# Constructing multi-level urban clusters based on population distributions and interactions

Wenpu Cao,<sup>1</sup> Lei Dong,<sup>1,\*</sup> Ying Cheng,<sup>2</sup> Lun Wu,<sup>1</sup> Qinghua Guo,<sup>1</sup> and Yu Liu<sup>1</sup>

<sup>1</sup>Institute of Remote Sensing and Geographic Information Systems,

School of Earth and Space Sciences, Peking University, Beijing 100871, China

<sup>2</sup> State Key Laboratory of Media Convergence Production Technology and Systems, Beijing, China

A city (or an urban cluster) is not an isolated spatial unit, but a combination of areas with closely linked socio-economic activities. However, so far, we lack a consistent and quantitative approach to define multi-level urban clusters through these socio-economic connections. Here, using granular population distribution and flow data from China, we propose a bottom-up aggregation approach to quantify urban clusters at multiple spatial scales. We reveal six 'phases' (i.e., levels) in the population density-flow diagram, each of which corresponds to a spatial configuration of urban clusters from large to small. Besides, our results show that Zipf's law appears only after the fifth level, confirming the spatially dependent nature of urban laws. Our approach does not need pre-defined administrative boundaries and can be applied effectively on a global scale.

Keywords: urban cluster, population distribution, commuting flow, multi-level, urban system, big data

## I. INTRODUCTION

A city (or an urban cluster) is an integrated area with a dense population and intensive internal socio-economic connections. It includes not only the core urban area but also the surrounding areas that interact closely with the core [\[9,](#page-7-0) [19\]](#page-7-1). However, in many countries, administrative divisions are still the most common way of delineating urban areas, which poses many urban planning and transportation problems [\[36\]](#page-7-2). For example, in China, many people work in Beijing but live in Sanhe (a small neighboring county close to Beijing) due to the low cost of living and relatively convenient transportation [\[5\]](#page-6-0). Those people who commute between Sanhe and Beijing contribute to Beijing's economic growth and have demands for Beijing's public services, but they are not part of Beijing's resident population according to the current definition of the city's boundaries.

Realizing that the administrative boundaries of cities are problematic, many studies turn to delineating urban areas through data-driven ways. These efforts include the identification of contiguous urban areas with a high density of human activity, i.e., density-based methods [\[4,](#page-6-1) [31,](#page-7-3) [41,](#page-8-0) [42,](#page-8-1) [47\]](#page-8-2). For example, the threshold methods define areas that exceed a certain density of population, infrastructure (e.g., road networks), or socioeconomic activity (e.g., nighttime light) as urban areas [\[6,](#page-7-4) [13,](#page-7-5) [33,](#page-7-6) [41\]](#page-8-0). Some machine learning methods extract impervious surfaces or high-density built-up areas as urban areas from remote sensing data [\[27,](#page-7-7) [43\]](#page-8-3).

With the rapid development of transportation and communication technology, the flows of people, goods, and information between areas are becoming increasingly frequent. From a flow perspective, a city can be defined in terms of its core urban areas and surrounding areas with intensive socio-economic interactions, known as functional urban areas (FUAs) [\[15,](#page-7-8) [21,](#page-7-9) [29\]](#page-7-10). Some developed countries have official definitions for FUAs, such as the metropolitan statistical areas (MSAs) in the United States, which encompass areas with high-density populations and the surrounding commuting zones [\[9\]](#page-7-0). Some recent studies have used similar flow-based methods to quantify urban areas in the UK [\[3\]](#page-6-2) and France [\[17,](#page-7-11) [18\]](#page-7-12). However, most previous works to identify FUAs need detailed census data or commuting survey data, which are difficult to obtain in developing countries. Meanwhile, the choice of population and commuting thresholds is arbitrary in most methods, making it difficult to compare urban clusters across different countries and periods [\[46\]](#page-8-4).

Another piece of the puzzle that has been overlooked by previous flow-based methods is the multi-level structure of urban systems. For example, the flow-based method by the OECD defines urban clusters on only one spatial level [\[21\]](#page-7-9), whereas urban systems have multiple levels, according to previous studies [\[1,](#page-6-3) [4,](#page-6-1) [8\]](#page-7-13). Although several data-driven approaches are proposed to derive urban hierarchies based on percolation theory [\[4\]](#page-6-1) or community detection [\[30,](#page-7-14) [32\]](#page-7-15), robust methods to delineate multi-scale urban clusters and reveal the different levels of urban systems require further research.

