

## MIT Open Access Articles

*Scaling of City Attractiveness for Foreign Visitors through Big Data of Human Economical and Social Media Activity*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Sobolevsky, Stanislav, Iva Bojic, Alexander Belyi, Izabela Sitko, Bartosz Hawelka, Juan Murillo Arias, and Carlo Ratti. "Scaling of City Attractiveness for Foreign Visitors through Big Data of Human Economical and Social Media Activity." 2015 IEEE International Congress on Big Data (June 2015).

**As Published:** <http://dx.doi.org/10.1109/BigDataCongress.2015.92>

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Persistent URL:** <http://hdl.handle.net/1721.1/101645>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike





## **Senseable City Lab :::: Massachusetts Institute of Technology**

This paper might be a pre-copy-editing or a post-print author-produced .pdf of an article accepted for publication. For the definitive publisher-authenticated version, please refer directly to publishing house's archive system

# Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity

Stanislav Sobolevsky, Iva Bojic, Alexander Belyi  
*SENSEable City Lab, MIT*  
*Cambridge, MA, USA*

*Email: stanly@mit.edu, ivabojoic@mit.edu, abely@mit.edu*

Juan Murillo Arias  
*New Technologies, BBVA*  
*Madrid, Spain*

*Email: juan.murillo.arias@bbva.com*

Izabela Sitko, Bartosz Hawelka  
*University of Salzburg*  
*Salzburg, Austria*

*Email: izabela.sitko@sbg.ac.at, bartosz.hawelka@sbg.ac.at*

Carlo Ratti  
*SENSEable City Lab, MIT*  
*Cambridge, MA, USA*

*Email: ratti@mit.edu*

**Abstract**—Scientific studies investigating laws and regularities of human behavior are nowadays increasingly relying on the wealth of widely available digital information produced by human activity. In this paper we use big data created by three different aspects of this activity (i.e., bank card transactions, geotagged photographs and tweets) in Spain for quantifying city attractiveness for foreign visitors. An important finding of this paper is a strong superlinear scaling law of city attractiveness with its population size. The observed scaling exponent stays nearly the same for different ways of defining cities and for different data sources, emphasizing the robustness of our finding. We also consider temporal variation of the scaling exponent in order to reveal seasonal patterns in the attractiveness. Finally, we propose a possible explanatory mechanism for the observed superlinear effect based on a simple discrete choice model.

**Keywords**-human mobility; urban attraction; Flickr; Twitter; bank cards; superlinear scaling; discrete-choice model.

## I. INTRODUCTION

As everyday people around the world leave more and more of digital traces behind them, scholars intensively use such data for studying human behavior, including mobile phone records [1]–[4], vehicle GPS traces [5], [6], smart cards usage [7], [8], social media posts [9]–[11] and bank card transactions [12]–[14]. Results of those studies can be applied to a wide range of policy and decision-making challenges, such as regional delineation [15], [16] or land use classification [17], [18] for instance. Moreover, some of them focus specifically on studying human mobility at urban [19]–[21], country [22] or even global scale [23], [24].

In this paper we investigate international human mobility patterns on an example of Spain by quantifying and analyzing the ability of different cities to attract foreign visitors (for various purposes) and proposing a mechanism that can explain observed patterns. For the purpose of the study we use information about interactions between people and businesses registered through bank card transaction

records and between people and urban spaces using geotagged photographs and tweets. By using Flickr and Twitter datasets, as well as bank card transaction dataset provided by Banco Bilbao Vizcaya Argentaria (BBVA) bank, our goal is to investigate how a city ability to attract foreign visitors depends on its size. City attractiveness is defined as the absolute number of photographs, tweets or bank card transactions made by foreign visitors in that city.

Discovering how to improve city attractiveness, which is seen differently in eyes of residents and tourists, can be used in several fields such as planning, forecasting flows, tourism, economics and transportation [25]. In the past photography has already been considered as a good proxy in architecture and urban planning for example when trying to understand landscapes [26]. Girardin et al. showed that it was possible to define city attractiveness by exploring big data from photo sharing websites [27]. Moreover, shopping patterns of tourists, including their specific preferences and satisfaction level, were analyzed with the overall purpose of accurate planning, marketing and management of sales strategies [28]–[32]. However, to the best of our knowledge, it has not been shown how city attractiveness can be quantified and explained from a multiple side perspective of diverse datasets created by different aspects of individual activity.

