indicators in finer resolution (Xi Li et al., 2013; Ou et al., 2015). The performance of urban area extraction at regional scale was verified by Shi et al. (2014). Although nighttime light data have relatively high spatial stability and guarantee the reliability of land parcel shaping, remote sensors are still incapable of recording socioeconomic attributes and human dynamics such as daily activities (Liu et al., 2015). For example, not only urban centers, but also roads, port regions, and industrial districts, emit glowing light at night, which can result in inaccurate estimates of population accumulation areas (Zhang et al., 2013).

In recent years, the rapid growth of location-based services and social media platforms has created new opportunities to discover the spatial characteristics of human behavior and activities (Jiang et al., 2016; Lee and Sumiya, 2010; Stefanidis et al., 2011). The high correlation between the check-in density of social media data and the human density distribution has been revealed by many studies (Cheng et al., 2011; Dunkel, 2015; Frias-Martinez et al., 2012; Steiger et al., 2015). Compared to conventional static data sources such as a population census, social media data are representative indicators with a much finer temporal-spatial scale that can depict the actual dynamics of the activities of urban dwellers (Hawelka et al., 2014). However, the check-in locations are so concentrated that most of the events occur around particular hot spots within a local region, resulting in serious spatial variability and regional instability. Therefore, there lacks a method that can provide both reliable land parcel shaping and quantitative analysis for the polycentric structure identification.

In this paper, we attempt to provide a method combining the advantages of nighttime light satellite images and social media check-in data for identifying the structure of polycentric cities. Three main steps are included: developing observation units, main center definition and subcenters definition. The Results section includes a test of our method across three cities and with different datasets and methods to demonstrate its effectivity. Two different methods are also provided to evaluate the accuracy of our results. The paper concludes with a summary of the method's advantages and the limitations of this study.

2. Study areas and data

2.1. Study areas

The three big cities, Beijing, Shanghai, and Chongqing, were selected as our study areas (Fig. 1). The geographical characteristics and urban morphology patterns vary greatly in these three cities, allowing us to verify the effectiveness and robustness of our method. Only municipal districts (city-controlled districts) are included in this study. Regions like the county-level cities or counties administratively belonging to these cities are removed.

As the capital of China, Beijing has undergone rapid suburbanization since the 1980s. It is densely populated with over 18.59 million residents (in 2015) living within the city-controlled districts of 12,046 km^2. The main built-up area of Beijing lies on the northern side of the North China Plain where the elevation ranges from 20 to 60 m. The plain topography means that the city's rapid expansion is not limited by natural conditions. The city has developed in a classic pie form, spreading out in concentric ring roads.

Shanghai is China's most populous city, where >24.15 million citizens (in 2015) are living in the city-controlled districts with area of 5462 km^2. The city sits in the Yangtze River Delta and the old urban and modern downtown are located on a vast alluvial plain that is divided by the Huangpu River. Due to the barrier of Yangtze River and the coastline, the city has mainly extended in the southwest direction.

Chongqing is the largest municipality under the direct administration of the Central Government and the only one in western China. The city-controlled area of Chongqing is about 15,162 km^2, with 15.7 population (in 2015) settling in. The city covers a large area crisscrossed by rivers and mountains, with great sloping areas at different heights. Therefore, the urban development is highly affected by topographical characteristics, leading to a relatively complex urban structure.

2.2. Data

2.2.1. Nighttime light imagery

In this study, nighttime light data are applied to characterize the textural features of urban built-up areas and to construct new observation statistical units. The Visible/Infrared Imager/Radiometer Suite (VIIRS), launched in October 2011, was designed to collect high-quality nighttime images in the day/night bands (DNBs), between 500 and 900 nm, with a ground spatial resolution of around 500 m (Miller et al., 2012). The VIIRS monthly composite data used in this study were obtained directly from the website of the Earth Observation Group, NOAA (http://ngdc.noaa.gov/eog/viirs/download_monthly.html), and the interference from stray light, lunar illumination, and cloud-cover is filtered out. The spatial resolution of the VIIRS product is 15 arc-seconds in the geographic grid covering the three cities' study areas.

2.2.2. Social media data

Compared to other spatial big data sources like mobile phone data, social media check-in record is believed to be more appropriate and commonly used for urban structure detection, because check-in events would normally be created when users are aware of something and stay in a particular position for relativity a long time (Kaplan and Haenlein, 2010). Weibo (microblogs) is one of the most popular social media platforms in China, whose monthly active users ("MAUs") reached 222 million in September 2015 and mobile MAUs represent 85% of the total MAUs (Weibo Corporation, 2015). In this study, the check-in map of Weibo is used as representative spatial information on human activities.

3. Methodology

3.1. Data preprocessing

Thirteen images covering the three study areas from April 2015 to April 2016 were collected for this study. An averaging map was calculated from those images to reduce noise and to use in null pixels. Finally, the data were projected into the Universal Transverse Mercator (UTM) grid with a resolution of 500 m.

With the movement of crowds and dynamic hotspots of interest, the check-in distribution of Weibo changes over time. The spatial

Please cite this article as: Cai, J., et al., Using multi-source geospatial big data to identify the structure of polycentric cities, Remote Sensing of Environment (2016), http://dx.doi.org/10.1016/j.rse.2017.06.039