With the development of information and communication technology, mobile phone data have shown great potential in mapping dynamic population distributions [\[14,](#page-7-16) [20\]](#page-7-17) and flow interactions [\[28,](#page-7-18) [35\]](#page-7-19). These advances in mobile phone data allow us to quantify multi-scale urban clusters in terms of both population distributions and flow interactions [\[15,](#page-7-8) [20,](#page-7-17) [28\]](#page-7-18). In addition, city science has shown that urban systems are typically complex systems and obey some universal patterns, such as Zipf's law [\[26,](#page-7-20) [48\]](#page-8-5) and allometric growth [\[12,](#page-7-21) [39\]](#page-8-6), which guide us to study effective multi-level urban clusters.

In this paper, we propose a bottom-up approach to quantify multi-level urban clusters based on the population distributions and flows inferred from mobile phone data. We first traverse all population density thresholds

<span id="page-0-0"></span><sup>∗</sup> [Corresponding author: leidong@pku.edu.cn](mailto:Corresponding author: leidong@pku.edu.cn)

and aggregate spatially contiguous populated units into separate clusters under each threshold. This process is similar to previous density-based methods [\[41,](#page-8-0) [42\]](#page-8-1), and we refer to these individual clusters as population clusters. We then link population clusters to form urban clusters based on different flow ratio thresholds. Through percolation analysis, we obtain six 'phases' in the population density-flow phase diagram, each of which corresponds to one level of urban systems. Results show that Zipf's law only appears after the fifth level. Our research proposes an effective method to characterize the population distributions and interactions of cities and to quantitatively understand urban clusters on different levels.

### II. DATA

The population distribution and flow data that we used in this work are derived from an anonymous mobile phone dataset provided by a large location service provider from China in 2015 [\[13,](#page-7-5) [23,](#page-7-22) [24\]](#page-7-23). Individual's home and work locations are detected by a machine learning method from the user's stay points; see Dong et al. [\[24\]](#page-7-23) for details. The population distribution is calculated based on the number of users whose homes are located within each  $0.5' \times 0.5'$  (approximately 1km at the equator) grid. The commuting flow data aggregate the number of home-work connections between any two  $0.5' \times 0.5'$  grids.

Different location service providers have different market shares in different regions. Therefore, we need to assess whether the population data derived from one single service provider can represent the real population dynamics in China. To assess the accuracy of the population distribution inferred from the mobile phone data, we compare the estimated data with the micro-census data of 2015 at the prefecture level. The correlation coefficient is 0.9, indicating that the estimated population distribution is consistent with the census. Note that the mobile phone dataset covers approximately 100 million people out of a total population of 1.4 billion in China. Thus, all reported population-related values are scaled by 13.9 (total population number / number of mobile phone users).

### III. METHODS

To construct multi-level urban clusters, we take the following steps (Fig. [1\)](#page-2-0). First, we merge the spatially adjacent populated grids into clusters at different population density thresholds. We refer to these clusters as population clusters. Second, we merge these population clusters into urban clusters by different flow ratio thresholds. Similar to the concept of FUAs, an urban cluster is a combination of populated areas with strong commuting connections. Finally, we use the percolation theory to analyze the urban clusters under different population density and flow ratio thresholds with the density-flow phase diagram.

## A. Clustering by population density thresholds

For a given population density threshold  $D_*$ , we label the cells with values greater than  $D_*$  as populated cells and then aggregate geographically contiguous populated cells into population clusters using the City Clustering Algorithm [\[41,](#page-8-0) [42\]](#page-8-1). Specifically, we start with any unprocessed populated cell and iteratively add its neighboring populated cells (eight nearest neighbors) until the neighbors of all cells in the cluster are either non-populated cells or within the cluster. Then we aggregate another population cluster until all populated cells belong to a specific cluster. The area and population of each population cluster is the sum of the areas and populations of all cells within that population cluster. In addition, each population cluster should meet certain area and population size criteria. Here, we set the minimum area  $A_*$ of the population cluster to  $10km^2$  (for reference, the land area of Sansha, the smallest city in China, is about  $20km^2$ ). We also set the minimum area to 5, 15, 20, and  $25 \; km^2$ , and the results are robust (Table A.1); see Appendix for details.

