Article

Investigating the Complexity of Spatial Interactions between Different Administrative Units in China Using Flickr Data

Wei Zhu [1], Ding Ma [1,2], Zhigang Zhao [1,2],* and Renzhong Guo [1]

[1] Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China; 1800325001@email.szu.edu.cn (W.Z.); dingma@szu.edu.cn (D.M.); guorz@szu.edu.cn (R.G.)

[2] Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources, Shenzhen 518060, China

* Correspondence: zhaozgrisc@szu.edu.cn

check for updates

Abstract: Location-based social media have facilitated us to bridge the gap between virtual and physical worlds through the exploration of human online dynamics from a geographic perspective. This study uses a large collection of geotagged photos from Flickr to investigate the complexity of spatial interactions at the country level. We adopted three levels of administrative divisions in mainland China—province, city, and county—as basic geographic units and established three types of topology—province–province network, city–city network, and county–county network—from the extracted user movement trajectories. We conducted the scaling analysis based on heavy-tailed distribution statistics including power law exponents, goodness of fit index, and ht-index, by which we characterized a great complexity of the trajectory lengths, spatial distribution of geotagged photos, and the related metrics of built networks. The great complexity indicates the highly imbalanced ratio of populated-to-unpopulated areas or large-to-small flows between areas. More interestingly, all power law exponents were around 2 for the networks at various spatial and temporal scales. Such a recurrence of scaling statistics at multiple resolutions can be regarded a statistical self-similarity and could thus help us to reveal the fractal nature of human mobility patterns.

Keywords: Flickr; complexity of human dynamics; spatial interactions; power laws; ht-index

1. Introduction

The impetus of information and communication technology has, to a great extent, reshaped our understanding of the geographic space and human activities therein [1]. The geography-related research in the last century has relied mainly on conventional data gathered by statistical or census authorities (so-called small data) and technologies of geographical information system (GIS) and remote sensing. It would be problematic for research to be based solely on conventional data collected in a top-down manner. For example, the census tracts, which are the aggregation of individual-based data based on geographic units, can avoid privacy concerns, but lead to a very low spatio-temporal resolution and can be hard to update on a yearly basis. Likewise, it is impossible to update the database of high-resolution remote sensing imagery every hour, let alone every minute or every second. This circumstance is now shifting, because the advent of location-based social media (LBSM) such as Flickr, Twitter, and Foursquare provide unprecedented opportunity to acquire the individual-based location data with a very fine-grained spatio-temporal scale [2]. The emerging big data, formulated in a bottom-up manner, has much greater potential to capture the essence of reality than small data.

Urban complexity is an important aspect of urban sustainable development [3]. One of its core parts—human dynamics—has attracted considerable and sustained interest both in academia and industry. The term of human dynamics is originated from physics and can be simply regarded, in the field of geography, as the collection of human motions and movements across different places [4]. Before the invention of LBSM, the seminal work of Barabasi [5] has quantitively examined the human dynamics using complex systems or a complex-network thinking. Human mobility patterns were then widely studied by taking the forms of, for example, travels of dollar bills [6], and mobile phone records [7]. Later, the arrival of LBSM made researchers reach a consensus that human dynamics always occur in a hybrid virtual-physical spaces [1]. Accordingly, many geography-related studies extensively leveraged check-in locations or other types of geo-tagged user-generated content from LBSM to inspect socio-geographic human activities and interactions [8–12]. Most of these studies found that human movements are complex and heterogenous, as they demonstrate strikingly fractal or scaling structures and nonlinear dynamics, which can be well-characterized by power law fitting parameters [5,13,14]. However, they were conducted mostly at the city level or within a single period and may be short of a holistic picture of complexity of human dynamics at different resolutions. As LBSM data is with finer spatial and temporal granularities, it has been possible for us to model the multi-scale interactions between human activities and geographic space.

In this article, we integrate human dynamics of both the virtual and physical spaces in a GIS environment from the perspective of underlying spatial interactions between three administrative levels (provinces, cities, and counties) in China through Flickr data with a 12-year timespan (2002–2014). Flickr is a popular LBSM platform that allows photographers or travelers to upload and manage geotagged photos to share their travel experiences. The location and timestamp of each photo of the same traveler makes it possible to follow the traveler's displacements. For a collection of movements based on all users, one can take the flows between places as the proxy to investigate the spatial interactions at a wide spatial setting [15]. The spatial interactions can be reflected using the complex network, which consist of nodes for individual geographic administrative unit, and links for their interactions. The motivation of using a complex network lies in the degree of nodes [16,17]. A network with a complex structure is inclined to possess a group of nodes in which the links are denser than ones outside the group. In other words, the popularity of nodes in a complex network varies greatly from one to another. Hence, in a complex network such as a small-world [18] or a scale-free network [19], the network structure tends to be very heterogeneous, wherein the degrees of nodes are inclined to exhibit a scaling hierarchy, indicating that there are far more less-connected nodes than well-connected ones, or far more light-weighted connections than heavy-weighted ones between nodes.

