Cities 72 (2018) 201–205

Contents lists available at ScienceDirect

Cities

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

The power law within a metropolitan area

Xiaoyan Huang[a,*], Christopher Yost-Bremm[b]

[a] Texas A & M University, Department of Landscape Architecture and Urban Planning, College Station, TX 77840, USA
[b] San Francisco State University, Department of Finance, San Francisco, CA 94132, USA

ARTICLE INFO

Keywords:
Urban spatial structure
Employment cluster
Power law
Zipf's law
Urban growth

ABSTRACT

While theoretical reasoning emphasizes the complexity in metropolitan spatial structures, empirical efforts still focus on measuring spatial structures using employment centers. This paper measures a metropolitan spatial structure as a system of employment clusters for 361 metro areas in the United States. We apply a log-log model to assess the relationship between a cluster's size and its rank in a metro area. We found that among the largest 50 metro areas: (1) cluster sizes in a metro area follow a power law distribution and (2) larger metro areas tend to have evener spatial structures. The results suggest that policymakers can better predict urban growth locations and sizes; may invest in the largest clusters for the biggest economic payoffs; and should consider holistically all elements' (e.g., clusters, infrastructure, socioeconomic interactions) respective scaling laws in the city for its healthy urban growth.

1. Introduction

A metropolitan area (i.e., "metro area" or “city” in this paper) has differently sized employment clusters; it is more than a dual system of employment centers and non-center areas. A number of studies (Anderson & Bogart, 2001; Forestall & Greene, 1997; Giuliano & Small, 1991) measure metropolitan spatial structures in the United States (US) by identifying employment centers, separating a metro area into employment centers and non-center areas. However, there are different sizes of employment clusters ranging from small clusters to employment centers (usually defined as a minimum size of 10,000 workers). A few studies (Bogart, 2006; Gordon & Richardson, 1996; Yang, Steven, Holt, & Zhang, 2012) describe the complexity of metropolitan spatial structures as “beyond polycentricity” (Gordon & Richardson, 1996), a network of “trading places” (Bogart, 2006), and “complex variation in density” (Yang et al., 2012). However, the patterns of cluster sizes remain unexamined.

To examine the mechanism by which a metro area has differently sized employment clusters, we may look into the varying city sizes in a country. Cities exist mainly to take advantage of agglomeration economies (Glaeser, 2008). In other words, the agglomerative force is a major contributor, if not the only one, to the formation of a city. These cities' sizes in a larger region (e.g., a country) follow a power law (e.g., Zipf's law) distribution (Gabaix, 1999). Similarly, employment clusters also mainly exist to take advantage of agglomeration economies. Do the sizes of employment clusters in a larger region (i.e., a metro area) also follow a power law distribution?

The literature on studying cities as a science (Anas, Arnott, & Small, 1998; Batty, 2008; Batty, 2012; Bettencourt, 2013; Gabaix, 1999) suggests, “cities are complex systems that mainly grow from the bottom up, their size and shape following well-defined scaling laws that result from intense competition for space” (Batty, 2008) (p.769). This claim is supported by empirical evidence of power law distributions at the city, sub-city, county, and firm levels (Berry, 1964; Chen & Wang, 2014; Gabaix, 2011; Rubiera-Morollón, del Rosal, & Díaz-Dapena, 2015). Employment clusters, as an agglomeration unit above the firm but below the city level, may also follow a power law distribution.

The significance of this study is threefold. First, it fills the gap in literature of urban economics on employment centers (i.e., job concentration). Second, it tests the theory of scaling laws on a different level (i.e., employment cluster). Third, it helps to inform urban policymaking (e.g., infrastructure planning). We discuss the contributions in the conclusion.

This paper is organized as follows. We begin by measuring the metropolitan spatial structure as a system of clusters for 361 US metro areas. Considering the US metro areas varying enormously in density, we identify each metro area's employment clusters by adopting a relative threshold (Lee, 2007; Wheaton, 2004). Thresholds are calculated based on each metro area's employment distribution. Next, we apply Gabaix's growth model, which analyzes a system of cities within a country, to a system of clusters within a city (Gabaix, 1999). Our study focuses on the largest 50 metro areas, because they provide a closer analogy to mature stable economies like countries. Conversely, smaller metro areas may be on dynamic paths evolving towards maturity. Then,

* Corresponding author.
E-mail address: huan6281@tamu.edu (X. Huang).

http://dx.doi.org/10.1016/j.cities.2017.08.001
Received 19 October 2016; Received in revised form 24 July 2017; Accepted 4 August 2017
0264-2751/

we relate the findings to existing literature and propose policy implications.

2. Data and method

2.1. Data

We use shape files from the Topologically Integrated Geographic Encoding and Referencing (TIGER), along with employment data from the Census Transportation Planning Products (CTPP). TIGER provides shape files with metropolitan boundaries and census tracts within the metro areas. The metro areas in this study are the Metros in the Core Based Statistical Area (MCBSA) system. In the US (and Puerto Rico), there are 370 metro areas in the Census 2000 TIGER shape file. However, we exclude the eight Puerto Rico metro areas because no employment data were available. Within the 362 MCBSAs, Bristol, VA became part of Kingsport-Bristol-Bristol, TN-VA. Therefore, we examine 361 metro areas.

CTPP provides employment data summarized at the census tract level. We join the year 2000 CTPP employment data to the year 2000 TIGER shape files to obtain employment data of each census tract. The year 2000 employment data were collected from the US Census Bureau.