### B. Clustering by flow ratio thresholds

We group the detected population clusters into urban clusters according to different flow ratio thresholds. For given population clusters  $i$  and  $j$ , we measure the commuting flow  $P_{ij}$  whose origin is located in cluster i and destination is located in cluster  $j$ . The flow is then normalized as follows:  $F_{ij} = P_{ij} / \sum_{j=0}^{N} P_{ij}$ . For a given flow ratio threshold  $F_*$  and any two population clusters, a link is established between the two clusters if the commuting ratio in either direction exceeds the threshold  $F_*$ . Besides, if a population cluster has multiple commuting flows to other population clusters that exceed the threshold  $F_*$ , the cluster is linked only to the cluster with the largest flow ratio. Finally, the linked population clusters are merged together to construct an urban cluster. This process is applied only once for all population clusters in parallel, and the merged clusters will not be iterated again. The area and population of each urban cluster is the sum of the areas and populations of all population clusters within that urban cluster.

### C. Percolation analysis on urban clusters

We analyze the urban clusters under different densityflow thresholds using the percolation theory. The percolation theory studies the transition of many small clusters merged into a large spanning cluster at the critical point [\[16\]](#page-7-24). Around the critical point, cluster sys-



<span id="page-2-0"></span>FIG. 1. The method. a) Spatially contiguous cells with values larger than the population density threshold are aggregated into population clusters. b) Population clusters with flow ratios greater than the flow ratio threshold are merged into urban clusters. In panel (b), each node corresponds to a population cluster, and each edge corresponds to the flow ratio from one population cluster to another. Line widths are plotted according to the flow ratio. Solid lines show the edges with weights exceeding the threshold, while dotted lines show the edges with weights less than the threshold.

tems exhibit power-law critical phenomena that also exist in urban systems (e.g., the size distribution of cities). Therefore, the percolation theory has been widely used to model urban growth [\[37,](#page-8-7) [38\]](#page-8-8) and quantify urban clusters [\[4,](#page-6-1) [13\]](#page-7-5). Here, we iterate over all possible population density thresholds (from non-zero to very dense) and obtain the population clusters under each density threshold. Then, we traverse the flow ratio thresholds from 0% to 100% and merge the population clusters into urban clusters under each flow threshold, where 0 means that any two population clusters with interactions are merged together, and 100% means that no population clusters are merged since the flow ratio of two clusters cannot exceed 100%. Finally, we obtain many collections of urban clusters, each of which corresponds to a specific population density threshold and a specific flow ratio threshold. Similar to the percolation model, we calculate the area of the largest urban cluster (ranked by land area) as a function of the density and flow thresholds and construct the density-flow phase diagram.

## IV. RESULTS

### A. Density-flow phase diagram

We rank urban clusters under each density-flow threshold by land area and obtain the phase diagram (Fig. [2a](#page-3-0)). The color of each grid cell in Fig. [2a](#page-3-0) indicates the area of the largest urban cluster under the specified population density-flow thresholds. From bottom left to top right, urban clusters gradually transition from sparsely

populated, weakly connected urban areas to densely populated, strongly connected urban areas as the thresholds increase. Remarkably, by transforming Fig. [2a](#page-3-0) into a histogram of the largest cluster areas under the different density-flow thresholds, we observe 6 groups (Fig. [2b](#page-3-0)). Using the breakpoints (vertical dashed lines) in Fig. [2b](#page-3-0), we further divide the density-flow phase diagram into 6 phases (solid lines in Fig. [2a](#page-3-0)), with the cluster systems remaining stable within each phase.

As the population density threshold and the flow ratio threshold increase, the largest urban cluster in the system remains stable until the urban system transitions to a lower-level spatial scale. Thus, we regard 6 phases as 6 spatial levels of one urban system. Table [I](#page-3-1) shows the population density threshold  $D_*$  (people $/km^2$ ), the flow ratio threshold  $F_*$  (%), the number of urban clusters  $N_{urban}$ , the total area  $A_{total}$  (km<sup>2</sup>) and the total population  $P_{total}$  (people) of the urban clusters, and the area  $A_{largest}$  (km<sup>2</sup>) and the population  $P_{largest}$  (people) of the largest cluster in each level from Level 1 to Level 6.

