Applied Geography 75 (2016) 36–48

Contents lists available at ScienceDirect

Applied Geography

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

Redefining Chinese city system with emerging new data

Ying Long

School of Architecture, Hang Lung Center for Real Estate, Tsinghua University, China

ARTICLE INFO

Article history:
Received 25 January 2016
Received in revised form
25 July 2016
Accepted 2 August 2016

Keywords:
Urban morphology
Urban function
Human activity
Street network
City evolution

ABSTRACT

Modern Chinese cities are defined from the administrative view and classified into several administrative categories, which makes it inconsistent between Chinese cities and their counterparts in western countries. Without easy access to fine-scale data, researchers have to rely heavily on statistical and aggregated indicators available in officially released yearbooks, to understand Chinese city system. Not to mention the data quality of yearbooks, it is problematic that a large number of towns or downtown areas of counties are not addressed in yearbooks. To address this issue, as a following study of Long et al. (2016), we have redefined the Chinese city system, using percolation theory in the light of newly emerging big/open data. In this paper, we propose our alternative definition of a city with road/street junctions, and present the methodology for extracting city system for the whole country with national wide road junctions. A city is defined as "a spatial cluster with a minimum of 100 road/street junctions within a 300 m distance threshold". Totally we identify 4629 redefined cities with a total urban area of 64,144 km^2 for the whole China. We observe total city number increases from 2273 in 2009 to 4629 in 2014. We find that expanded urban area during 2009 and 2014, comparing with urban areas in 2009 are associated with 73.3% road junction density, 25.3% POI density and 5.5% online comment density. In addition, we benchmark our results with the conventional Chinese city system by using yearbooks.

1. Introduction

Cities, used to be as simple as spatial entities surrounded by city walls in ancient time, have been getting increasingly complex. For instance, the total number of Chinese cities has increased from about 100 in 1949 to 653 in 2014, not to mention the frequent spatial adjustments of a single city (Wu, Long, Mao, & Liu, 2015). A universal difficulty for urban studies is how a city can be defined properly (Batty, 2006; Krugman 1996). There are several lenses to look at a city, morphologically, functionally, and administratively. **From the morphology dimension**, a city can be defined as an area with a minimum of population or continuous built-up area. For instance, Densely Inhabited District (DID) in Japan indicating a city whose population density is over 4000 people per km^2 is a similar concept. Urban areas in UK are derived from constructions-built areas where certain real-estate densities are detected in satellite images or other datasets (Hu, Wu, Xiong, & Pan, 2008). **From the functional dimension**, a city can be regarded as a functional urban area with a dense core and peripheral area as its commuting catchment (Cottineau, Hatna, Arcaute, & Batty, 2016). For instance, labor force markets and commuter sheds are utilized to represent Metropolitan Areas (MA) in US (Berry, Goheen, & Goldstein, 1969). The morphological and functional dimensions have been widely adopted by cities in western countries, and administrative dimension representing for the political and management territory applies to almost all cities on the planet. Rather than from morphological or functional dimensions, modern Chinese cities are defined **from the administrative view** and classified into several administrative categories ranging from municipalities, prefectural level to county level, due to lack of sufficient data and their historical path dependence. Hence, it makes Chinese cities inconsistent with their counterparts in western countries. While the administrative level and extent of each Chinese city was consistent with its spatial power and form in its initial stage, we see increasing mismatches between the administrative definition and the spatial dimension of Chinese cities.[1]

Concerning the administrative dimension governing the definition of existing Chinese cities and being not able to access fine-scale

---

[1] We also notice the notion of "Natural Cities" proposed by Jiang and Miao (2015). Natural city is a product of the bottom-up thinking in terms of data collection and geographic units or boundaries, which can be extracted from remote sensing images, GPS and location-based social media data.

E-mail address: ylong@tsinghua.edu.cn.

http://dx.doi.org/10.1016/j.apgeog.2016.08.002
0143-6228/

urban data, researchers have to rely heavily on statistical and aggregated indicators available in officially released yearbooks (Deng, Huang, Rozelle, & Uchida, 2010; Mao, Long, & Wu, 2015), as well as remote sensing images (Liu, Zhan, & Deng, 2005), to understand the Chinese city system. On one hand, the indicators of yearbooks in China are generally gathered in a bottom-up manner and reported by the staff in each city, thus making the inconsistent data quality across cities due to various reasons like the variations on training level of staffs or technology developing levels among cities. Not to mention the everlasting criticism on data quality of yearbooks, it is problematic that a large number of towns or downtown areas of counties, which should be regarded as cities from a general view of built-up area continuity and population scale, are not addressed in yearbooks (termed as "invisible Chinese cities" or "neglected Chinese cities"). Some large cities like Beijing should be regarded as several separated cities in the commonly accepted definition on cities. In addition, city size is not directly linked to its administrative level as well. On the other hand, those studies using remote sensing-interpreted expansion on understanding urbanizing China still rely on the aforementioned problematic administrative dimension of Chinese cities. Therefore, to address these issues on existing Chinese city system, redefining Chinese city system is urgent for both researchers and decision makers.

The spatial dimension (urban area pattern) is an important view for understanding a city system. An overall review on existing methods for mapping urban areas is available in the second section of Long, Shen and Jin et al. (2016) and the first section of Zhou (2015). These methods range from remote sensing images, population census, settlement and building data, points of interest (POIs),[2] volunteered geographical information (e.g. Flickr photos and Twitters with geo-tags) to road network. Most of the applications of these existing methods in China, if not all, map urban areas for the existing administrative-oriented city system and do not consider redefining Chinese city system. For instance, Long et al. (2016) provide a straight-forward method for allocating urban areas according to urban area totals prescribed in not-always-reliable yearbooks. In this paper, our methodology will not rely on urban area totals prescribed in yearbooks and extract national wide urban areas using road junctions with a bottom-up method.

In this paper, we develop a framework for redefining the Chinese city system from the views of urban morphology, function and human activities by using percolation theory in the light of big/open data. Nowadays, big data and ubiquitous open data from the governments, commercial and social network websites are increasingly available to researchers, and the new data environment is providing more avenues for quantitative urban studies (see Long and Liu (2015) for a review). In such a background, aiming at providing an alternative view on the Chinese city system, we propose our definition of a city with road junctions. Our road-junction oriented city system using percolation theory is more like population center classification system, if we consider a city with many road junctions can be regarded as a city (similar with the minimum population control for a city). We then present the methodology for extracting city system for the whole country with national wide road junctions as well as the results of redefined Chinese city systems in 2009 and 2014, respectively. In addition, we discuss the findings on (1) evaluating the redefined Chinese city system in terms of size, pattern, scaling, hierarchy and temporal evolution, (2) understanding the evolution of the Chinese city system during 2009–2014, and (3) benchmarking our results with the conventional Chinese city system by using yearbooks.

This paper is structured as follows. Section 2 describes the adopted methodology. The used datasets and their results are introduced in Section 3 and 4. Section 5 makes concluding remarks and propose discussion on this study.

2. Method

2.1. Deriving urban areas with road junctions

Considering the data availability and approach reproducibility, we refer to the method proposed by Masucci et al. (2015) for deriving urban areas using road junctions. The approach has been successfully applied for the city system in Britain. In this paper, all road junctions in a large region are partitioned into various clusters by using an elementary clustering technique, which considers two road junctions belonging to the same cluster if their distance is below a given distance threshold (Fig. 1). An increasing threshold enlarges the size of generated clusters, until eventually a very large cluster appears spanning the whole road network. We then calculate the size of the maximum cluster in terms of number of road junctions as a function of the threshold. For all the cities, the size of the maximum cluster has proved to grow exponentially. Eventually, the growth slows down and the curve condensates to a certain value (Masucci et al., 2015). The condensation threshold can be used to derive urban areas of various cities (clusters actually). The condensation threshold of about 300 m was observed in the Greater London. The details for this process are available in the appendix of Arcaute et al. (2015).

2.2. Redefining city system

In this sense, we redefine a city is a spatial entity which (1) consists of a cluster of adjacent road junctions within a predefined distance (morphological requirement for a city system). The condensation threshold is used as the clustering distance; (2) associates with at least 100 road junctions (minimum size requirement, 4 km generally). Each city is then associated with a group of road junctions. The size of each city can be regarded as the total number of road junctions, which has been proved to be significantly correlated with the built-up area of the city (Jiang & Jia, 2012; Masucci, Stanilov, & Batty, 2013).

To derive city spatial distribution, we should convert road junctions of each city into polygon(s) to represent urban areas of the city. The polygons stand for the triangle network composed by road junctions. This can be done by using the toolbox Aggregate Points in ArcGIS. In this process, we use 500 m as the aggregation distance, which is defined according to a trial-error procedure.