The present paper intends to conduct a multi-scale, complex-network-based analysis of the spatial interactions for illustrating the complexity of human dynamics reflected by LBSM. The contribution of this study is three-fold. (1) We extracted tens of thousands of trajectories within mainland China from the Yahoo database, which is originally composed of 100 million records at the global coverage, in order to detect the imbalanced distribution of human activities over space. (2) We constructed networks at three spatial scales: the county–county network, the city–city network, and the province–province network, and introduced an area-based cartogram for avoiding visual clutters (see details in Section 2.2). All built networks had a strikingly scaling feature with respect to node degrees and edge weights. (3) We examined three types of network at different temporal granularities (whole timespan or annually); the power law exponent of each network was close to 2. Such statistical self-similarity at multiple resolutions reveals the fractal structure of the spatial interactions.

The remainder of this paper proceeds as follows. Section 2 displays the datasets and proposes the methodological framework. Section 3 presents the visualization and statistical results regarding the scaling analysis of user trajectories, spatial distribution of photo locations, and constructed network metrics. Section 4 discusses the results and Section 5 presents the conclusion and ideas for future research.

2. Methodology

2.1. Data and Data Processing

We selected China as the study area and adopted three different scales of administrative division (provinces, cities, and counties) as modeling units. Administrative boundary data were sourced from the National Resources and Environment Database of the Chinese Academy of Sciences [20]. Chinese urbanization has developed rapidly in the past few decades, but the manner in which this urbanization has occurred has differed between the eastern and western sides, as delimited by the "Hu Huangyong" Line [21] (Figure 1a). More specifically, unlike the outward urban expansion on the western side, the urbanization in the eastern side has resulted in several urban agglomerations, such as The Yangtze River Delta along with the southeastern coastline of China, which includes metropolises like Shanghai, Hangzhou, and Nanjing.
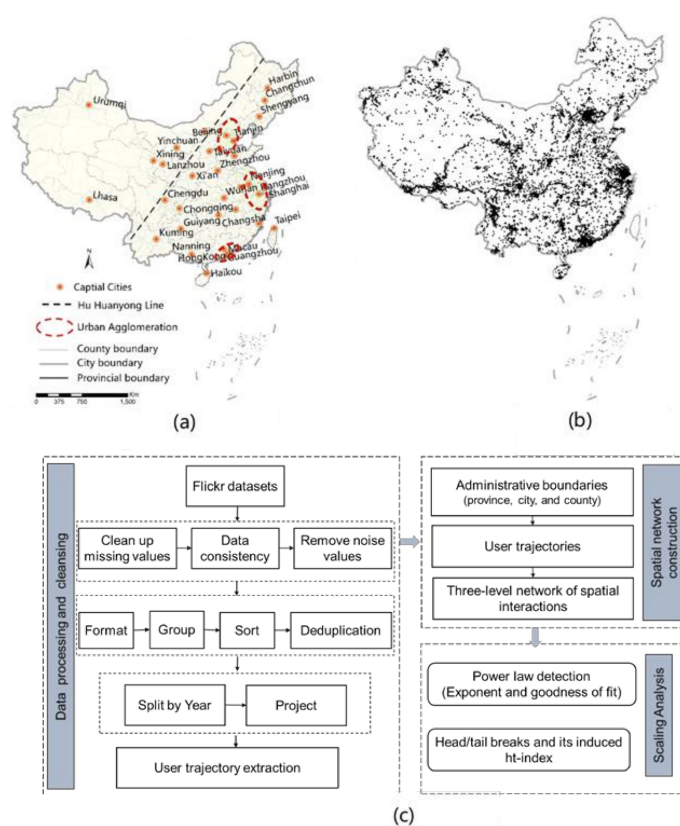

Figure 1. (Color online) The three administrative level boundaries in China (**a**), the spatial distribution of geo-tagged photos (**b**), and the workflow of data processing and analysis in this study (**c**).

Flickr Data comes from the Yahoo Flickr Creative Commons 100 Million Dataset [22], which contains the videos and pictures collected by Yahoo from 2002 to 2014. There are in total 100 million records, compressed into 10 files by the bzip2-compressed method. Each file contains roughly 10 million rows. Each row contains 23 columns containing different types of geo-tagged photo attributes, such as user ID, photo identifier, time when the photo was taken and uploaded, and location information like longitude and latitude. It should be noted that not all the columns were filled with values.

As Flickr data were gathered from millions of users who are from all over the world and with diverse backgrounds, it is necessary to clean and reformat the data before the analysis. Firstly, we omitted all photos whose locations are outside the study area, and we then deleted users who have only one geotagged photo because a trajectory contains at least two locations; in the meantime, we "deduplicate" geo-tagged photos within 10 m and 1 h of a single user to keep the noise of data at the

minimum extent. According to the above three procedures, we extracted the 708,191 geotagged photos of mainland China (excluding Taiwan) out of 100 million records (Figure 1b represents their spatial distribution). We further used only the related information, such as user ID, point ID, coordinate-pair, and timestamp for trajectory extraction. The structure of a single trajectory is represented as:

$$Trajectory(uID) = \{(x_1, y_1, t_1), (x_2, y_2, t_2) \dots (x_i, y_i, t_i)\} \tag{1}$$

where by each trajectory is built using a user ID and a list of photo locations sorted by photo uploading time. After trajectory extraction, we convert all geographic coordinates into plane coordinates under the China Lambert Conformal Conic projection for subsequent spatial calculations; for example, we performed spatial join to project each location onto different levels of administrative units. Figure 1c presents the workflow of data processing.