2.2. Method

2.2.1. Delimiting urban and rural areas

To ensure homogeneity of data units (i.e., Census tracts) in a metro area, we delimit urban and rural areas (or land), because spatial indices are "sensitive to the presence of large, unpopulated census tracts in outlying areas due to the well-known mismatch of administrative boundaries and functional areas." (Lee, 2007) (485) There are absolute and relative threshold methods to delimit urban and rural areas. The absolute threshold method adopts a universal cutoff for all metro areas, such as 1000 persons per square mile (as defined by the US Census). This definition may inappropriately exclude large areas of developed land (Lopez & Hynes, 2003). Conversely, the relative threshold method applies a cutoff percentage to a metro area's total population, e.g., 98% (Wheaton, 2004), or 95% (Lee, 2007).

We choose the relative over the absolute threshold method, considering the large variation in employment density among the US metro areas. We also use employment instead of population, because employment is better than population (based on primary residences) in capturing agglomeration economies. For example, activity centers such as the Central Business Districts usually have few residents.

After testing three percentages, we use 98% as the most reasonable threshold to delimit urban and rural areas. Because population is generally larger than employment (especially in rural areas), we initially try to define urban areas as having 99% of employment. However, the resulting polygons include large swathes of mountainous terrain that are not urban. We then try to use 95%, but the resulting polygons exclude many urban tracts that are near urban center. Finally, we produce a reasonable result by using 98% of employment.

We separate the census tract (polygons) with the lowest year 2000 employment density from a metro area one by one until the total employment is as close to (but not less than) 98% of the year 2000 total employment. The remaining census tracts within the metro area are defined as urban tracts. The rest of the census tract polygons, which contain roughly 2% of the metro's total employment, are defined as rural tracts.

2.2.2. Identifying employment clusters

Similar to delimiting urban and rural areas, identifying employment clusters involves either absolute or relative threshold criteria (Giuliano, Redfearn, Agarwal, Li, & Zhuang, 2005; Lee, 2007). One method adopts an absolute density threshold (e.g., 10 workers/acre) and a total employment threshold (e.g., 10,000 workers) (Giuliano & Small, 1991). A second method adopts a regression model (e.g., locally weighted regression) using distance starting from the Central Business District to identify secondary density peaks. It then uses a geographic window to ensure the secondary peak has statistically significantly higher density than its surrounding area (McMillen & Smith, 2003). A third method adopts a percentile threshold to identify a metro area's densest areas as potential employment centers and then applies a total employment threshold. The density threshold may be at the 90-percentile (Lee, 2007) or the 87.7-percentile (Pan & Ma, 2006) depending on the study areas.

We adopt the third method, as it meets the following criteria: (1) is capable of capturing each metro area's structural characteristics, (2) is easy to use for a large sample, and (3) uses a consistent density threshold throughout a metro area. Conversely, the first and second methods have the following drawbacks: The first method can only capture the spatial characteristics of high-density metro areas, and is not applicable for large sample studies. The second method results in inconsistent density thresholds within a metro area, and thus complicates the comparison analysis in a metro area. Furthermore, the second method is arbitrary in choosing "significance level, geographic window size, and weight of distance" (Matsuo, 2008) (p.27).

To apply the third method, we identify clusters by selecting high-density urban tracts and merge the neighboring tracts into a zone. High-density urban tracts have a density above the metro area's cluster density threshold. Each metro area's cluster density threshold is two standard deviations above the metro area's mean employment density in urban land. In Eq. (1), the probability (P) is a census tract's area proportional to a metro area's total urban land area, $\mathrm{P} = \sum_{i=1}^{n_j} \frac{S_{i,j}}{\sum_{i=1}^{n} S_{i,j}}$. The mean density (U) is a metro area's total employment in the urban land divided by the urban land area, $\mathrm{U} = \frac{\sum_{i=1}^{n_j} S_{i,j} d_{i,j}}{\sum_{i=1}^{n_j} S_{i,j}}$. A metro area's total urban land area is the sum of all census tracts in the urban land, $S_{i,j} = \sum_{i=1}^{n_j} S_{i,j}$. For metro area j, we have:

$$D_j = 2 \times \sqrt{\mathrm{P}(d_{i,j} - \mathrm{U})^2} + \mathrm{U} \tag{1}$$

Where,

D_j—Metro area j's cluster density threshold; $j = 1, 2, 3, \ldots, 361$;
$S_{i,j}$—Urban tract i's area in metro area j; $i = 1, 2, 3, \ldots, n_j$ (n_j = metro area j's total number of urban tracts);
$d_{i,j}$—Urban tract i's density in metro area j.

3. The power law hypotheses

We apply Gabaix's growth model (Gabaix, 1999) to a system of clusters in a city. As Gabaix pointed out, the power law phenomena exist because of scale invariance. In other words, the growth process of clusters is the same as that of cities. Furthermore, the underlying principle of forming clusters is the same as that of forming cities—agglomerative forces. Therefore, we hypothesize that the sizes of employment clusters within a metro area follow a power law distribution.

We use a log-log model (i.e., Eq. (2)) to represent the relationship between the cluster sizes and the cluster ranks in a metro area. As Fig. 1 shows, the largest cluster in a metro area has a rank value of one; the second largest cluster has a rank value of two, and so on. The absolute value of the slope k, in Eq. (2) can represent the degree of evenness of a metropolitan spatial structure. The smaller the absolute value, the evener the spatial structure. The value of k is expected to be negative, similar to the coefficient value for the system of cities (i.e., −1.005) (Gabaix, 1999). However, given that the asymptotic distribution does depend on the number of clusters, it is not necessarily the case that the coefficient should be close to −1.

$$logRank = k \log(Size\ of\ employment\ cluster) + b \tag{2}$$