2.3. Evaluating redefined city system and its evolution with open data

The redefined city system can be evaluated from geometric, morphological, functional and social dimensions sequentially. From the geometric dimension, the size, boundary, and scaling characteristics of redefined cities can then be evaluated. Since the morphological evaluation focuses on the spatial organization of road junctions in every city, indicators like road junction density can be used to reflect the urban block size pattern (urban design) to gain a big picture of urban spatial structure. Considering that every POI reflects one type of urban function like dining, shopping, education, public administration, etc, the functional dimension pays more attention to using POIs in order to understand urban functions of each city. We can use the total POI density and the POI density of each type to evaluate the degree of maturity for every

---

[2] According to Wikipedia, "a point of interest, or POI, is a specific point location that someone may find useful or interesting" (https://en.wikipedia.org/wiki/Point_of_interest).

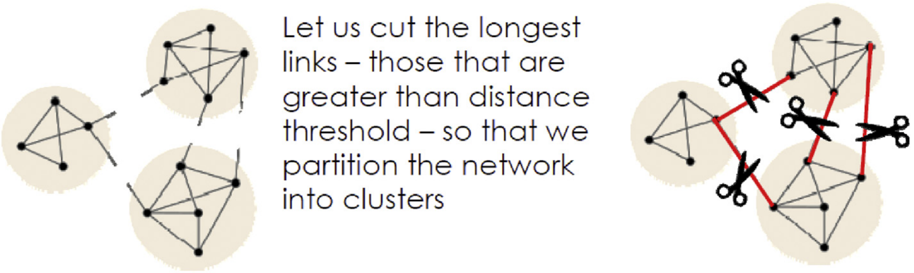


Fig. 1. Clustering road junctions with a threshold of edge length (from Masucci et al., 2015).

city (e.g. functional vibrant or not), which supplements to the geometric and morphological evaluations. Last but not least, we evaluate these cities from the view of human activities (social dimension). Social network records can be equipped for the evaluation, which, in addition to functional evaluation, further extends our understanding on the redefined city system. The evolution of our redefined cities can be evaluated via applying the multi-dimensional evaluations to cities at different time points.

Urban expansion of the redefined city system also deserves an in-depth investigation for comparing with urban expansion from the perspective of the conventional city system. **First**, urban expansion in a time period (t_1 to t_2) can be derived using the polygons (urban areas) of redefined cities at the starting (t_1) and end (t_2) time points of the period, which can be achieved by using the overlay function in almost all GIS software. **Second**, the magnitude of urban expansion can be measured using the geometrical size of expanded urban areas during t_1 and t_2. The ratio of magnitude and the urban area size at the start time point (t_1) indicates the spatial urban expansion speed. **Third**, the quality of urban expansion during the period can be evaluated from these aforementioned four dimensions, via comparing the corresponding density of each dimension for the expanded urban areas and urban areas in the redefined city system at t_1. More knowledge on urban expansion can then be gained via benchmarking urban expansion evaluation results with those of old developments (city system at the beginning of a time period).

We also compare our redefined city system with existing data sources for benchmarking our methodology and research framework for re-understanding the city system from the spatial entity perspective. Urban areas in the other data sources are extracted and overlaid with urban areas in our redefined city system. Similar with urban expansion identification, this process can also be achieved by using the overlay function in most of GIS software. The match ratio is calculated by dividing the size of overlaid urban areas of two layers with the size of urban areas of our redefined city system.

3. Data

3.1. The Chinese city system from the administrative view

The Chinese city system has long been defined from the administrative view, and most of statistical data correspond to the administrative Chinese cities, not the concepts like metropolitan statistical areas (MSA) in the United States or functional urban areas in European Union countries. In total, there are 653 Chinese cities in 2014 according to MOHURD (2014) (Fig. 2). On the basis of Chinese administrative system, there are mainly five levels of cities classified in this way including: four municipalities directly leaded by the nation, 15 sub-provincial cities, 17 other provincial cities, 250 prefecture-level cities, and 367 county-level cities. This system generally reflects hierarchy of these cities in terms of city size and population. More information regarding administrative divisions of China is available at WIKIPEDIA (https://en.wikipedia.org/wiki/Administrative_divisions_of_China). The statistics of city type and urban area are in Table 1, as a profile of Chinese city system. The study area of this paper covers the whole China territory, not being limited to administrative areas of cities (city proper) in Fig. 1. The administrative boundaries are used to extract our analysis results in the whole country for benchmarking with conventional indicators available in yearbooks.

Based on the city administrative boundaries, statistics of urban area are extracted from MOHURD (2009 and 2014) to compare with our redefined city system at the administrative city level. In this paper, we focus on the timespan from 2009 to 2014 to make it consistent with the used open data. Until 2014, total urban area of 653 cities in China has reached 49,743 km^2 from 41,175 km^2 in 2009 (see Table 1 for each administrative category). MOHURD (2014) also says that there were 1596 county towns with total urban areas of 20,100 km^2, and 20,401 towns with total urban areas of 38,000 km^2. Urban areas in county towns and towns are not accounted in the 653 cities. However their urban areas are greater than that of 653 cities.

3.2. Road junctions in 2009 and 2014

To derive the most important data of this study (road junctions), we obtain road networks in 2009 and 2014 of the whole China from a Chinese navigation company. Urban streets, regional roads and many other detailed roads are encompassed in the road networks, and the quality of the data is guaranteed by our previous study, Liu and Long (2016) which used the 2011 version of the same dataset.

We conduct necessary data pre-processing on road networks to derive more practical road junctions using ESRI ArcGIS (Fig. 3 as an example). We first remove foot paths from the raw streets/roads,[3] and merge the remaining roads (Fig. 3a) to single-line road features using the “Merge Divided Roads” tool. Matched pairs of roads are merged if they are in the same road class, generally parallel to each other, and are within the merge distance between each other (Fig. 3b). We then use “Must not have dangles”, “Must not have pseudo nodes”, and “Must not intersect” as topology rules for deleting dangling roads, pseudo nodes, and breaking roads into road segments at road junctions respectively (Fig. 3c&d).

We then derive 2.98 m (million) road junctions in 2009 and 8.24 m road junctions in 2014 (Fig. 4). These road junctions are then used for deriving redefined cities.

3.3. Points of interest (POIs) in 2009 and 2014

POIs in 2009 and 2014 are gathered and geo-coded by business

---

[3] We use the term “road” in the following context to represent both roads and streets.

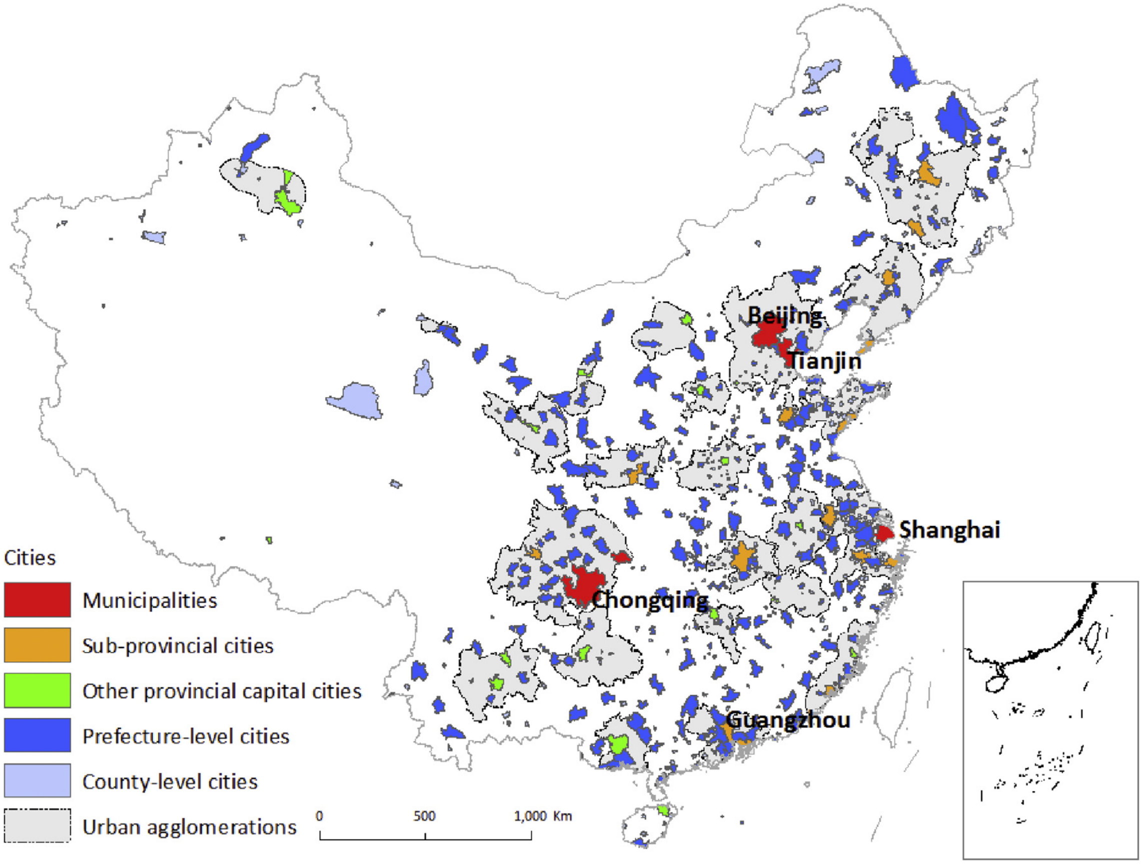


Fig. 2. Administrative areas of Chinese cities (the polygon in color stands for the administrative region of a city).