2.2. Network Construction and Visualization

The research built up networks based on each user's trajectory and from the geographic perspective to model the spatial interaction patterns. There are three types of networks in this study in relation to three levels of Chinese administrative units: county–county network, city–city network, and province–province network. In the network, administrative units (i.e., county, city, and province) represent nodes, while the edges between them denote the movements from one administrative unit to another. We first joined the location of each geotagged photo with each type of administrative unit, then linked administrative units *i* and *j* based on two consecutive locations in each extracted user trajectory. As the connection between two administrative units can be repeatedly counted, the network is a directed and weighted graph, where the weights w_{ij} equal to the number of duplicated connections (Equation (2)).

$$w_{ij} = \begin{cases} \geq 1, & \text{if } i \text{ and } j \text{ are connected} \\ 0, & \text{else} \end{cases} \tag{2}$$

It is equally important to visualize the network after it is built. Many specific layout algorithms can effectively meet this end, such as the circular layout [23] and the force-directed layout [24]; however, they did not prioritize each node's absolute geographic position as well as its relative one to other nodes. Given that each node embeds inherently with a specific geographic information (such as the boundary of an administrative unit), the simplest way is to assign each node the centroid of the boundary as its location to visualize the network in a geographic manner. As the number of check-in activities probably differs greatly from one place to another, such a visualization may not be capable of representing the dense area (small boundary size, but numerous in- and out-flows) where visual clutters may occur. In this respect, we reshaped the boundary of each administrative unit in the form of area cartograms using the diffusion-based method [25] to consider the effect of the number of geotagged locations. The enlarged space of populated administrative units enables a clearer display of denser or more complex spatial interactions. Besides, the relative geographic position of each node to other nodes can be properly preserved in the area-based cartogram.

2.3. Scaling Analysis on the Network Metrics

A power law model can be expressed by an accurate expression as Equation (3) or a probability distribution as Equation (4):

$$y = x^{-\alpha} \tag{3}$$

or

$$(x) \propto x^{-\alpha} \tag{4}$$

To identify a power law distribution, it is natural to first rank the data increasingly according to the frequency or quantity and then log-log plot the sorted sequence to have a preliminary verdict.

A distribution is likely to obey the power law model if the log-log graph of its data points appears as a straight line in the plot. In practice, however, most datasets are tailed with messy up-and-downs so that it is labored to deal with [13]. This problem could be addressed by the maximum likelihood estimation [14], which facilitates the power law detection in a more accurate way. This method adopts two parameters: power law exponent, α, and the goodness-of-fit index, p-value. The exponent α is denoted as Equation (5):

$$\alpha = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}} \right]^{-1} \tag{5}$$

where x_{min} is the smallest value from where the data is power-law distributed. Typically, it is acceptable for α to be located in [1,3]. The goodness-of-fit index p-value, ranging from 0 to 1, indicates the extent of fitness for the data to a power law distribution. The greater a p-value is, the better the fitness would be. The baseline of p-value is set as 0.1 in this study.

Commonly, a power law distribution demonstrates a substantial heterogeneity. However, a heterogenous characteristic could also be embedded in others such as lognormal and exponential distributions, which belong to the kin of heavy-tailed distributions. In order to resolve this issue, Hanel et al. [26] provides a decision tree to determine power law exponents under different heavy-tailed data assumptions. Another way to enclose more extended distributions instead of the power law alone could be the head/tail breaks method [27]. In short, given data with a heavy-tailed distribution, it can be divided into bisections (that is, a head and a tail) at the arithmetic mean in a recursive way. The recursions occur in each head on a ratio of 40–60 percent until the head can no longer be divided. The number of acquired partitions is defined as the ht-index [28], an indicator of the extent of heterogeneity. In other words, the ht-index indicates the number of times a scaling pattern of far more smalls than larges recurs, thereby capturing the scaling property of a dataset. Basically, a distribution is heterogeneous if its ht-index is greater than 3. The higher the ht-index, the more heterogeneous the data are. In sum, we computed three parameters for conducting scaling analysis on both network node degrees and edge weights: the power law exponent α, the goodness-of-fit index p-value, and the ht-index value ht.

3. Results and Discussion

The findings from the analysis of interactions between different Chinese administrative units are represented through three aspects. First, user trajectories are complex at both the individual and collective levels, as its length and number of contained locations follow long-tailed distribution statistics, such as power law exponent and p-value. Second, the distribution of photo locations is imbalanced among places at each spatial scale. Such an imbalance of populated-to-unpopulated places can also be characterized by power law statistics. Third, the constructed networks can effectively capture the province–province, city–city, and county–county interactions in China, as network node degrees and edge weights correlate highly with the level of development of each administrative division. More interestingly, the multiscale networks reveal a statistical self-similarity, and thereby hold an obviously fractal nature.