The novelty of this paper is threefold: applying new multi-layered data (i.e., Flickr, Tweeter and bank card transactions) for quantifying urban attractiveness for foreign visitors, detecting a strong and robust pattern - superlinear scaling of attractiveness with a city size, and finally proposing a possible explanatory mechanism for it. More specifically, our study focuses on the way how attractiveness of cities depends on their size. Validating the robustness of our findings, we use different city definitions and different datasets to quantify the attractiveness. Finally, we propose a simple discrete choice model that provides a possible explanation for the observed effect using very basic assumptions.

## II. DATASETS

In our analysis we combine three different datasets: the first one containing more than 100 million publicly shared geotagged photographs on Flickr that people took around the world during a period of several years, the second containing geotagged tweets posted on Twitter worldwide during 2012, and the last one containing a set of bank card transactions of domestic and foreign users recorded by Banco Bilbao Vizcaya Argentaria (BBVA) during 2011, all over Spain. The aforementioned data allows us to analyze activity of foreign visitors in Spain from three different perspectives - while making purchases, taking photographs or sharing their thoughts on interesting places they visited.

### A. Flickr dataset

By merging two Flickr datasets [33], [34] we created a new dataset containing more than 130 million photographs/videos. Both datasets are publicly available and can be downloaded using the given links. Each dataset consists of over 100 million photographs or videos taken by more than one million users. The records in two datasets partially overlap, but since each photograph/video has its id, we were able to merge them by omitting duplicates and choosing only those entries that were made within a ten year time window (i.e., from 2005 and until 2014).

In order to determinate users who are acting as foreign visitors in a certain location, for each user in the merged dataset we define his/her home country by using the following criteria: user is considered to be a resident of a certain country if that is the country where he/she took the highest number of photographs/videos over the longest time period (calculated as the time between the first and last photograph taken in that country) compared to all other countries he/she visited. Using our criteria we were able to determine user home country for over 500 thousand users. Those users took almost 80% of all the photographs/videos in the dataset (i.e., more than 90 million in total), while the rest of users, for which their home country could not have been defined, mostly belong to a low active group taking photographs only occasionally. In our study we only consider those users with defined home countries. From the total of over 3.5 million photographs taken in Spain, over 400 thousand were taken by over 16 thousand of foreign visitors coming from 112 countries all over the world.

### B. Twitter dataset

The second dataset consists of geotagged messages posted during 2012 and collected from the digital microblogging and social media platform Twitter. Data was collected with Twitter Streaming API [35] and cleaned from potential errors and artificial tweeting noise as previously described in [23]. Globally, the dataset covers 944 million tweets sent by 13 million users [23]. The number of tweets posted in Spain in 2012 is almost 35 million messages sent by 641 thousand

Twitter users. To differentiate between Spanish residents and foreign visitors, we used a similar technique as it was used in the case of Flickr dataset. We found out that 2% of the total number of tweets posted in Spain in 2012 was sent by 80 thousand foreign visitors from 180 countries.

### C. BBVA dataset

The third dataset used in this study is a set of bank card transactions registered by the Spanish bank BBVA during 2011. Those transactions are of two types: i) made using debit or credit cards issued by BBVA, or ii) made using cards issued by other banks in any of over 300 thousand BBVA card terminals. For the latter transactions dataset includes the country of origin where the card was issued. In our study we focus on this second group, in particular on 17 million transactions made by 8.6 million foreigner visitors from 175 different countries.

Due to the sensitive nature of bank data, our dataset was anonymized by BBVA prior to sharing, in accordance to all local privacy protection laws and regulations. Therefore, cards are identified only by randomly generated IDs, and all the details that could allow re-identifying a card holder were removed. The raw dataset is protected by the appropriate non-disclosure agreement and is not publicly available.

## III. SCALING OF CITY ATTRACTIVENESS IN SPAIN

Cities are known not only to be places where people live, but also environments that are transforming human life. A bigger city boosts up human activity: intensity of interactions [36], creativity [37], economic efficiency (e.g., measured GDP [38]), as well as certain negative aspects: crime [37] or infectious diseases [39]. Due to agglomeration effects and intensified human interactions, many aggregated socioeconomic quantities are known to depend on a city size in a form of superlinear scaling laws, meaning that those quantities are not simply growing with the city size, but are actually growing faster compared to it. However, other studies on urban infrastructure dimensions (e.g., total road surface) reveal a sublinear relation to the city size [38]–[41].