### B. Urban clusters at different levels

We next investigate the spatial distributions of the urban clusters at different levels. Fig. [3](#page-4-0) shows the urban clusters at Level 1 and Level 2, and Fig. [4](#page-4-1) shows the urban clusters from Level 3 to Level 6 in Beijing-Tianjin-Hebei, the Yangtze River Delta, and the Pearl River Delta (mapped with black lines in Fig. [3b](#page-4-0)).

Level 1 corresponds to the non-zero population density-flow thresholds; that is, areas with a population



<span id="page-3-0"></span>FIG. 2. a) The density-flow phase diagram. Each cell in the diagram indicates the size of the largest urban cluster at the corresponding density and flow thresholds. Solid lines indicate the boundaries of each phase. b) Histogram of the largest cluster areas of all urban systems. We observe clear breakpoints that are demarcated by dotted lines.

<span id="page-3-1"></span>



>0 are marked as populated areas, and any two clusters with flows  $>0$  are linked together. At Level 1, the areas of the top 5 largest clusters account for 87% of the total area of all clusters. The largest urban cluster has an area of  $6.60 \times 10^5 km^2$  and a population of  $7.65 \times 10^8$ , including most of Central and East China and parts of North and Northeast China. The 2nd and 5th largest urban clusters cover South China. The 3rd and 4th largest urban clusters correspond to the major urban areas in Southwest and Northwest China, respectively.

Level 2 corresponds to the thresholds of 140 people per  $km^2$  and a 0% flow ratio. Level-2 clusters represent the development poles in each region. The largest urban cluster covers most of Hebei Province, Henan Province, and Shandong Province, and has an area of  $1.45 \times 10^5 km^2$ and a population of  $3.19 \times 10^8$ . Beijing, the capital of China, is within the cluster. The 2nd largest urban cluster covers Jiangsu Province, Anhui Province and Zhejiang Province and includes the Yangtze River Delta; the 3rd cluster covers Guangdong Province and includes the Pearl River Delta; the 4th and 5th clusters correspond to Hunan Province and Sichuan Province, respectively (Fig.

[3b](#page-4-0)).

Level 3 corresponds to the thresholds of 280 people per  $km^2$  and a 2% flow ratio, and Level 4 corresponds to the thresholds of 700 people per  $km^2$  and a 2% flow ratio. The urban clusters at these two levels are consistent with the concept of urban agglomeration, which is the spatial form of several highly integrated cities [\[25\]](#page-7-25). The largest urban cluster at Level 3 is located in the Yangtze River Delta, including Shanghai and major cities in southern Jiangsu Province and northern Zhejiang Province. At Level 4, Hangzhou and its surrounding cities are separated from this cluster. The largest urban cluster at Level 4 is located in the Pearl River Delta, corresponding to the Greater Bay Area, and mainly includes Guangzhou and Shenzhen (both cities with a population of more than 10 million). The Yangtze River Delta and the Pearl River Delta are all well-developed regions in China.

Level 5 corresponds to the thresholds of 2, 940 people per  $km^2$  and a 5% flow ratio. At Level 5, each urban cluster has a dense core urban area, as well as surrounding areas that are closely connected to the core, which is similar to the definition of an FUA. However,



FIG. 3. Urban clusters at Level 1 (a) and Level 2 (b). The top 5 largest clusters are colored by rank. The black lines show the area of Beijing-Tianjin-Hebei, the Yangtze River Delta, and the Pearl River Delta.

<span id="page-4-0"></span>

<span id="page-4-1"></span>FIG. 4. a-d) Urban clusters in Beijing-Tianjin-Hebei, the Yangtze River Delta, and the Pearl River Delta at Levels 3 (a), 4 (b), 5 (c), and 6 (d), respectively. The top 5 largest clusters in each region are colored by rank.

instead of using pre-defined parameters (population density and flow ratio thresholds), our method obtains the thresholds objectively through percolation. The number of Level-5 urban clusters (349) is similar to the number of prefecture-level cities (333) in China. As shown in Fig. [4,](#page-4-1) one prefecture-level city corresponds to one urban cluster in most regions. Yet, in some well-developed regions, intensive socio-economic connections make urban clusters cross administrative boundaries. For example, Sanhe (the red circle in Fig. [4c](#page-4-1)) and Beijing are merged together due to the large number of commuting, although Beijing and Sanhe are within different administrative city boundaries.