Table 1
Statistical information for every administrative level of cities.

Administrative category	# Cities in 2014	Total urban areas in 2009 (km^2)	Total urban areas in 2014 (km^2)
Municipalities	4	5211	6318
Sub-provincial cities	15	6119	7201
Other provincial cities	17	3353	4434
Prefecture-level cities	250	18,190	22,245
County-level cities	367	8302	9546
Total	**653**	**41,175**	**49,743**

Note: there were 654 cities in 2009.

cataloguing websites. The initial twenty POI types are aggregated into eight more general assemblies: commercial sites account for most POIs, followed by business establishments, transportation facilities, and government buildings (see Liu and Long (2016) for the 2011 version of POIs). The POIs data quality is ensured through manual checking randomly selected POIs. There are 1.56 m POIs in 2009 and 10.6 m POIs in 2014 (Fig. 4). These POIs are used for evaluating the redefined Chinese city system and its evolution from the functional dimension.

3.4. Other datasets

For benchmarking our redefined city system of China, we gather the urban areas in 2009 and 2014, respectively. The 2009 data are obtained from the dataset developed by He, Liu, Tian, and Ma (2014), which are interpreted from DMSP/OLS and associated with a spatial resolution of 1 km. The average overall accuracy is 95.2%, and the Kappa value is 0.66. For urban areas in 2009, the total area is 52251.0 km^2 for 2166 patches (the mean patch size is 24.1 km2). The 2014 data are retrieved from VIIRS/DNB (the updated version of DMSP/OLS[4]) and with a spatial resolution of 500 m (Xu, He, Liu, & Dou, 2016). There are 12,752 patches for 59530.5 km^2 in total in urban areas in 2014, and the mean patch size is 4.7 km^2. A total urban area of 13,779.7 km^2 in 2009; 26.4% of urban areas in 2009 is not urban in 2014 due to different approaches used for image interpretation and spatial resolution of raw images.

For evaluating the redefined city system of China and its evolution from the social dimension, we gather the Dianping[5] comments for catering establishments in China (Liu, 2014). Totally 13.0 m Dianping POIs are collected, and 1.9 m POIs are with at least one comment. Totally there are 47.0 m comments for all Dianping establishments.

4. Results

4.1. Redefined the Chinese city system in 2014

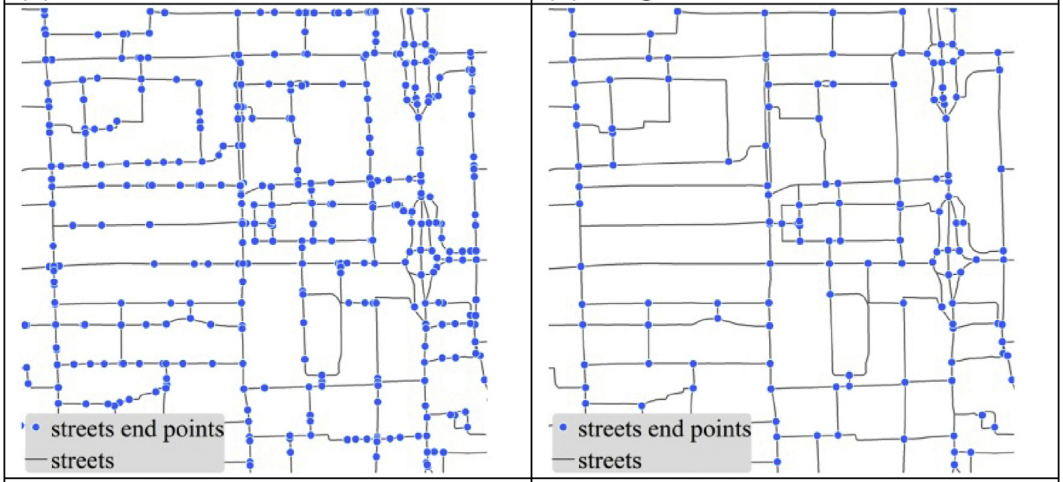
In this paper, we repeat the method in Section 2.1 to partition road junctions of China to derive urban areas. The edge length thresholds ranging from 50 to 10,000 m at an interval of 50 m have been adopted for obtaining the condensation threshold (Fig. 5). Each cluster has at least one road junction. We find that 300 m is

---

[4] DMSP/OLS has been updated by VIIRS/DNB since 2012. VIIRS/DNB has several improvements over DMSP/OLS like no saturation and collecting low light imaging data.

[5] The Chinese version of Yelp.

(a) Raw roads (b) Merged roads



(c) Before topology processing (d) After topology processing

Fig. 3. Deriving road junctions from raw road networks within an exemplified area.

the first convergence point, like the condition in the Greater London (Masucci et al., 2015). This indicates the universal law applied in the western European countries also works in the Far East country. We then use the 300 m for deriving urban areas.

Various patterns can be observed while we plot the top ten clusters at different edge length thresholds in the whole China (Fig. 6). For instance, the extensively-discussed east-west divide (Huhuanyong Line, see Tian et al., 2004) appears at 5000 m. City regions can be detected between the results at 2000 m and 1000 m, and the profile of cities sprouts from 500 m.

Urban areas are then extracted in China. At the edge length threshold of 300 m, totally 275,049 clusters are derived. Considering the redefinition of cities proposed in Section 3.2, we derive 4629 cities which have at least 100 road junctions (Fig. 7a). Totally 4.98 m junctions exist in these cities among all 8.24 m junctions. These cities correspond to a total urban area of 64,144 km^2 (49,743 km^2 for 643 official cities in 2014). We also plot each city as a point to reflect the city hierarchy of China (Fig. 7b) by using the head/tail division proposed by Jiang (2013) for determining the categories. The urban area pattern highlights the three most important city regions (Beijing-Tianjin-Hebei, Yangtze River Delta and Pearl River Delta) and the city hierarchy indicates the underestimated city system in the Yangtze River Delta. It should be noted that urban developments in Pearl River Delta have already merged into one mega city, which is different from the polycentric structure in the other two city regions.

We have shared the redefined Chinese city system in 2014 online. It can be accessed via http://geohey.com/apps/dataviz/517d9c0ad0d94e3b956d2f69f987909e/share?ak=ZmYzNmY0ZWJhYjcwNGU2ZGExNDgxMWUxNmZiOWNhNGY.

4.2. Evaluating the redefined Chinese city system in 2014

We evaluate all the redefined 4629 cities in 2014 mainly from the morphological, functional and social dimensions. For the morphological dimension, the mean value of road junction densities of all redefined cities is 78.0 per km^2 at the redefined city level, exhibiting a skewed normal distribution for all cities. For the functional dimension, the mean value of POI densities of all redefined cities is 92.5 per km^2 at the redefined city level, exhibiting a long-tailed distribution for all cities. For the social dimension, the average value of Dianping comment densities of all redefined cities is 686 comments per km^2 at the redefined city level, and cities are also long-tailed distributed. The cities with greater Dianping comment density tend to distribute within the existing administrative boundaries of cities. The differences among probability distribution of various dimensions indicate that physical world like transportation infrastructure tends to be normally organized and the social space like urban function and human activities is more scaling. We find no significant correlation among road junction