3.1. Scaling Properties of User Trajectories and the Distribution of Photo Locations

After data filtering, 708,191 photo locations and 10,820 users remained. By grouping the photo locations according to individual user and publishing time, 8191 trajectories were obtained. We started by inspecting each extracted user trajectory from two aspects: trajectory length and number of geo-tagged photos. Both datasets show apparent scaling patterns indicated by either power law fitting metrics or a big ht-index value. Specifically, trajectory length followed a power-law distribution with exponent = 2.56 and p-value = 0.16, both of which meet the condition of the power-law fitting; the number of photo locations did not pass the power law test, as the p-value = 0, but it still holds a strikingly scaling pattern as its ht-index value is eight, which indicates that the imbalanced ratio

between many-to-few photos occurs seven times. As seen in Table 1, a low head percentage occurs repeatedly at almost each level (below 35 percent). The scaling property of user trajectories indicates that only a minority of users travel a lot, and it is most likely that they contributed the majority of geo-tagged photos.

Table 1. Statistics of head/tail breaks on the number of geo-tagged photos per trajectory (Note: # = number).

#Photos	#head	%head	#tail	%tail	mean
8191	1550	18%	6641	82%	78
1550	384	24%	1166	76%	339
384	104	27%	280	73%	881
104	33	31%	71	69%	1854
33	13	39%	20	61%	3158
13	4	30%	9	70%	4247
4	1	25%	3	75%	5786

We then examined how geo-tagged photos were distributed in space; that is, how many photos were in each province, city, and county. The results show that the spatial distribution of all photos at each scale is perfect-power-law distributed (Figure 2d), indicating far more less-populated administrations than well-populated ones. The results further imply a high variation of attractiveness from one administration to another. In this way, we can spot the complexity of the movement dynamics among different spatial units. At the city level, for instance, the number of photo locations ranged from 1 to 144,209 and the ht-index value of those numbers is 5. We made use of the area-based cartogram to visualize such variation, from which we can see the top five cities are Beijing, Hongkong, Shanghai, Guangzhou, and Zhuhai. All those cities have high economic or political status. The ranks of provinces and counties also correlate with that of cities.
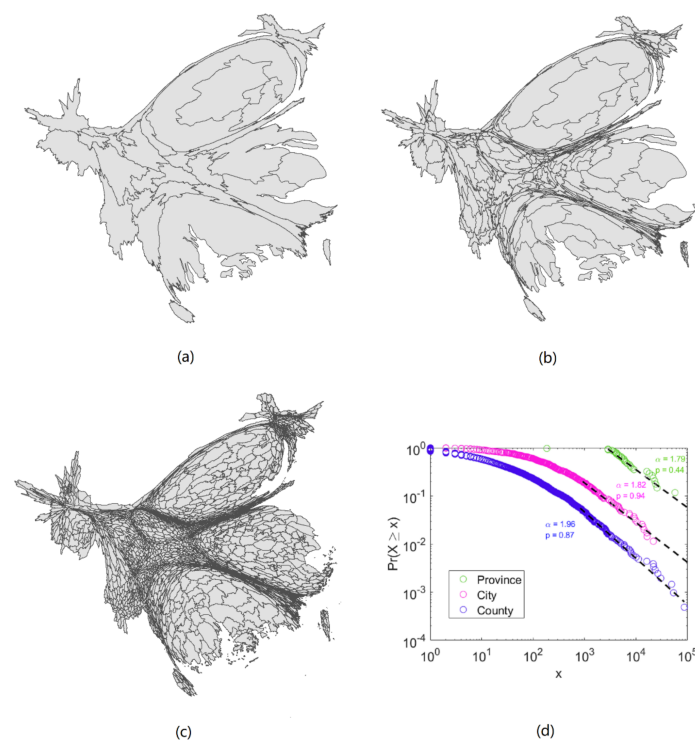

Figure 2. The cartogram showing the spatial distribution of geo-tagged photos in China at province (**a**), city (**b**), and county (**c**) level, respectively, and the power-law distributions of #photos per administration (**d**).

3.2. Analysis of the Constructed Networks

With the extracted user trajectories and administrations at different levels as of 2014, we then established the province–province network, city–city network, and county–county network to investigate the underlying complexity of spatial interactions between different administrative divisions in China. The corresponding metrics of three types of networks are given in Table 2. In sum, all networks are scale-free networks, indicated by either power law metrics or ht-index values. For nodes, there are far more small provinces/cities/counties than big ones regarding location number. For edges, there are far more light-weighed flows than heavy-weighted ones between each type of administration. Such a structure is greatly consistent with the uneven spatial distribution of photo locations, or place attractiveness. Figure 3 shows the network at each level with two types of geographic layout. We can observe that the network contains four or five hierarchical levels (as indicated by ht-index in Table 2) regarding both the node degrees (dot size) and edge weights (line color and width). Besides the scaling pattern illustrated by those hierarchical levels, we can see that most heavy-weighted connections were concentrated within three highly developed regions in China (Figure 1a): the Yangtze River Delta, Beijing–Tianjin–Hebei, and the Pearl River Delta (or Guangdong–Hong Kong–Macao Greater Bay Area).