All of the above mentioned quantities are mainly related to the processes that are happening inside cities. In this paper we pose another closely related, but slightly different question about city external appearance seen as its attractiveness for foreign visitors. Worth mentioning is that by "attractiveness" we do not only mean that some place is attractive as touristic destination, but we also do consider its ability to attract foreign visitors for whatever reasons – touristic, business, or any other personal matter-related ones.

In this study we focus on Spain as it is the country which economy heavily depends on international tourism giving a paramount importance to the ability of its cities to attract foreign visitors. Namely, even if they are coming for a primary purpose other than tourism, visitors still often act like tourists do, making their contribution to the touristic

sector. Many tourist rankings only consider the number of people visiting a city, consequently often leading to a fairly obvious conclusion that larger cities are more attractive as they can accommodate more tourists. The question we are interested in is how the total amount of visitor activity typically scales with the city size. Understanding this kind of scaling allows us to predict the expected performance for a particular city and to estimate if it is actually under- or over-performing compared to the average expectation for other cities of that size.

The first step of conducting an analysis of city attractiveness is to decide what should be considered to be a city. There is a number of ways of how to define a city and selecting an appropriate city definition is important for analyzing aggregated urban performance [38], [40], [42]. For the purpose of our study we use definitions proposed by European Urban Audit Survey (EUAS) [43], European Spatial Planning Observation Network (ESPON) [44] and Afeas Urbanas de Espana (AUDES) [45] project. On the most fine grain level, AUDES project defines 211 conurbations (CONs) in Spain. On a more aggregate level, ESPON defines 40 Functional Urban Areas (FUAs), and finally on the most aggregated level EUAS defines 24 Large Urban Zones (LUZs). Population for LUZ and FUA were obtained from Eurostat [46] and National Statistics Institute of Spain [47], and for CONs from the AUDES project. In addition to those three city definitions, we also perform our analysis for 52 Spanish provinces in order to see if our conclusions could be actually extrapolated from a city level to more general consistent geographical entities.

#### A. Aggregated city attractiveness

To explore the overall city attractiveness for foreign visitors we focus on three different aspects of their activity - bank card transactions, taking photographs and twitting during their visit. Rather than simply counting the number of visitors, we use the total amount of the described activity to quantify an attractiveness measure. Namely, unlike the quantity of people who visited a city at least once, the amount of activity also takes into an account the length of their stay and the intensity of exploring the city. We believe that this can serve as a more relevant proxy for the average visitor activity in the observed city at a certain time.

Figure 1 shows the results of fitting a power-law dependence  $A \sim ap^b$  to the observed pairs of attractiveness  $A$  and population  $p$ . Fitting is performed on a log-log scale where it becomes a simple linear regression problem  $\log(A) \sim \log(a) + b \cdot \log(p)$ . The fitted scaling trends are substantially superlinear for all three datasets (i.e., BBVA, Flickr and Twitter) and for all types of city definitions (i.e., CONs, FUAs, LUZs) as well as for the provinces. Regardless of the spatial scale used in the analysis, the scaling exponent  $b$  remained approximately the same for all three datasets considered (i.e., around 1.5 with the highest

level of fluctuation for provinces, compared to the lowest one for LUZs) confirming the robustness of the pattern. Moreover, this pattern seems to be quite significant – such a high exponent indicates that for example attractiveness of one city that is 3 times bigger than the other one, is expected not to be 3 times, but on average 5 times higher.

In order to double-check if the average scaling trend is really consistent with the fitted power-law, we perform a binning process of the data by considering average attractiveness of all the cities falling into each of the five population ranges, evenly splitting the entire sample on the log-scale. As one can see from Figure 1, the binned trend in all cases goes pretty much along the fitted power-law, confirming the scaling shape. Finally, the analysis of statistics fit confirms statistical significance of the trends —  $R^2$  values for the binned trends fitted by power laws vary as  $97.6 \pm 1.8\%$  for all three ways of city definition and all three datasets, quantitatively confirming the observed visual similarity between the trends and the superlinear power laws. Quantifying  $R^2$  for the power-law fits to the scattered plots including the entire variation of all the individual original city data still keeps  $R^2$  high enough, reporting  $57.5 \pm 12.9\%$  of the total data variation being explained by the superlinear trends. At the same time  $p$ -values are always below 1% (usually much lower) in all cases considered, serving as an ultimate evidence of the trend significance.