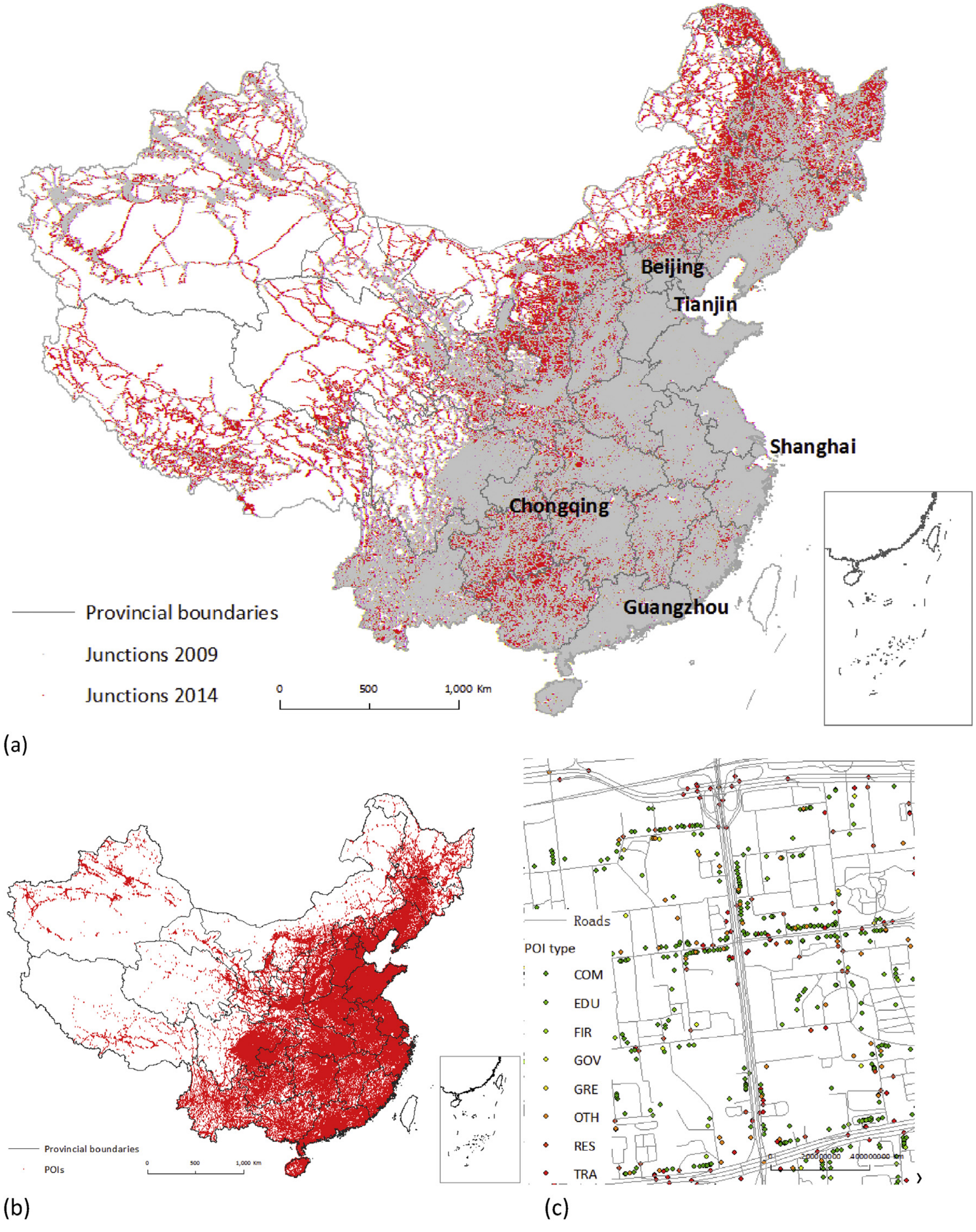


Fig. 4. Road junctions and POIs of China (a) road junctions in the whole China, (b) POIs in 2014 (the whole China, (c) POIs in 2014 (a part of the central city of Beijing).

density, POI density, Dianping comment density and city size in terms of road junction count at the redefined city level ($N = 4629$). In addition, we find that the 4629 cities cover 61.2% road junctions, 83.4% POIs, and 94% Dianping comments among all of them in the whole China, which indicates that our redefined city system is more like human activity agglomerated areas than urban function and physical developments.

Considering the morphological and functional evaluation dimensions, we classify all the Chinese redefined cities in 2014 into four types, including 726 vibrant cities, 1660 sprawl cities, 1061 vibrant and sprawl cities, as well as 1182 “ghost” cities, considering each redefined city's road junction and POI densities. The mean values aforementioned are used for classifying all these cities. Vibrant cities are associated with both above average road junction and POI densities, indicating the block size is small and with high urban function development. A large number of vibrant cities are presented in Fig. 8 and their size is large as well. Beijing and Guangzhou are the two largest vibrant cities. As the contrary of vibrant cities, sprawl cities have large blocks and are associated with very low urban function development. Most of sprawl cities

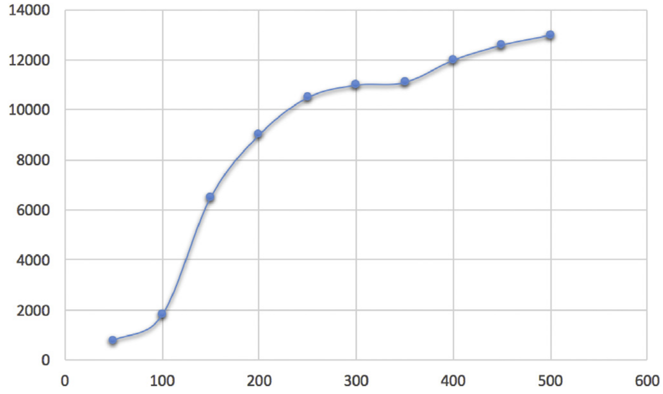


Fig. 5. Edge length threshold measured in meters for clustering road junctions (*x* axis) versus the maximum cluster size in terms of road junctions (*y* axis).

are not large as shown in Fig. 8, and Qingdao, Nanjing, and Shantou dominate the sprawl cities and the other largest sprawl cities are not within the existing administrative city boundaries. Vibrant and sprawl cities are those cities whose blocks are large but urban function is sufficient, which represent the other type of urban development model. The largest ones are Suzhou and Shaoxing in Yangtze River Delta. For the last type of cities, the so called "ghost cities" are these cities with small blocks but with less developed urban functions. That is, the city structures are deliberate designed and developed, but the necessary urban functions like shopping, dining, services and transport are very limited and lie behind the physical development. Most of these "ghost" cities are very small in size and the largest one is locating at Dongying, Shandong province. The other reason of these "ghost cities" may be the delayed development and urban function maturity.

4.3. *Evolution of the Chinese city system during 2009–2014*

The city system evolution of China during 2009–2014 has been detected via applying the methodology to the 2009 road junctions in China. The total city number increases from 2273 in 2009 to 4629 in 2014 (doubled), and the total urban areas increase from 28,405 km^2 in 2009 to 64,144 km^2 in 2014. The city size in terms of road junction count and city rank exhibit Zipf distribution in both years, and we see significant city size increasing in this period (Fig. 9a). The predefined cities provide us a new perspective for understanding urban expansion and redevelopments during 2009–2014 in China. For instance, urban redevelopments can be identified via analyzing the spatial pattern of road junctions in 2009 and 2014 for the existing urban areas in 2009 (Fig. 9b). Both physical redevelopments (road junction densification) and functional adjustments can then be identified. Urban redevelopments, as the prevailing discussion topic in China (appealing for urban redevelopment rather than urban expansion) is reserved for our future studies.

We pay more attention to urban expansion during 2009–2014 since it is the main form of Chinese urban developments and has been extensively studied. We aim to re-understand urban expansion process in China with the lens of the redefined city systems in both 2009 and 2014. As shown in Fig. 10, the land with a total area of 38,479 km^2 is expanded when we overlay urban areas in 2009 and 2014. Note that not all urban areas in 2009 are developed in 2014, thus making the expanded number being unequal to the gap between urban areas in 2009 and 2014. The yearbooks say urban areas increase from 41,175 km^2 in 2009 (654 cities) to 49,743 km^2 in 2014 (653 cities), and urban expansion areas are 8568 km^2 (20.8% of urban areas in 2009). We find that urban expansion speed in yearbooks has been largely underestimated when comparing it with our results, which may rise from the limited data quality of yearbooks, inconsistent statistical standards between 2009 and 2014 in yearbooks, as well as differed statistical scope (all cities versus only official cities) and overlap issue between our redefined cities in 2009 and 2014. Nevertheless, this finding deserves more attention for reexamining the urban expansion process in China.

In addition to the scale and speed of urban expansion in China, we are able to evaluate the urban expansion quality via comparing the urban areas in 2009 (old developments) and expanded urban areas during 2009–2014 (new developments). We find expanded urban areas are associated with 73.0% road junction density, and 25.3% POI density and 5.5% Dianping comment density, in contrast to existing urban areas in 2009. This finding indicates that the new developed areas, versus the old developed areas, are associated with large city blocks, lower urban function and much lower human activities. This is consistent with our common sense on "physical development first, followed by function fulfillment, and lastly people come". The spatial pattern of the quality evaluation on urban expansion may inspire the alternative findings on "ghost cities" (Chi, Liu, & Wu, 2015).