Table 2. General parameters of the three constructed networks.

	Province–Province	City–City	County–County
#Nodes	33	349	1957
#Edges	886	5245	16,142
Ht-index (Ndeg)	5	4	4
Power-law (Ndeg)	2.06 (0.46)	1.92 (0.94)	1.71 (0.09)
Ht-index (Eweg)	4	4	5
Power-law (Eweg)	2.08 (0.1)	2.34 (0.26)	2.21 (0.25)

(Note: # = number, Ndeg = Node degree, Eweg = Edge weight, decimal numbers out/in brackets = alpha/*p*-value).

We dug further into the power-law-fitting metrics of the constructed networks. We could remark that the power law exponent of each network, for either node degrees or edge weights, remains close to 2. We then examined the networks of previous years in order to see whether such an exponent value in terms of node degrees always persists on an annual basis. The results can be described in two aspects: (1) average node degrees first increased from 2002 to 2010, then decreased dramatically in the following years (Figure 4a–c); (2) although the average node degree varies every year, the node degrees of each year's network pass the power law test (with p-value ≥ 0.1). Moreover, the variations of scaling exponents were very small: the largest fluctuation was 2 ± 0.4 for province–province networks (Figure 4d), while for city–city or county–county networks the fluctuation was 2 ± 0.2 (Figure 4e,f).
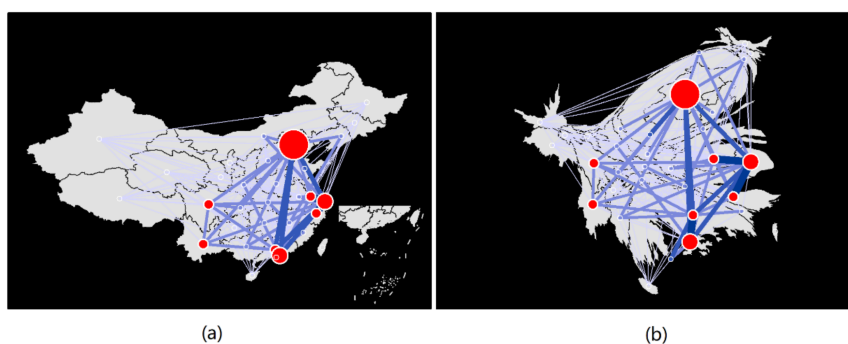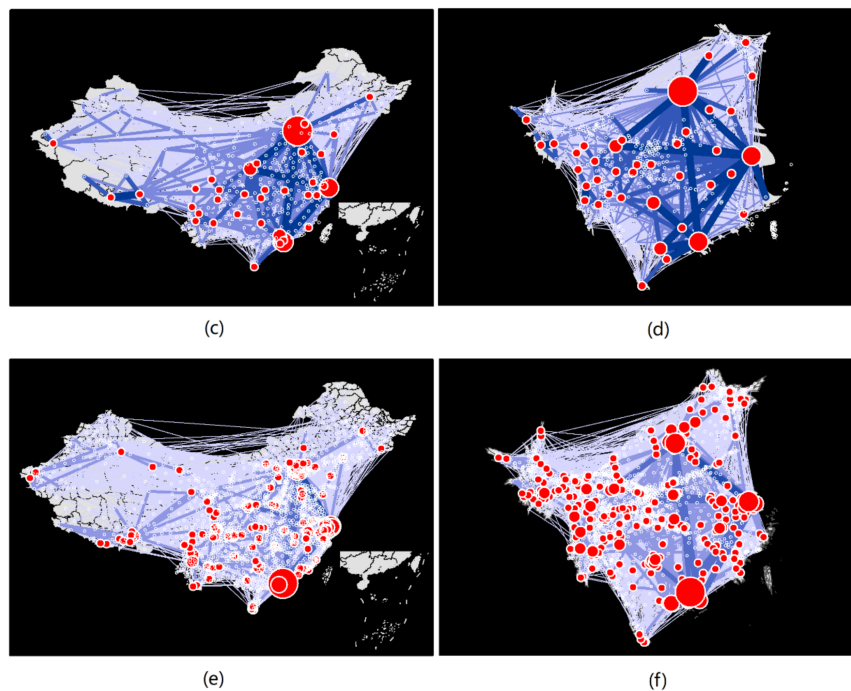

(a)　　　　　　　　　　　　　　　　　　　　　(b)

