Scaling of foreign attractiveness for countries and states

Iva Bojic [a, b, *], Alexander Belyi [b, c], Carlo Ratti [a], Stanislav Sobolevsky [a, d]

[a] Senseable City Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[b] SENSEable City Laboratory, SMART Centre, 1 Create Way, Singapore
[c] Belarusian State University, 4 Nezavisimosti Avenue, Minsk, Belarus
[d] Center for Urban Science and Progress, New York University, 1 MetroTech Center, Brooklyn, NY 11201, USA

ARTICLE INFO

ABSTRACT

People's behavior on online social networks, which store geo-tagged information showing where people were or are at the moment, can provide information about their offline life as well. In this paper we present one possible research direction that can be taken using Flickr dataset of publicly available geo-tagged media objects (e.g., photographs, videos). Namely, our focus is on investigating attractiveness of countries or smaller large-scale composite regions (e.g., US states) for foreign visitors where attractiveness is defined as the absolute number of media objects taken in a certain state or country by its foreign visitors compared to its population size. We also consider it together with attractiveness of the destination for the international migration, measured through publicly available dataset provided by United Nations. By having those two datasets, we are able to look at attractiveness from two different perspectives: short-term and long-term one. As our previous study showed that city attractiveness for Spanish cities follows a superlinear trend, here we want to see if the same law is also applicable to country/state (i.e., composite regions) attractiveness. Finally, we provide one possible explanation for the obtained results.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

It is true that advancement of transport systems in the last 100 years provided people with more mobility options, but it was only a little bit more than 5 years ago, with the emergence of social network platforms, shared economy and multiple mobile devices, that digital records on human mobility became public in a real-time. Namely, in a society relying on shared economy traveling cost per capita are reduced which resulted in a higher number of people who actually move and while doing that they use their smartphones equipped with Global Position Systems (GPS) enabling them to capture their geo-location and publicly share it on their social networks. Although it is still not clear how *living online* will affect us in future, we can be sure for one thing — the huge digital data produced by people all around the world represents almost an inexhaustible resource that can be used in different scientific studies.

Different types of such data are now being increasingly utilized for estimating human mobility, including cell phone data (Calabrese, Di Lorenzo, Liu, & Ratti, 2011; Gonzalez, Hidalgo, & Barabasi, 2008; Hoteit, Secci, Sobolevsky, Ratti, & Pujolle, 2014; Kung, Greco, Sobolevsky, & Ratti, 2014), vehicle GPS traces (Kang, Sobolevsky, Liu, & Ratti, 2013; Santi et al., 2014) or credit card transactions (Lenormand et al., 2015; Sobolevsky et al., 2014). However, so far geo-tagged social media was one of a few sources allowing to approach human mobility on a global scale (Belyi et al., 2016; Hawelka et al., 2014; Paldino, Bojic, Sobolevsky, Ratti, & González, 2015). Applications of those studies include diverse examples such as urban transportation (Santi et al., 2014), determining economical potential of the cities (Sobolevsky, Massaro, Bojic, Arias, & Ratti 2015), land use classification methods (Grauwin, Sobolevsky, Moritz, Gódor, & Ratti, 2015; Pei et al., 2014), or regional delineation (Amini, Kung, Kang, Sobolevsky, & Ratti, 2014; Ratti et al., 2010).

Geo-location part of information on human mobility allows scientific studies to focus on specific areas, such as on cities which are important as today more than 50% of the world population lives in them. The trend of the urban population growth is expected to be continued in future as well and it is estimated that in the next

30–40 years the number of city dwellers might nearly double (United Nations, 2014). However, cities are not only studied because of the number of people living there, but also because of their long and rich history. Many cities around the world started their development back in medieval or classical times and during that long span of time experienced many demographic, economic, political and spatial transformations (Hall & Raumplaner, 1998). Nevertheless, cities are not only exceptional and unique, but they also do share certain common properties as shown in scientific studies across different disciplines (e.g., economics, engineering, complex systems). These effects are known as agglomeration or scaling relation between macroscopic properties of cities (e.g., employment rate) and its size (Bettencourt, 2013).

Although over the years urban scaling theory shed lights on superlinear behavior of socioeconomic (e.g., intensity of interactions (Schläpfer et al., 2014), creativity (Bettencourt, Lobo, Strumsky, & West, 2010), economic efficiency (Bettencourt, 2013)) and sublinear scaling of urban infrastructure (Batty, 2008), it is still not clear if this behavior could be extended to other scales, such as provinces or the whole states and countries. In this study we focus on investigating scaling laws of attractiveness, defined as the number of digital media objects created by foreign visitors from Flickr dataset or number of foreign citizens/foreign born population living in different countries across the world captured in dataset created by United Nations. Our previous studies presented results on city attractiveness in Spain, showing strong superlinearly dependence (with scaling exponent around 1.5) (Bojic et al., 2015; Sobolevsky, Bojic, et al., 2015; Sobolevsky, Sitko, et al., 2015) and here we extend our analysis to other spatial scales as well.