4.4. *Comparison of our results with existing ones*

We compare our redefined cities with urban areas in other data sources for examining our methodology. We find that 18,976 km^2 (66.8%) among 28,405 km^2 in the 2009 redefined cities is covered by the DMSP interpreted urban areas in 2009, and that 31,488 km^2 (49.1%) among all 64,144 km^2 in the 2014 redefined cities is covered by the VIIRS/DNB interpreted urban areas in 2014. This process is enabled by overlaying our redefined cities with the remote sensing images derived urban areas. In addition, considering the data resolution of urban areas from DMSP in 2009, we use the interpreted urban areas from Landsat TM images in the year 2010[6] by Liu et al. (2014) for comparing with our 2009 cities. There are 13,283 patches for 63,425 km^2 in total, and the mean patch size is 4.8 km^2, which is similar with that of urban areas from VIIRS/DNB in 2014. We find 72% of the cities in 2009 is covered by the urban areas in 2010.

We also compare our results with existing administrative cities from the hierarchical aspect. We visualize city hierarchies of the 2014 official and redefined cities using head/tail division as shown in Fig. 11ab. Our redefined city system and official one are both associated with a ht-index of 5. This may indicate that our redefined city system is as complex as official one, considering the definition of ht-index, which is proposed by Jiang and Yin (2014) for quantifying the fractal or scaling structure of geographic features. The higher the ht-index, the more complex the geographic feature. Significant mismatch can be observed for the first-tier cities in both settings. We plot the city rank versus city size for both systems as shown in Fig. 11cd. We find that our redefined city system follows the Zipf's law better than the official ones, since small official cities do not follow such distribution very well (Zipf, 1949, p. 1). This finding suggests that our redefined city system tends to be a more natural system compared with the existing official one which has been manipulated by the administrative process.

We overlay the redefined cities with the administrative boundaries of existing cities in China. Totally 602 existing cities have at least one redefined cities in 2014, and the other 51 cities do

---

[6] The 2009 or 2014 version is not available at the moment.

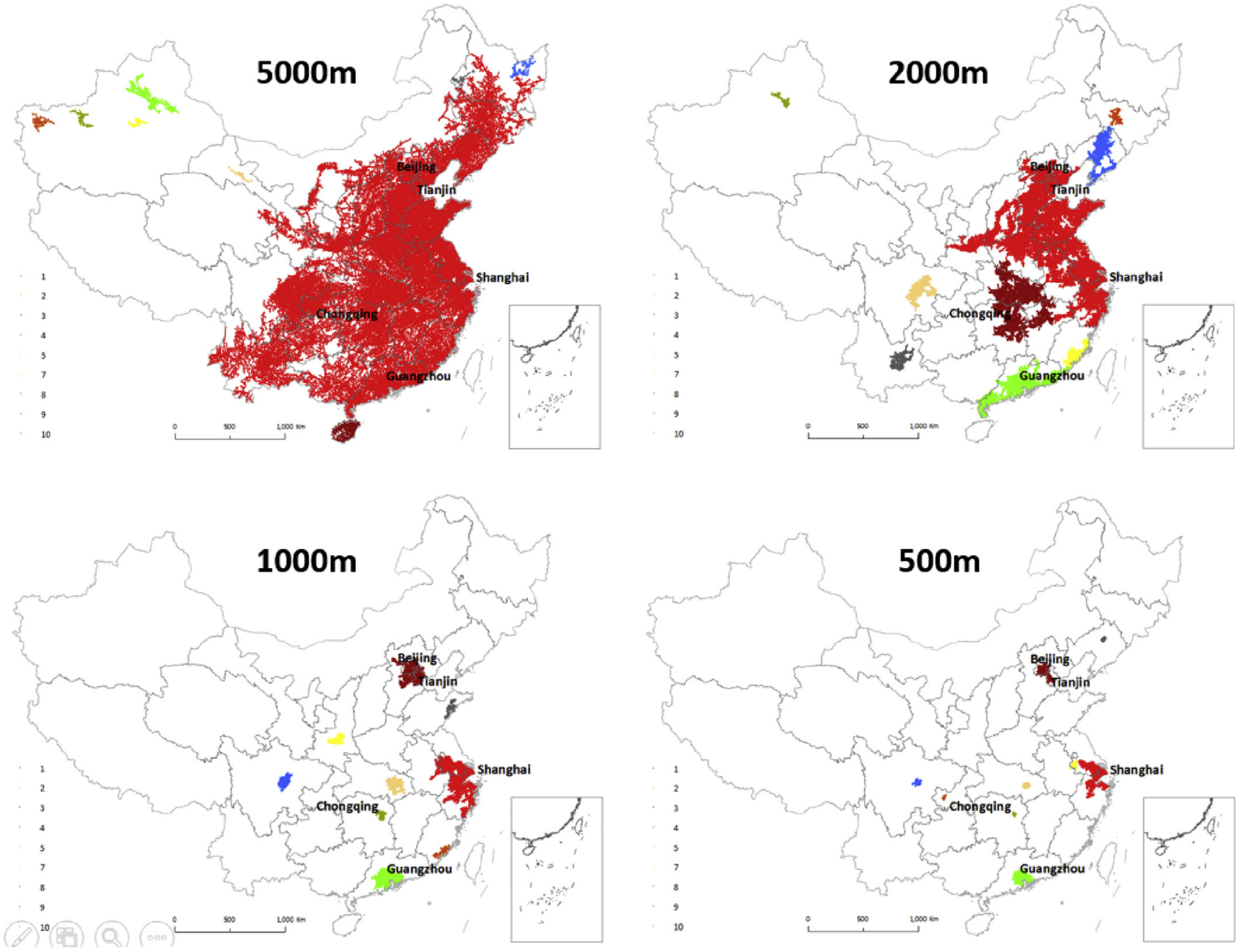


Fig. 6. Top ten clusters with different edge length thresholds (a color indicates a cluster). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

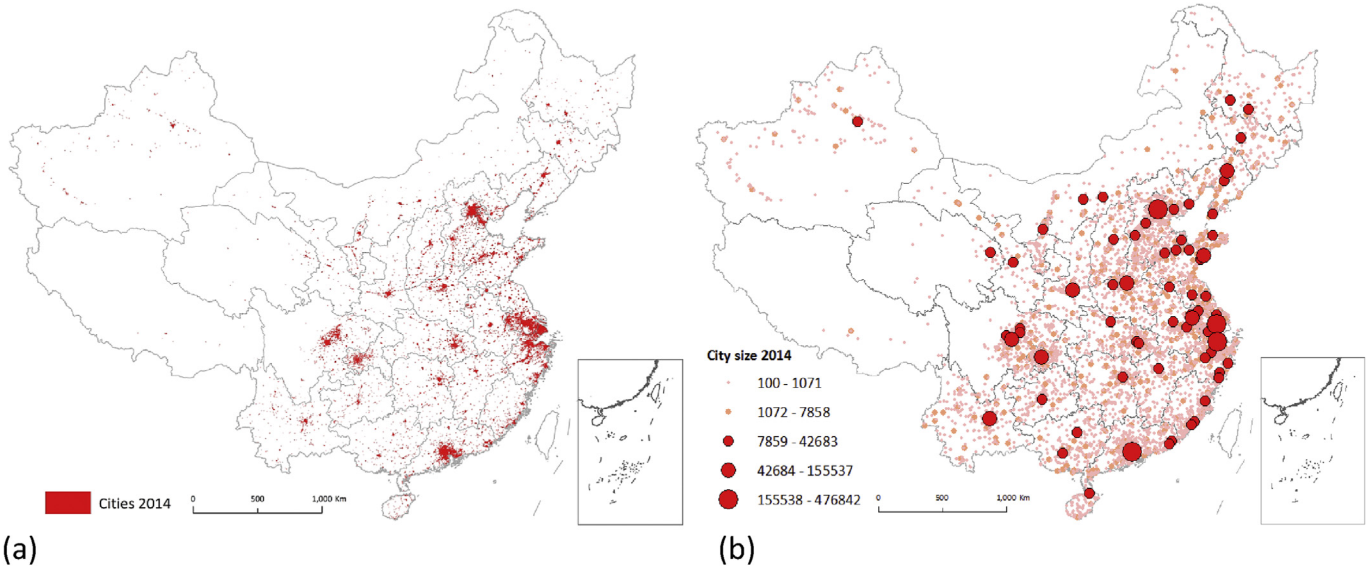


Fig. 7. Redefined Chinese city system in 2014 (a) urban areas; (b) city hierarchy in terms of road junctions.