Figure 3. *Cont.*

Figure 3. (Color online) Constructed networks showing the spatial interactions between three-level administrative divisions (Note: Panels **a,c,e** are the networks with geographic-correct administrative boundaries, while panels **b,d,f** are the networks with geographic-distorted (area-based cartogram) ones. The distorted boundaries lead to a more discernible display of networks in the dense areas.).
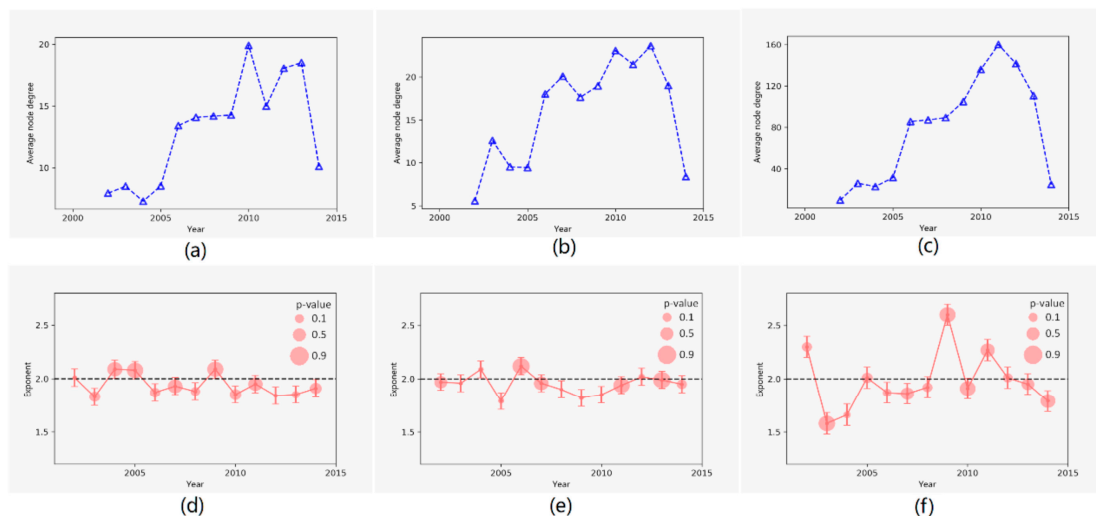

Figure 4. (Color online) The curve graph of average node degree at county, city, and province levels (Panels **a–c**), power law exponent with its related *p*-value (Panels **d–f**) of each type of network on a yearly basis from 2002 to 2014.

4. Further Discussion of This Study

Supported by the mobile devices embedded with the GPS module, LBSM such as Flickr provide enormous spatio-temporal data from tens of thousands of users with diverse backgrounds and with a very high resolution in terms of meters in space and seconds in time. As one typical type of the geospatial big data, LBSM data is essentially bottom-up and has three distinctive characteristics: all, measured, and individual-based [29]. Therefore, LBSM data provides a better alternative to the conventional data for studying the complexity of geographic space and human dynamics therein.

Having presented the results in Section 3, the following content of this section further discusses some implications of the results.

There is a great complexity of the Flickr data from the user trajectories to the number of photo locations in each administration, and further to the spatial interactions between administrations. The complexity can be well-reflected by the maps of built networks, and long-tailed distribution statistics including power law fitting metrics and big ht-index values. Overall, the obtained results indicated an imbalanced concentration of human activities in China from one place to another. Such an imbalance was also consistent with the spatial differentiation of Chinese economic developments [30,31]. If we look at cities on the two sides of "Hu Huanyong" Line, as Figure 1a shows, which splits the area of China into two approximately equal parts (western and eastern), it can be observed that the most active three regions, illustrated by the networks (Figure 3), reside in the biggest urban agglomerations the eastern part, while the western part hardly contains nodes or edges at a higher scaling hierarchical level.

The Flickr data, that is; geo-tagged photos with fine-grained spatio-temporal information, facilitates the development of insights into the nonlinear dynamics of big data. Normally, a complex urban system, such as finance and economics, often demonstrates such nonlinear statistics which is, to some extent, uncovered by the exponential, especially power-law distribution [32]. It might not be of big surprise that the spatial interactions exhibit power-law distribution statistics, as such a scaling pattern is likely to be determined by the city size distribution, which has been confirmed a power-law in China [30]. Meanwhile, the asymmetric connections between cities, revealed by the transportation network [33], can also explain, to a certain extent, the formation of highly imbalanced ratio of large-to-small flows in the constructed networks. However, they are still unable to grasp the entire picture behind the power law of spatial interactions represented by LBSM. To be specific, the power law exponent of city-city network degrees (1.92) shows not much difference from that of the city-size distribution (1.82), but the flows between cities are with a larger power exponent (2.34). This implies that the distribution of interactions is related to, but simultaneously more complex than, the city-size distribution. We conjecture that it is because of the effect of endogenous interactions within a city or region. In this connection, the spatial interactions from LBSM follow the strategy of the endogenous network formation [34–37].