Level 6 corresponds to the population density threshold of 4,620 people per  $km^2$  and the flow ratio threshold of 28%. Due to the high flow ratio threshold, most urban clusters in this level consist of one single population cluster. Compared to the administrative boundaries, the obtained urban clusters mainly correspond to core urban areas within prefecture-level cities. There are also some urban clusters that cross the boundaries of prefecturelevel cities. For example, the central bodies of western Shenzhen and Dongguan are geographically continuous with high population density, forming an urban cluster across the boundaries, while eastern Shenzhen is not adjacent to this urban cluster due to geographical barriers (e.g., mountains and forests), and the flow ratio is also not high enough, forming another urban cluster. As a validation, we compare the largest urban cluster (Beijing) detected by our method to the official released built-up area data in 2015. Our results show that Beijing at Level 6 has an area of  $1,381km^2$ , while the official number is  $1,401km^2$ , and the two values are very close.

We also perform a sensitivity test by applying our method to Guangdong Province in China. Although fewer levels are derived, the density-flow phase diagram and the derived urban clusters at different levels are similar; see Fig. A.1-A.2 in the Appendix.

### C. Zipf's law and allometric growth

To investigate the relationship between area and population in the urban clusters, we test Zipf's law and allometric growth. Zipf's law and allometric growth reflect the self-organized properties of urban complex systems and have been widely tested across different countries in different periods [\[11,](#page-7-26) [12,](#page-7-21) [44\]](#page-8-9). Zipf's law states that when cities are ranked by population size, the population size S and the rank R of cities satisfy  $S = 1/R$ . This can also be expressed as the probability density function (PDF) of urban cluster sizes following a power-law distribution where  $P(S) \sim S^{-\alpha}$  and  $\alpha = 2$  [\[26,](#page-7-20) [48\]](#page-8-5). Allometric growth refers to the sub-linear power-law relationship between the population  $P$  and the urban area  $A$ , which can be expressed as  $A = P^{\beta}$  with  $\beta \approx 0.6 - 0.8$  [\[11,](#page-7-26) [39\]](#page-8-6). Allometric growth represents economies of scale, meaning that larger urban clusters can result in more efficient land

use per capita, since the population is more concentrated [\[12\]](#page-7-21).

Figure [5a](#page-6-4)-b shows that the population sizes of the urban clusters from Level 1 to Level 6 all follow power-law distributions. We further use the method proposed by Alstott et al. [\[2\]](#page-6-5) to fit the data, and the power-law exponents  $\alpha$  for Level 5 and Level 6 are around 2, which means that Zipf's law holds well at these two levels. However, the exponents  $\alpha$  for Levels 1-4 deviate from 2, indicating that Zipf's law does not hold at higher levels. It should be noted that there is a long-standing debate on whether Zipf's law holds for urban clusters at different scales [\[10\]](#page-7-27), and Zipf's law is not a validation of the derived urban clusters. As a complement, we demonstrate that Zipf's law does not hold on all spatial scales; it only holds after Level 5. Interestingly, as shown in the previous section, the Level-5 urban clusters have similar spatial scales to the prefecture-level cities, indicating that our results are consistent with previous studies on the empirical test of Zipf's law [\[26,](#page-7-20) [44\]](#page-8-9).

Figure [5c](#page-6-4)-d shows the power-law relationship between the populations and the areas of the urban clusters at different levels. The exponents  $\beta$  are all  $\lt 1$ , indicating the existence of allometric growth. However, similar to Zipf's law, only after Level 5, the exponents begin to approach the previous theoretical and empirical values [\[11\]](#page-7-26). This finding also adds evidence that the allometric growth depends on the definition of urban clusters [\[10,](#page-7-27) [40\]](#page-8-10).

#### V. DISCUSSION

We propose a bottom-up approach to construct multilevel urban clusters using population distribution and flow data. The density-flow thresholds are derived according to the critical characteristics of urban systems based on the percolation theory. We reveal six levels of urban clusters from the density-flow phase diagram, and show that Zipf's law holds only for urban clusters after Level 5, providing empirical evidence for the spatially dependent nature of urban laws. Regarding applications, our method does not rely on any pre-defined geographical units and requires only the density-flow thresholds obtained by percolation, allowing us to efficiently derive multi-level urban clusters.