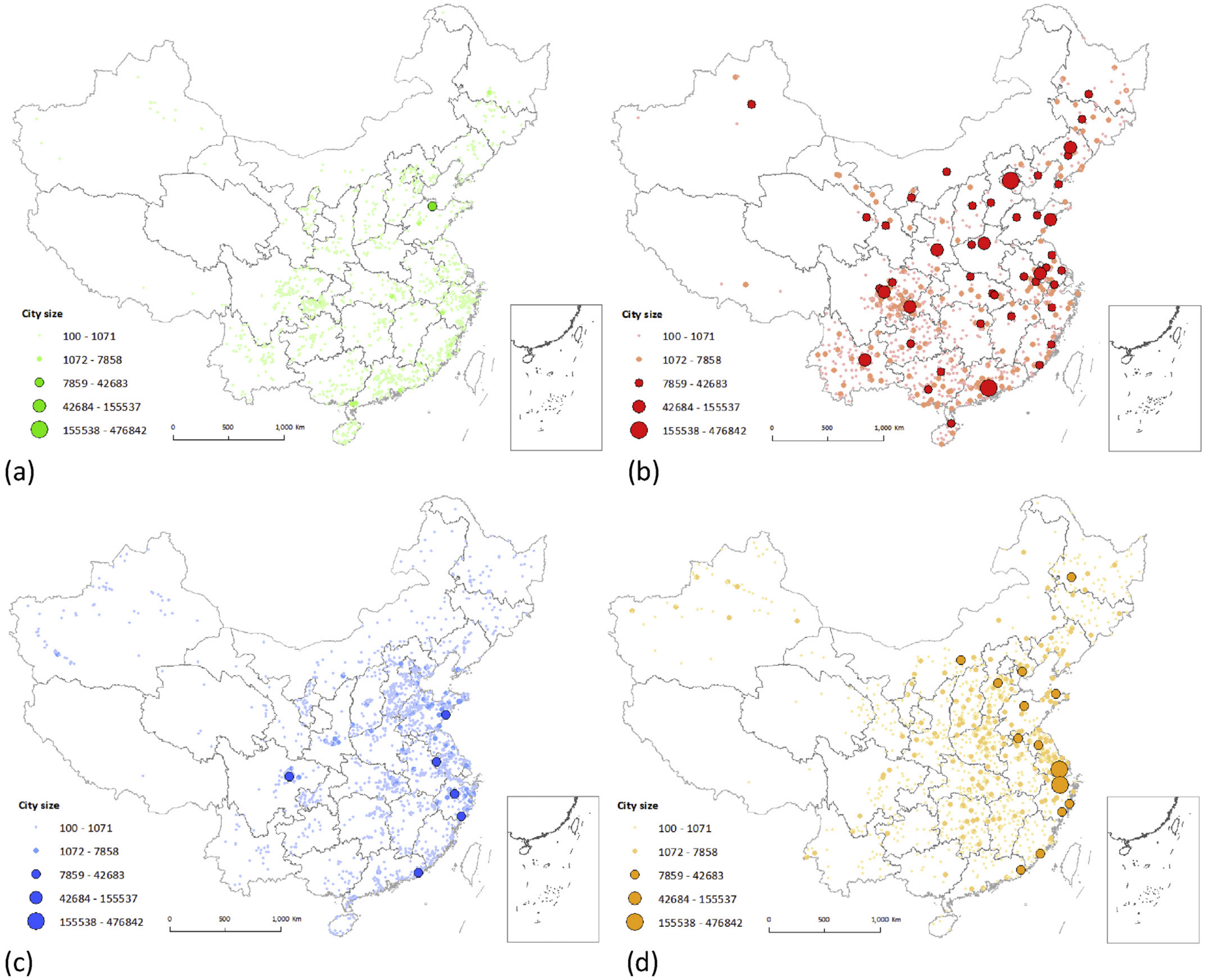


Fig. 8. Four types of redefined cities of China in 2014 based on the road junction and POI densities. (a): “ghost” cities; (b) vibrant cities; (c) sprawl cities; (d) vibrant and sprawl cities. Note: the city size is measured with of road junction count.

not have associated redefined cities. Most of these 51 cities are county-level cities which do not have a large urban development scale. Chongqing, Beijing and Wuhan have 87, 40 and 34 redefined cities, respectively, indicating the largest mismatches between spatial developments and administrative settings. Tianjin, Nanjing, Yangzhou, Qingyuan, Jinan, Kunming, Changsha, Shaoxing, and Shantou have over ten redefined cities as well, suggesting that these cities should also be divided as some separated cities. We also find that some prefectural level cities in the top three city regions Beijing-Tianjin-Hebei, Yangtze River Delta and Pearl River Delta, e.g. Langfang, Wuxi, Changzhou and Dongguan, are forming the same redefined cities with the central cities in these city regions, suggesting that the spatial developments in these regions go beyond the existing administrative boundaries.

5. Concluding remarks and discussion

5.1. Summary of the study

In this paper, we propose a methodological framework for redefining, identifying and evaluating Chinese city system in the light of open data, without relying on conventional yearbooks which have been extensively used by researchers interested with China in previous studies. A city is defined as “a spatial cluster with a minimum of 100 road/street junctions within a 300 m distance threshold” according to the iteratively spatial clustering process using various edge length thresholds by percolation theory.

We apply the methodology to the Chinese city system using national-wide road junctions in 2014. Totally we identify 4629 redefined cities with a total urban area of 64,144 km^2 for the whole China. Meanwhile, there are 653 official cities with a total urban area of 49,763 km^2 documented in statistical yearbooks, representing the existing administrative oriented Chinese city system. The differences between the redefined and official Chinese city systems are due to the mismatches between the administrative and spatial dimensions of cities. We then evaluate all the redefined cities from methodological, functional and social dimensions by using road junction, POI and online comment densities, respectively, and all cities are categorized into four types: vibrant, sprawl, vibrant and sprawl, as well as “ghost” cities.

The evolution of Chinese city system during 2009 and 2014 is

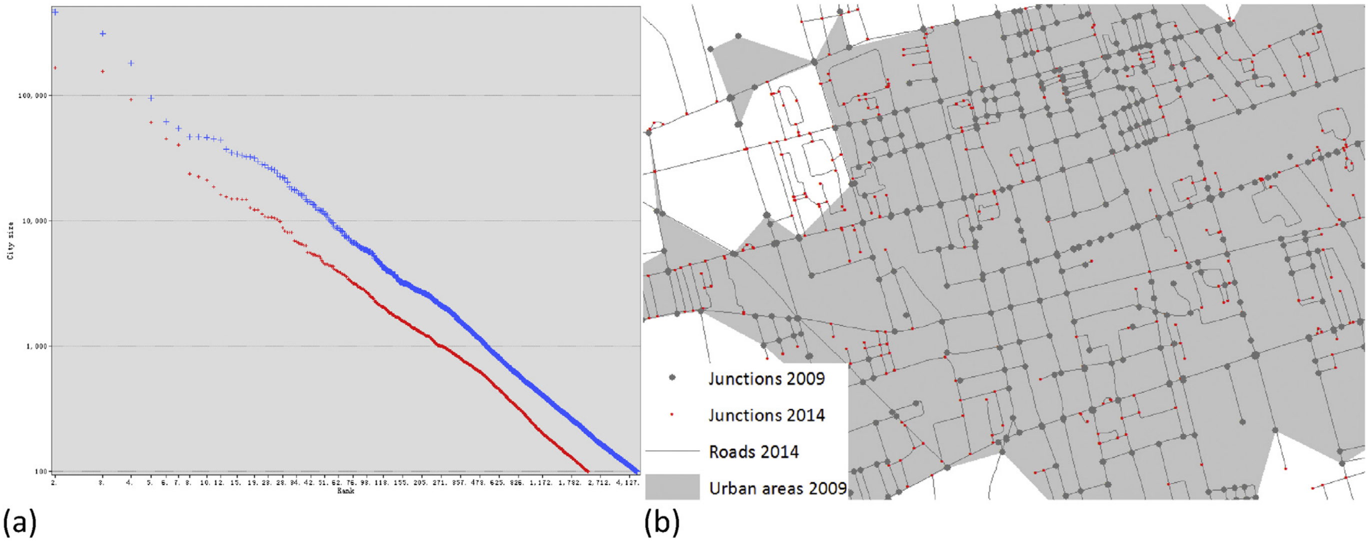


Fig. 9. Evolving city system of China during 2009–2014. (a) Rank size distributions; (b) A local area.