Another interesting finding is that we found that power law exponents of networks at three spatial scales were around 2 (with a deviation of 0.2 in most cases), meaning that they also follow Zipf's law [38]. Previous studies have mostly confirmed such a statistical regularity at the city scale across countries [13,39,40], and used Zipf's law for cities as evidence of the self-organized behavior of a complex urban system [41]. The present study further found that Zipf's law held not only for cities, but also for provinces and counties, as well as their interactions. Furthermore, the similar power law exponents of network degrees at different levels of administrations exhibit a statistical self-similarity spanning over a range of spatial scales, which enables us to see the incredible fractal structure of human activities in social media. Such a fractal structure of human activities reflected by the multiscale networks is akin to that of a biological system. A recent study by Zheng et al. [42] uncovered the self-similarity of the multiscale structure of the human connectome. Similarly, García-Pérez et al. [43] stated that the structure of human metabolic network also reveals self-similarity across a series of scales. In essence, such a fractality of the system is a result of collective behaviors based upon a set of similar movement principles and constructs connections to its surroundings within the system in a long run [44]. The sustainability of a system relies greatly on such biophilic structures and configurations. Therefore, we believe that the geographic space and human activities therein are self-organized and need to be better studied through the reference of a biological entity or system in terms of its form and function for the sake of making sustainable cities and societies.

5. Conclusions

The complexity of geographic space is not only manifested as fractal shapes in the physical space such as rivers, mountains, and coastlines [45], but also reflected by the nonlinear dynamics of human

activities in the virtual space such as LBSM, the Internet, and the World Wide Web. The present paper has explored the complexity of spatial interactions between different administration divisions in China through the lens of human activities in the LBSM platform Flickr. We extracted each user's trajectories and used them to build up the networks to explore the countrywide spatial interactions at multiple spatio-temporal resolutions. We found that the interactions had a striking spatial-temporal heterogeneity, supported by power law fitting metrics and ht-index values regarding both node degrees and edge weights. The networks at multiple scales hold similar power law exponents with accepted goodness of fit index values, which enables us to characterize the fractal structure of spatial interactions. Further research could consider other social media data sources, such as Weibo, for a more comprehensive study of movement dynamics in China. In the meantime, we could extend this methodological framework using finer administration units, such as at the town or village level.

References

1. Shaw, S.-L.; Yu, H. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *J. Transp. Geogr.* **2009**, *17*, 141–149. [CrossRef]
2. Li, S.; Dragićević, S.; Castro, F.A.; Sester, M.; Winter, S.; Çöltekin, A.; Pettit, C.; Jiang, B.; Haworth, J.; Stein, A.; et al. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 119–133. [CrossRef]
3. Salat, S.; Bourdic, L. Urban complexity, scale hierarchy, energy efficiency and economic value creation. In *The Sustainable City VII: Urban Regeneration and Sustainability*; Pacetti, M., Passerini, G., Brebbia, C.A., Latini, G., Eds.; WIT Press: Boston, MA, USA, 2012; Volume 155, pp. 97–107.
4. Shaw, S.-L.; Tsou, M.-H.; Ye, X. Editorial: Human dynamics in the mobile and big data era. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1687–1693. [CrossRef]
5. Barabási, A.-L. The origin of bursts and heavy tails in human dynamics. *Nat. Cell Biol.* **2005**, *435*, 207–211. [CrossRef]
6. Brockmann, D.; A Hufnagel, L.; Geisel, T. The scaling laws of human travel. *Nat. Cell Biol.* **2006**, *439*, 462–465. [CrossRef]
7. González, M.C.; Hidalgo, C.A.; Barabási, A.-L. Understanding individual human mobility patterns. *Nat. Cell Biol.* **2008**, *453*, 779–782. [CrossRef]
8. Kulshrestha, J.; Kooti, F.; Nikravesh, A.; Gummadi, P.K. Geographic Dissection of the Twitter Network. In Proceedings of the 6th International AAAI Conference on Web and Social Media (ICWSM), Dublin, Ireland, 4–7 June 2012; pp. 202–209.
9. Takhteyev, Y.; Gruzd, A.A.; Wellman, B. Geography of Twitter networks. *Soc. Netw.* **2012**, *34*, 73–81. [CrossRef]
10. Wu, L.; Zhi, Y.; Sui, Z.; Liu, Y. Intra-Urban Human Mobility and Activity Transition: Evidence from Social Media Check-In Data. *PLoS ONE* **2014**, *9*, e97010. [CrossRef]
11. Ma, D.; Sandberg, M.; Jiang, B. A Socio-Geographic Perspective on Human Activities in Social Media. *Geogr. Anal.* **2017**, *49*, 328–342. [CrossRef]
12. de Bruijn, J.A.; de Moel, H.; Jongman, B.; Wagemaker, J.; Aerts, J.C. TAGGS: Grouping tweets to improve global geoparsing for disaster response. *J. Geovisualization Spat. Anal.* **2018**, *2*, 2. [CrossRef]
13. Newman, M.E.J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **2005**, *46*, 323–351. [CrossRef]

14. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-Law Distributions in Empirical Data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]

15. Zhong, C.; Arisona, S.M.; Huang, X.; Batty, M.; Schmitt, G. Detecting the dynamics of urban structure through spatial network analysis. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2178–2199. [CrossRef]

16. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [CrossRef]

17. Newman, M.E.J. The Structure and Function of Complex Networks. *SIAM Rev.* **2003**, *45*, 167–256. [CrossRef]

18. Watts, D.J.; Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nat. Cell Biol.* **1998**, *393*, 440–442. [CrossRef]

19. Barabasiand, A.-L.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509–512. [CrossRef]

20. National Resources and Environment Database of the Chinese Academy of Sciences. Available online: http://www.resdc.cn/ (accessed on 11 August 2020).