#### B. Temporal aspects of city attractiveness

In Section III-A we considered the overall aggregated city attractiveness over the entire time frame of data availability (i.e., one year in case of Twitter and BBVA datasets or ten years in a case of Flickr data). However, one should be aware that this pattern changes over different months. Therefore, in this section we investigate whether or not this seasonal change affects the observed attractiveness scaling. In order to reduce the possible negative effect of low data volumes on the stability of the results, we consider the aggregated activity over a three-month period. For example, a data point that shows city attractiveness value for April denotes an average for March, April and May.

Figure 2 shows a substantial dependence between the observed exponent and the time of year (i.e., month). Although, the trend appears to be mostly consistent across different ways of city definition, it shows a slight variation depending on the dataset. Nevertheless, the main pattern is always the same and is confirmed by all trends considered – it always drops over summer period as it seems that people tend to explore Spain more extensively and visit more different locations. This could be easily explained by a higher touristic activity over summer, especially focused on beach tourism being spread over a number of smaller destinations along the coast, while the rest of a year, especially in spring and autumn, there is a larger number of business foreign visitors who are primarily attracted to the major cities.

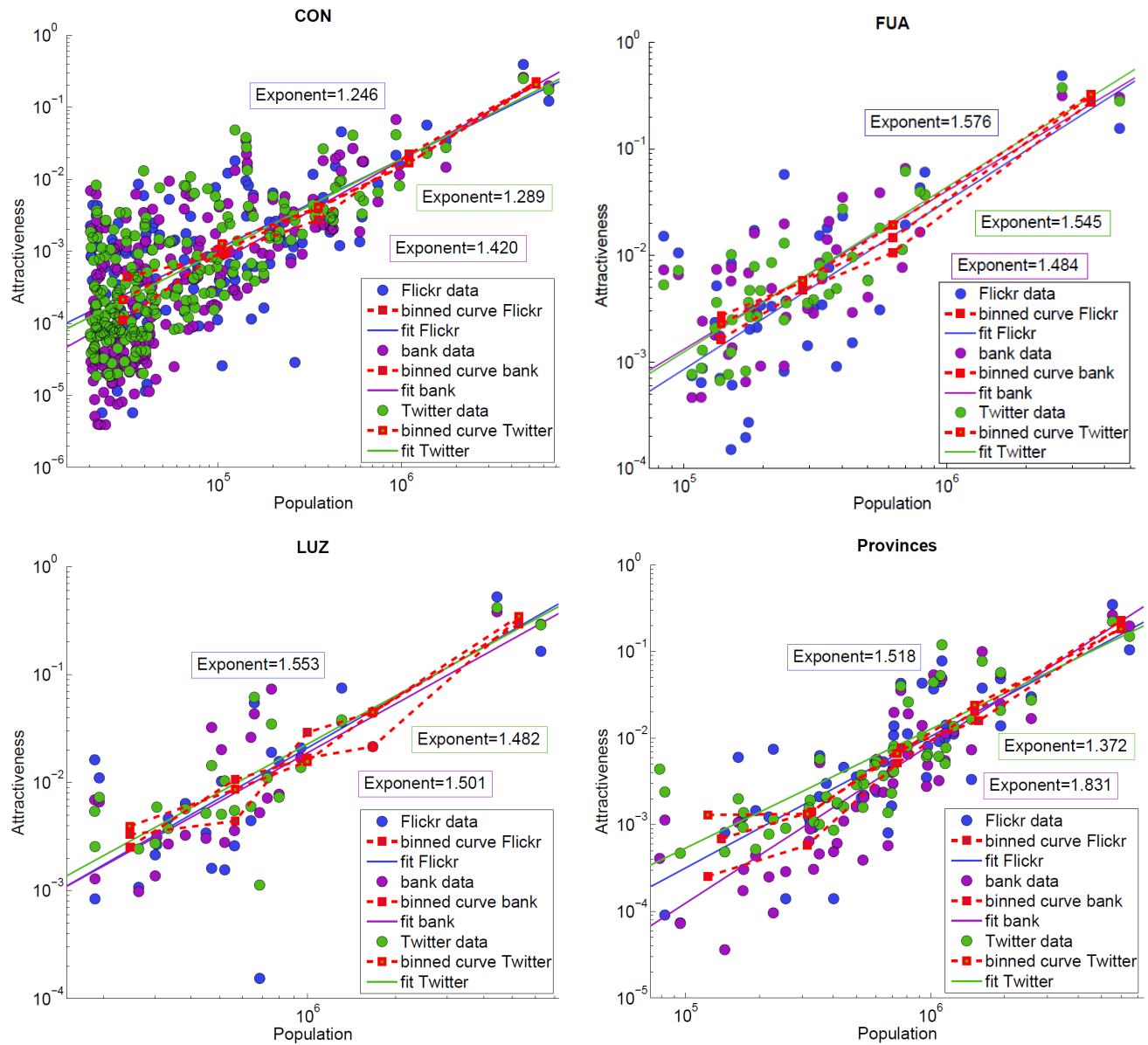


Figure 1. Scaling of city attractiveness with population size observed through three different datasets for different Spanish city definitions as well as for the provinces. X-axis represents the number of people living in the city, while Y-axis represents the fraction of the number of photographs/videos, tweets or bank card transactions registered in the city versus the total amount registered in all cities.

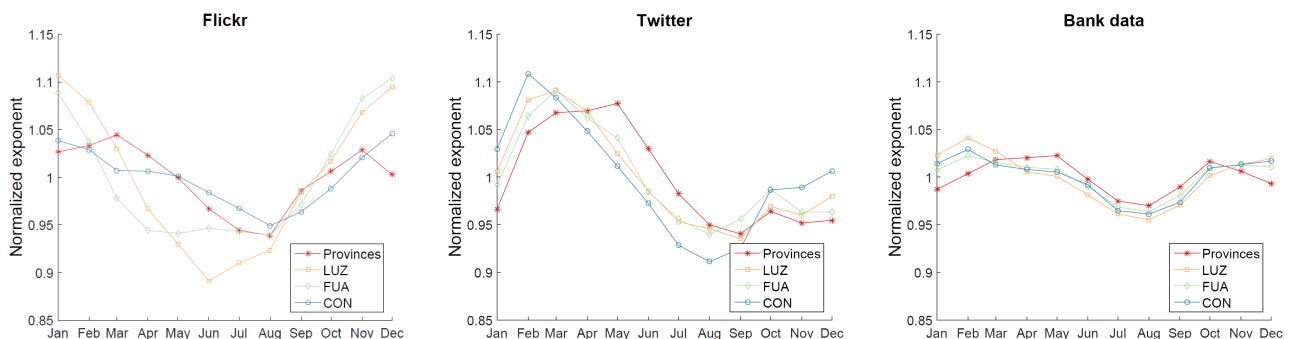


Figure 2. How scaling exponent changes over a year, normalized by the yearly average.