In this paper we discuss possible reasons why and at which level attractiveness starts to scale sublinearly. Namely, city attractiveness scales superlinearly in contrast to country level that shows a sublinear behavior. In order to find a turning point between cities and countries, in our analysis we included the US states as an intermediate level and showed that on that level the scaling exponents indeed lay between cities' and countries' exponents. The rest of the paper is organized as follows. Section 2 describes two datasets used to check country/state scaling laws. Furthermore, Section 3 shows results for country attractiveness and scaling of attractiveness of the US states. Finally, Section 4 discusses our results and provides final conclusion marks.

2. Datasets

In this study we use publicly available Yahoo Flickr Creative Commons dataset of the largest public multimedia collection that has ever been released and created as a part of the Yahoo Webscope program[1]. It contains 100 million media objects that were created between 2004 and 2014 where each media object in the dataset is represented by its metadata: object identifier, user identifier, time stamp when it was taken, location (i.e., latitude and longitudinal coordinates) where it was taken (if available), and CC license it was published under (Thomee et al., 2016). Additionally, the metadata for some objects also contains object title, user tags, machine tags and description, as well as a direct link for downloading the content. In addition to the amount of media objects and supporting metadata for each object, the main advantage of this dataset compared to the other similar ones is that all media objects are licensed under one of the Creative Commons copyright licenses, and as such can be used for benchmarking purposes as long as media objects are credited for the original creation.

Once when downloaded the raw dataset, we pruned the data by

first omitting records that were not geo-tagged (i.e., more than 50%) and then by omitting those that came with the wrong date format (less than 0.01%). Since the focus of the paper is to investigate how attractiveness scales on a country level, we had to perform reverse geo-coding so that every geo-location (i.e., latitude and longitude coordinates) is translated into an area we observed (e.g., state or country). We used shapefiles that define boundaries of world countries[2] and the US states[3]. After both pruning and reverse geo-coding we were left with 40 million objects created in 238 countries around the world. In addition to Flickr dataset, as another source for verification of a sublinear country scaling law, we are using official migration statistics provided by United Nations[4]. This statistic is provided in a form of an origin/destination (OD) matrix which represents an estimate of how many foreign citizens had lived in each country in July of 2010.

3. Attractiveness scaling laws

As defined in our previous paper (Sobolevsky, Bojic, et al., 2015) foreign attractiveness for Flickr dataset is quantified as the number of media objects created by people visiting the considered area (e.g., country) from other countries. Similarly, attractiveness for migration dataset is represented with the number of the current country residents who moved to that country from abroad. Since Flickr dataset does not include information about where people live, we used the following methodology of inferring the user's home country — we say that a particular user lives in a particular country if there he/she created the maximum number of media objects and spend the maximal number of days compared to all other countries across the world (if maximal activity and activity period give inconsistent results, we consider that home country is undefined). More detailed discussion on the importance of choosing the right home definition method can be found in (Bojic, Massaro, Belyi, Sobolevsky, & Ratti, 2015). Once when finished with inferring the home location, we took the activity only of those users for whom we detected where they lived (we were left with 72.4% of original objects) and calculated aggregated attractiveness as the number of media objects taken by foreign users (15.2% of all objects). Aggregated attractiveness means that we took the total aggregated user activity over the entire timeframe of the data availability without taking into consideration time when it as made.

Flickr dataset and dataset provided by United Nations (i.e., migration dataset), which are described in the previous section, reveal different aspects of country attractiveness—short-term and long-term one. Namely, rather than showing mostly a touristic behavior, a migration dataset counts in for people who decided to temporarily or more permanently move to another country. Even before trying to find scaling law patterns for both datasets, in Fig. 1 we show distribution of country attractiveness as well as its population. Both types of attractiveness are distributed similarly in shape and follow the same log-normal patterns as country population distribution does—although variance for them is a slightly different: a higher one for attractiveness. Now let us consider how attractiveness of each specific country is actually related to its population.

The results of fitting a power-law dependence $A \sim ap^{\beta}$ to the total number of media objects denoting aggregated attractiveness A and population p are shown in Fig. 2. Fitting process is performed

---

دانلود مقاله

http://daneshyari.com/article/83130