21. Chen, D.; Zhang, Y.; Yao, Y.; Hong, Y.; Guan, Q.; Tu, W. Exploring the spatial differentiation of urbanization on two sides of the Hu Huanyong Line—Based on nighttime light data and cellular automata. *Appl. Geogr.* **2019**, *112*, 102081. [CrossRef]

22. Yahoo Flickr Creative Commons 100 Million (YFCC100m) Dataset. Available online: http://projects.dfki.uni-kl.de/yfcc100m/ (accessed on 15 July 2020).

23. Bertin, J. *Semiology of Graphics: Diagrams, Networks, Maps*; University of Wisconsin Press: Madison, WI, USA, 1983.

24. Fruchterman, T.M.J.; Reingold, E.M. Graph drawing by force-directed placement. *Softw. Pr. Exp.* **1991**, *21*, 1129–1164. [CrossRef]

25. Gastner, M.T.; Newman, M.E.J. From The Cover: Diffusion-based method for producing density-equalizing maps. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7499–7504. [CrossRef]

26. Hanel, R.; Corominas-Murtra, B.; Liu, B.; Thurner, S. Fitting power-laws in empirical data with estimators that work for all exponents. *PLoS ONE* **2017**, *12*, e0170920. [CrossRef] [PubMed]

27. Jiang, B. Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution. *Prof. Geogr.* **2013**, *65*, 482–494. [CrossRef]

28. Jiang, B.; Yin, J. Ht-Index for Quantifying the Fractal or Scaling Structure of Geographic Features. *Ann. Assoc. Am. Geogr.* **2014**, *104*, 530–540. [CrossRef]

29. Mayer-Schonberger, V.; Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*; Eamon Dolan/Houghton Mifflin Harcourt: New York, NY, USA, 2013.

30. Anderson, G.; Ge, Y. The size distribution of Chinese cities. *Reg. Sci. Urban Econ.* **2005**, *35*, 756–776. [CrossRef]

31. Berliant, M.; Watanabe, H. Explaining the size distribution of cities: Extreme economies. *Quant. Econ.* **2015**, *6*, 153–187. [CrossRef]

32. Gabaix, X. Power Laws in Economics and Finance. *Annu. Rev. Econ.* **2009**, *1*, 255–294. [CrossRef]

33. Berliant, M.; Watanabe, A.H. A scale-free transportation network explains the city-size distribution. *Quant. Econ.* **2018**, *9*, 1419–1451. [CrossRef]

34. Jackson, M.O.; Wolinskyb, A. A Strategic Model of Social and Economic Networks. *J. Econ. Theory* **1996**, *71*, 44–74. [CrossRef]

35. Jackson, M.O.; Watts, A. On the formation of interaction networks in social coordination games. *Games Econ. Behav.* **2002**, *41*, 265–291. [CrossRef]

36. Jackson, M.O.; Watts, A. The Evolution of Social and Economic Networks. *J. Econ. Theory* **2002**, *106*, 265–295. [CrossRef]

37. Jackson, M.O.; Yves, Z. *Economic Analyses of Social Networks*; Edward Elgar Publishing: Cheltenham, UK, 2013.

38. Zipf, G.K. *Human Behavior and the Principles of Least Effort*; Addison Wesley: Cambridge, UK, 1949.

39. Giesen, K.; Südekum, J. Zipf's law for cities in the regions and the country. *J. Econ. Geogr.* **2011**, *11*, 667–686. [CrossRef]

40. Mitzenmacher, M. A brief history of generative models for power law and lognormal distributions. *Internet Math.* **2004**, *1*, 226–251. [CrossRef]

41. Gabaix, X. Zipf's Law for Cities: An Explanation. *Q. J. Econ.* **1999**, *114*, 739–767. [CrossRef]

42. Zheng, M.; Allard, A.; Hagmann, P.; Alemán-Gómez, Y.; Serrano, M. Ángeles Geometric renormalization unravels self-similarity of the multiscale human connectome. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 20244–20253. [CrossRef] [PubMed]

43. García-Pérez, G.; Boguñá, M.; Serrano, M. Ángeles Multiscale unfolding of real networks by geometric renormalization. *Nat. Phys.* **2018**, *14*, 583–589. [CrossRef]

44. Heppenstall, A.; Crooks, A.; See, L.M.; Batty, M. *Agent-Based Modelling of Geographical Systems*; Springer: Dordrecht, The Netherlands, 2012.

45. Mandelbrot, B. How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Science* **1967**, *156*, 636–638. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.