The prevalence of mobile phones allows us to obtain population distribution and flow data at a lower cost and faster update rate. Thus, our approach can be easily applied to countries that have mobile phone data but do not have recent census/survey data (e.g., some developing countries). Even for those regions where mobile phone data are unavailable, we could also consider using human mobility models to infer population flows from population distributions [\[7\]](#page-7-28), which are usually easy to obtain, such as the WorldPop dataset [\[45\]](#page-8-11).

Some extensions can be explored in further studies. First, in this paper, we use commuting flows to mea-



<span id="page-6-4"></span>FIG. 5. a-b) The PDFs of the population sizes of the urban clusters from Level 1 to Level 6. Solid lines represent the power-law fitting results. c-d) The sub-linear relationship between the populations and the areas of the urban clusters from Level 1 to Level 6. Solid lines represent the fitting results.  $\pm$  represents one standard deviation.

sure the socio-economic connections between areas. However, commuting accounts for only approximately 1/3 of all human movements [\[22\]](#page-7-29), and a large number of noncommuting movements also reflect the socio-economic connections. The impact of different types of flow interactions on quantifying urban clusters is worth exploring. Second, the movements of individuals vary substantially across different periods. For example, resident commuting dropped significantly during the COVID-19 pandemic [\[34\]](#page-7-30). This might lead to changes in the boundaries of

- <span id="page-6-3"></span>[1] Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. Nature, 587(7834):402–407, November 2020.
- <span id="page-6-5"></span>[2] Jeff Alstott, Ed Bullmore, and Dietmar Plenz. powerlaw: a Python package for analysis of heavy-tailed distributions. PLoS ONE, 9(1):e85777, January 2014.
- <span id="page-6-2"></span>[3] Elsa Arcaute, Erez Hatna, Peter Ferguson, Hyejin Youn, Anders Johansson, and Michael Batty. Constructing cities, deconstructing scaling laws. Journal of The Royal Society Interface, 12(102):20140745, January 2015.

flow-based urban clusters, and these temporally dynamic boundaries are worth exploring in our future work.

## VI. DATA AND CODES AVAILABILITY

The dataset and codes to reproduce the results of this paper are available via [https://github.com/](https://github.com/caowenpu56/densityflow) [caowenpu56/densityflow](https://github.com/caowenpu56/densityflow).