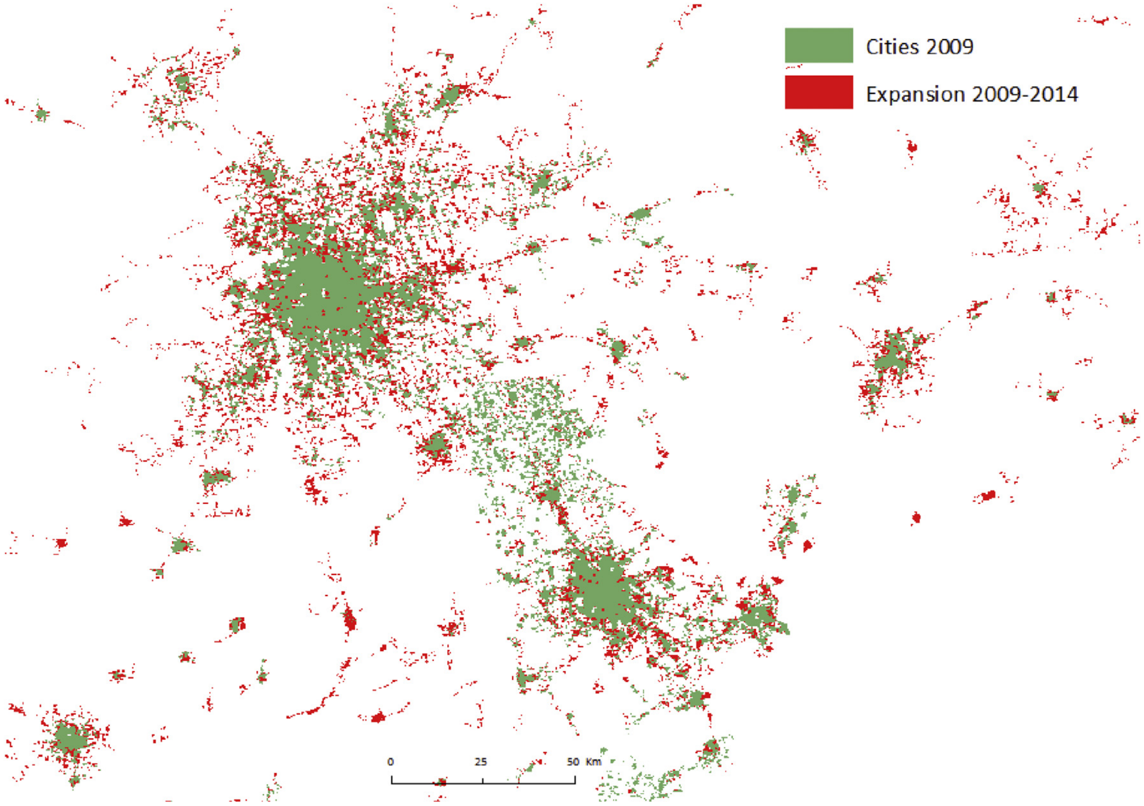


Fig. 10. Urban expansion during 2009–2014 in China (limited to the Beijing-Tianjin-Hebei area for better visualization in a limited page).

detected enabled by our repeating the experiments using the national wide road junctions in 2009. We observe total city number increases from 2273 in 2009 to 4629 in 2014, and an urban expansion of 38,479 km^2 in space. We notice that the urban expansion speed has been under estimated significantly according to the comparison between our "redefined" urban expansion and conventional urban expansion from yearbooks. In addition, we find that expanded urban area during 2009 and 2014, comparing with urban areas in 2009 are associated with 73.3% road junction density, 25.3% POI density and 5.5% online comment density. This indicates the newly developed urban areas have larger blocks, much lower urban function development, and much lower human activities than old ones, thus providing a big picture for the quality of urban expansion in China. In addition to urban expansion, urban redevelopments are also possible to be detected via comparing the redefined city system of China in 2009 and 2014 at the road junction level.

The redefined Chinese city system is also benchmarked with conventional city system in China for examining our methodology and its applications. The redefined city system of China in 2009

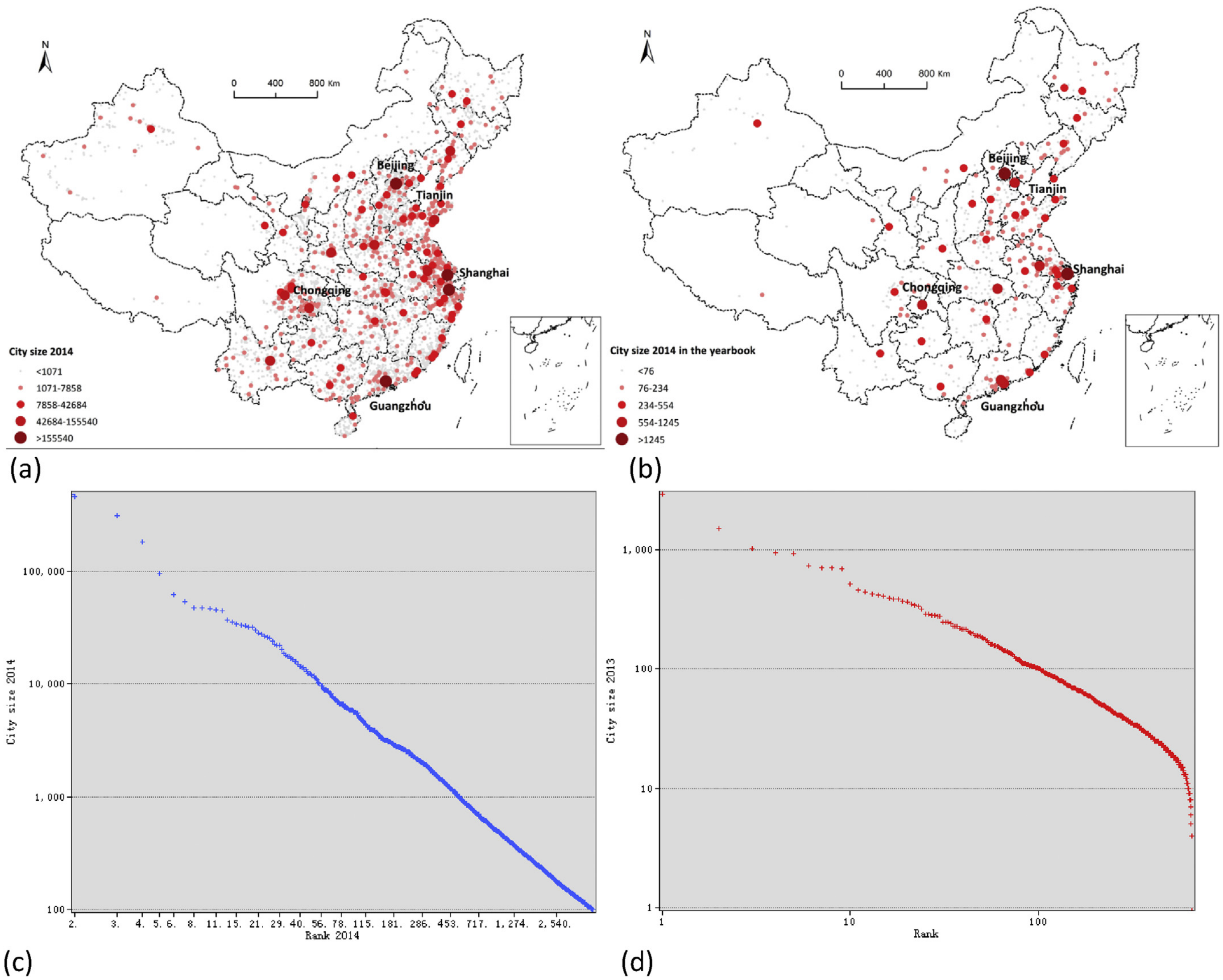


Fig. 11. Redefined Chinese city system versus official one in 2014 (a) City hierarchy of the redefined cities in 2014 (city size in terms of road junction count); (b) City hierarchy of official cities (city size in terms of urban areas in km^2); (c) Rank-size distribution of the redefined cities in 2014; (d) Rank-size distribution of official cities.

Table 2
The comparison between the redefined Chinese city system and the official one (2014).

Indicators	Redefined cities	Official cities
# cities	4629	653
Urban area in km^2	64,144	49,763
Probability distribution	Perfect Zipf	Zipf (excluding the tail part)
Frequency of information updating	Weekly or at least monthly	One year
Thinking philosophy	Bottom up	Top down
Available data for analysis	Big/open data	Yearbooks

overlaps 70% with urban areas interpreted from remote sensing images in 2010, indicating an acceptable degree between the two sources. We compare the redefined city system of China with the official one from various dimensions in Table 2. Our redefined city system in 2014 exhibits a better Zipf's law in contrast to the official city system in 2014, which may indicate a more beautiful and natural city system. The spatial relationship between the redefined Chinese city system in 2014 with the official one is also analyzed as well, and we find that the administrative boundaries of several large cities contain some redefined cities.