- <span id="page-6-1"></span>[4] Elsa Arcaute, Carlos Molinero, Erez Hatna, Roberto Murcio, Camilo Vargas-Ruiz, A. Paolo Masucci, and Michael Batty. Cities and regions in Britain through hierarchical percolation. Royal Society Open Science, 3(4): 150691, April 2016.
- <span id="page-6-0"></span>[5] Guang Bai and Chaohe Rong. From the view of field theory, the basic role of cross urban rail commuting service in metropolitan area — A case study of Beijing, Sanhe and Yanjiao. In 2016 International Conference on Industrial Economics System and Industrial Security Engineering (IEIS), pages 1–4, Sydney, Australia, July 2016. IEEE.
- <span id="page-7-4"></span>[6] Kathryn Baragwanath, Ran Goldblatt, Gordon Hanson, and Amit K. Khandelwal. Detecting urban markets with satellite imagery: An application to India. Journal of Urban Economics, 125:103173, September 2021.
- <span id="page-7-28"></span>[7] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R. James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J. Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. Physics Reports, 734(6):1–74, March 2018.
- <span id="page-7-13"></span>[8] Michael Batty. The New Science of Cities. The MIT Press, 2013. ISBN 978-0-262-31823-5.
- <span id="page-7-0"></span>[9] B.J.L. Berry, P.G. Goheen, and H. Goldstein. Metropolitan Area Definition: A Re-evaluation of Concept and Statistical Practice, volume 28. Washington, D.C.: U.S. Bureau of the Census, 1969.
- <span id="page-7-27"></span>[10] Brian J.L. Berry and Adam Okulicz-Kozaryn. The city size distribution debate: Resolution for US urban regions and megalopolitan areas. Cities, 29:S17–S23, March 2012.
- <span id="page-7-26"></span>[11] Luís M. A. Bettencourt. The origins of scaling in cities. Science, 340(6139):1438–1441, June 2013.
- <span id="page-7-21"></span>[12] Luís M. A. Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. Growth, innovation, scaling, and the pace of life in cities. Proceedings of the National Academy of Sciences, 104(17):7301–7306, April 2007.
- <span id="page-7-5"></span>[13] Wenpu Cao, Lei Dong, Lun Wu, and Yu Liu. Quantifying urban areas with multi-source data based on percolation theory. Remote Sensing of Environment, 241: 111730, May 2020.
- <span id="page-7-16"></span>[14] Jie Chen, Tao Pei, Shih-Lung Shaw, Feng Lu, Mingxiao Li, Shifen Cheng, Xiliang Liu, and Hengcai Zhang. Fine-grained prediction of urban population using mobile phone location data. International Journal of Geographical Information Science, 32(9):1770–1786, September 2018.
- <span id="page-7-8"></span>[15] Zifeng Chen and Anthony Gar-On Yeh. Delineating functional urban areas in Chinese mega city regions using fine-grained population data and cellphone location data: A case of Pearl River Delta. Computers, Environment and Urban Systems, 93:101771, April 2022.
- <span id="page-7-24"></span>[16] Kim Christensen and Nicholas R. Moloney. Complexity and Criticality, volume 1. Imperial College Press, 2005.
- <span id="page-7-11"></span>[17] Clémentine Cottineau, Erez Hatna, Elsa Arcaute, and Michael Batty. Diverse cities or the systematic paradox of urban scaling laws. Computers, Environment and Urban Systems, 63:80–94, May 2017.
- <span id="page-7-12"></span>[18] Clémentine Cottineau, Olivier Finance, Erez Hatna, Elsa Arcaute, and Michael Batty. Defining urban clusters to detect agglomeration economies. Environment and Planning B: Urban Analytics and City Science, 46(9):1611– 1626, November 2019.
- <span id="page-7-1"></span>[19] Garrett Dash Nelson and Alasdair Rae. An economic geography of the United States: From commutes to megaregions. PLOS ONE, 11(11):e0166083, November 2016.
- <span id="page-7-17"></span>[20] Pierre Deville, Catherine Linard, Samuel Martin, Marius Gilbert, Forrest R. Stevens, Andrea E. Gaughan, Vincent D. Blondel, and Andrew J. Tatem. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111(45):15888–15893, November 2014.
- <span id="page-7-9"></span>[21] L. Dijkstra, H. Poelman, and P. Veneri. The EU-OECD definition of a functional urban area. OECD Regional

Development Working Papers, December 2019.

- <span id="page-7-29"></span>[22] Chuan Ding, Chao Liu, Yi Zhang, Jiawen Yang, and Yunpeng Wang. Investigating the impacts of built environment on vehicle miles traveled and energy consumption: Differences between commuting and noncommuting trips. Cities, 68:25–36, August 2017.
- <span id="page-7-22"></span>[23] Lei Dong, Sicong Chen, Yunsheng Cheng, Zhengwei Wu, Chao Li, and Haishan Wu. Measuring economic activity in China with mobile big data. EPJ Data Science, 6(1): 29, December 2017.
- <span id="page-7-23"></span>[24] Lei Dong, Paolo Santi, Yu Liu, Siqi Zheng, and Carlo Ratti. The universality in urban commuting across and within cities, 2022. URL [https://arxiv.org/abs/2204.](https://arxiv.org/abs/2204.12865) [12865](https://arxiv.org/abs/2204.12865).
- <span id="page-7-25"></span>[25] Chuanglin Fang and Danlin Yu. Urban agglomeration: An evolving concept of an emerging phenomenon. Landscape and Urban Planning, 162:126–136, June 2017.
- <span id="page-7-20"></span>[26] X. Gabaix. Zipf's law for cities: An explanation. The Quarterly Journal of Economics, 114(3):739–767, August 1999.
- <span id="page-7-7"></span>[27] Virgilio Galdo, Yue Li, and Martin Rama. Identifying urban areas by combining human judgment and machine learning: An application to India. Journal of Urban Economics, 125:103229, September 2021.
- <span id="page-7-18"></span>[28] Amnir Hadachi, Mozhgan Pourmoradnasseri, and Kaveh Khoshkhah. Unveiling large-scale commuting patterns based on mobile phone cellular network data. Journal of Transport Geography, 89:102871, December 2020.
- <span id="page-7-10"></span>[29] Peter Hall. Looking backward, looking forward: The city region of the mid-21st century. Regional Studies, 43(6): 803–817, July 2009.
- <span id="page-7-14"></span>[30] Ye Hong and Yao Yao. Hierarchical community detection and functional area identification with OSM roads and complex graph theory. International Journal of Geographical Information Science, 33(8):1569–1587, August 2019.
- <span id="page-7-3"></span>[31] Marc L. Imhoff, William T. Lawrence, David C. Stutzer, and Christopher D. Elvidge. A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. Remote Sensing of Environment, 61(3):361– 370, September 1997.
- <span id="page-7-15"></span>[32] Tao Jia, Xi Luo, and Xin Li. Delineating a hierarchical organization of ranked urban clusters using a spatial interaction network. Computers, Environment and Urban Systems, 87:101617, May 2021.
- <span id="page-7-6"></span>[33] Bin Jiang and Tao Jia. Zipf's law for all the natural cities in the United States: A geospatial perspective. International Journal of Geographical Information Science, 25 (8):1269–1281, August 2011.
- <span id="page-7-30"></span>[34] Moritz U. G. Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M. Pigott, Open COVID-19 Data Working Group†, Louis du Plessis, Nuno R. Faria, Ruoran Li, William P. Hanage, John S. Brownstein, Maylis Layan, Alessandro Vespignani, Huaiyu Tian, Christopher Dye, Oliver G. Pybus, and Samuel V. Scarpino. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science, 368(6490):493–497, May 2020.
- <span id="page-7-19"></span>[35] Kevin S. Kung, Kael Greco, Stanislav Sobolevsky, and Carlo Ratti. Exploring universal patterns in human home-work commuting from mobile phone data. PLoS ONE, 9(6):e96180, June 2014.
- <span id="page-7-2"></span>[36] Shuang Ma and Ying Long. Functional urban area delineations of cities on the Chinese mainland using mas-

sive Didi ride-hailing records. Cities, 97:102532, February 2020.

- <span id="page-8-7"></span>[37] Hernán A. Makse, Shlomo Havlin, and H. Eugene Stanley. Modelling urban growth patterns. Nature, 377(6550): 608–612, October 1995.
- <span id="page-8-8"></span>[38] Hernán A. Makse, José S. Andrade, Michael Batty, Shlomo Havlin, and H. Eugene Stanley. Modeling urban growth patterns with correlated percolation. Physical Review E, 58(6):7054–7062, December 1998.
- <span id="page-8-6"></span>[39] Stig Nordbeck. Urban allometric growth. Geografiska Annaler: Series B, Human Geography, 53(1):54–67, April 1971.
- <span id="page-8-10"></span>[40] Erneson A. Oliveira, José S. Andrade, and Hernán A. Makse. Large cities are less green. Scientific Reports, 4 (1):4235, December 2014.
- <span id="page-8-0"></span>[41] H. D. Rozenfeld, D. Rybski, J. S. Andrade, M. Batty, H. E. Stanley, and H. A. Makse. Laws of population growth. Proceedings of the National Academy of Sciences, 105(48):18702–18707, December 2008.
- <span id="page-8-1"></span>[42] H. D. Rozenfeld, D. Rybski, Xavier Gabaix, and H. A. Makse. The area and population of cities: new insights from a different perspective on cities. American Economic Review,  $101(5):2205-2225$ , August 2011.
- <span id="page-8-3"></span>[43] Annemarie Schneider, Mark A. Friedl, and David Potere. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. Remote Sensing of Environment, 114(8):1733–1746, August 2010.
- <span id="page-8-9"></span>[44] Kwok Tong Soo. Zipf's law for cities: a cross-country investigation. Regional Science and Urban Economics, 35(3):239–263, May 2005.
- <span id="page-8-11"></span>[45] Andrew J. Tatem. WorldPop, open data for spatial demography. Scientific Data, 4(1):170004, April 2017.
- <span id="page-8-4"></span>[46] Wenjia Wu, Hongrui Zhao, and Shulong Jiang. A Zipf's law-based method for mapping urban areas using NPP-VIIRS nighttime light data. Remote Sensing, 10(1):130, January 2018.
- <span id="page-8-2"></span>[47] Yuyu Zhou, Steven J. Smith, Christopher D. Elvidge, Kaiguang Zhao, Allison Thomson, and Marc Imhoff. A cluster-based method to map urban area from DMSP/OLS nightlights. Remote Sensing of Environment, 147:173–185, May 2014.
- <span id="page-8-5"></span>[48] G. K. Zipf. Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Cambridge: Addison-Wesley, 1949.