5.2. Academic contributions

This study aims to build a new paradigm on city system, rather than relying on conventional one. We see great opportunities of booming urban studies for the redefined city system with emerging new data. The contributions of our study mainly lie in the following aspects. **First**, the methodology we proposed is straightforward in redefining city system from the spatial dimension using ubiquitous open data that are increasingly available to researchers. Cities in the system can be benchmarked at the same horizon. It helps to mitigate the mismatch between the administrative dimension and the spatial dimension of cities. Referring to C. A. Doxadias's general hierarchical scale with fifteen levels of Ekistic Units,[7] the open data make it possible to view the Ekistic Unit 13 (small eperopolis) or above from the Ekistic Unit 3 (house), which was not possible at the time of proposing the Ekistic theory. We also notice that similar philosophy between the Ekistic theory and percolation theory, which are both looking at the world from local to global. **Second**, the size, scale, pattern and hierarchy of redefined cities can be evaluated together with understanding these cities from various dimensions like morphology, function and social activity. This provides a big picture for a city system, and can be tested and extended in empirical urban studies in future. **Third**, city evolution at different temporal spans like as short as several months and as long as several years can be detected with the proposed framework. For instance, urban expansion is enabled to be compared and benchmarked with the existing developed areas through morphological, functional and social dimensions, thus providing an avenue for evaluating the quality of urban expansion. **Last but not least**, the application of the proposed methodology in China provides us a novel lens for better understanding the Chinese city system and its dynamics and transitions, thus supplementing the existing means by yearbooks. This would be extremely helpful for objectively evaluating the current urbanization, identifying urban problems and designing development strategies for China.

5.3. *Potential applications*

As we have discussed in the first section of the paper, the mismatches between the administrative aspect and the spatial entities of existing Chinese cities is increasing in the background of spatial adjustment of administration. In addition to the academic contributions, we expect that our study would help the institutional design for planning a spatial and administrative matched city system, and contribute to the following applications in the real world like China. **First**, this study contributes to our understanding on these neglected Chinese cities like county cities and large towns in developed regions, which are absent from the official Chinese city system. We suggest include these cities into city management in China. **Second**, each of some large cities corresponds to several redefined cities, which suggests the existing city administrative system requires necessary adjustment to mitigate the mismatch in between. **Third**, as a supplement to yearbooks, city status could be monitored through the proposed framework, enriched with these merging new data accumulated for months, years or even decades. Their evolvement impact can be detected with other open data as well. **In sum**, not to mention its academic inspiration on other studies, the potential applications of this study are not limited to the aforementioned items for practice.

5.4. *Potential bias and future steps*

This study proposes a novel paradigm for understanding city system with ubiquitous open data. The increasingly available open data and the straightforward methodology used in this framework promote the merits of this approach. Nevertheless, several limitations still exist in the current study, which would be highlighted in our future research. **On one hand**, we use the city size threshold as 100 road junctions (about 4 km^2 in space), which we will sought for a more objective means for determining. **On the other hand**, more open data sources are expected to test our proposed framework to redefine cities from other perspectives. For instance, remote sensing based urban mapping at regional/global scale like Liu (2016) and Zhang, Li, and Hu (2016) presented at IGARSS is a potential new data source for replicating the proposed framework.

Acknowledgments

This work is supported by a grant from National Natural Science Foundation of China (No.51408039). The author would like to thank Mr Peng Cao for his data processing. The thanks also go to Dr Kang Wu for his comments on the draft of this paper, and Dr Qingxu Huang for providing the urban areas interpreted from night images in 2009 and 2014. Most importantly, the discussion with Prof Qizhi Mao from Tsinghua University and Mr Yao Shen from University College London greatly stimulated and inspired this work. The mapping results are available for downloading at the Beijing City Lab website (www.beijingcitylab.com).

References

Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, P., et al. (2015). *Hierarchical organisation of Britain through percolation theory*. arXiv preprint arXiv:1504.08318.

Batty, M. (2006). Rank clocks. *Nature, 444*(7119), 592–596.

Berry, B. J. L., Goheen, P. G., & Goldstein, H. (1969). *Metropolitan area definition: A re-evaluation of concept and statistical practice* (Vol. 28). Washington: US Bureau of the Census.

Chi, G., Liu, Y., & Wu, H. (2015). *Ghost cities analysis based on positioning data in China*. arXiv preprint arXiv:1510.08505.

Cottineau, C., Hatna, E., Arcaute, E., & Batty, M. (2016). Diverse cities or the systematic paradox of urban scaling laws. *Computers, Environment and Urban Systems* (in press).

Deng, X., Huang, J., Rozelle, S., & Uchida, E. (2010). Economic growth and the expansion of urban land in China. *Urban Studies, 47*, 813–843.

He, C., Liu, Z., Tian, J., & Ma, Q. (2014). Urban expansion dynamics and natural habitat loss in China: A multi-scale landscape perspective. *Global Change Biology, 20*(9), 2886–2902.

Hu, Y., Wu, Z., Xiong, W., & Pan, C. (2008). Study of identifying urban built-up area: Taking Wuhan as an example. *City Planning Review, 32*(4), 88–92.

Jiang, B. (2013). Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer, 65*(3), 482–494.

Jiang, B., & Jia, T. (2012). Zipf's law for all the natural cities in the United States: A geospatial perspective. *International Journal of Geographic Information Science, 25*, 1269–1281.

Jiang, B., & Miao, Y. (2015). The evolution of natural cities from the perspective of location-based social media. *The Professional Geographer, 67*, 295–306.

Jiang, B., & Yin, J. (2014). Ht-index for quantifying the fractal or scaling structure of geographic features. *Annals of the Association of American Geographers, 104*(3), 530–540.

Krugman, P. (1996). Urban concentration: The role of increasing returns and transport costs. *International Regional Science Review, 19*(1-2), 5–30.

Liu, X. (2014). Visualizing urban gastronomy in China. *Environment and Planning A, 46*(5), 1012–1013.

Liu, L. (2016). *Monitoring urbanization dynamics in the Southeast U.S. using DMSP/OLS nightlight time series*. IGARSS. http://igarss2016.org/Papers/PublicSessionIndex3.asp?Sessionid=1088.

Liu, J., Kuang, W., Zhang, Z., Xu, X., Qin, Y., Ning, J., et al. (2014). Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s. *Journal of Geographical Science, 69*(1), 3–14.

Liu, X., & Long, Y. (2016). Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design, 43*(2), 341–360.

Liu, J., Zhan, J., & Deng, X. (2005). Spatio-temporal patterns and driving forces of urban land expansion in China during the economic reform era. *AMBIO: a journal of the human environment, 34*, 450–455.

Long, Y., & Liu, L. (2015). Big/open data in Chinese urban studies and planning: A review. *ISOCARP Review, 11*.

Long, Y., Shen, Y., & Jin, X. (2016). Mapping block-level urban areas for all Chinese cities. *Annals of the Association of American Geographers, 106*(1), 96–113.

Mao, Q., Long, Y., & Wu, K. (2015). Spatio-temporal changes of population density and exploration on urbanization pattern in China: 2000-2010. *City Planning Review, 39*(2), 38–43.

Masucci, A. P., Arcaute, E., Wang, J., Hatna, E., Stanilov, K., & Batty, M. (2015). *Logistic growth and ergodic properties of urban forms*. arXiv preprint arXiv:1504.07380.

Masucci, A. P., Stanilov, K., & Batty, M. (2013). London's street network dynamics since the 18th century. *PLoS ONE, 8*, e69469.

Ministry of Housing and Urban-rural Development of the People's Republic of China (MOHURD). (2009). *Chinese city construction statistics yearbook 2009*. Beijing: China Planning Press.

Ministry of Housing and Urban-rural Development of the People's Republic of China

---

[7] See https://en.wikipedia.org/wiki/Ekistics.

(MOHURD). (2014). *Chinese city construction statistics yearbook 2014*. Beijing: China Planning Press.

Tian, Y., Chen, S., Yue, T., Zhu, L., Wang, Y., Fan, Z. M., et al. (2004). Simulation of Chinese population density based on land use. *Acta Geographica Sinica, 59*(2), 283–292.

Wu, K., Long, Y., Mao, Q., & Liu, X. (2015). Mushrooming Jiedaos, growing Cities: An alternative perspective on urbanizing China. *Environment and Planning A, 47*, 1–2.

Xu, M., He, C., Liu, Z., & Dou, Y. (2016). How did urban land expand in China between 1992 and 2015? A multi-scale landscape analysis. *PloS one, 11*(5), e0154839.

Zhang, X., Li, P., & Hu, H. (2016). *Regional urban extent extraction using multi-sensor data and decision rules*. IGARSS. http://igarss2016.org/Papers/PublicSessionIndex3.asp?Sessionid=1066.

Zhou, Q. (2015). Comparative study of approaches to delineating built-up areas using road network data. *Transactions in GIS, 19*(6), 848–876.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, Massachusetts: Addison-Wesley.