### C. Learning from deviations - scale-free city attractiveness

The aforementioned superlinear power-law trends describe the way city attractiveness scales with the city size on average. However, each particular city performance can be different from the trend prediction. This actually opens up a possibility for the scale-independent scoring of the city attractiveness by considering the log-scale residual of the actual attractiveness value vs. the trend prediction:  $\text{res} = \log(A) - b \cdot \log(p) - \log(a)$ . When this residual is positive, it points out towards the city over-performance vs. the average trend, while the negative value points out that the city is under-performing.

Just to give an example, Figure 3 visualizes residuals for Spanish LUZs ordering the cities from the most over- to the most underperforming ones according to the bank card transactions data. One can notice that although residuals from different datasets are different, the patterns are generally consistent – cities strongly over/under-performing according to one dataset usually do the same according to the others. However, there are some interesting exceptions for example in cases of Malaga and Alicante. Namely, from bank card transactions and Twitter data it seems that foreign people visit those cities, but it appears that they have relatively lower motivation for taking pictures of those cities. This could point out to more specific types of visitor activity in those places justifying our choice of combining different datasets to get more a precise picture of city attractiveness. Furthermore, the two island cities: Santa Cruz de Tenerife and Las Palmas are attracting foreign visitors to spend their money, but do not seem to encourage them enough to perform both online activities: Twitter or Flickr. Nevertheless, general consistency of the residual values defined through different datasets also holds on scales of FUAs, CONs, as well as provinces. Table I presents the pairwise correlation values between those residuals that typically fall in the range between 50 and 80%.

Table I  
CORRELATIONS (%) BETWEEN CITY/PROVINCES RESIDUALS DEFINED THROUGH DIFFERENT DATASETS.

	Bank/Twitter	Bank/Flickr	Twitter/Flickr
Provinces	84.45	46.77	52.89
LUZs	62.95	52.72	77.09
FUAs	73.68	59.98	77.90
CONs	80.07	56.90	58.56

### IV. POSSIBLE EXPLANATORY MECHANISM

We found a strong empirical evidence for a highly super-linear nature of how city attractiveness scales with the city size. Being confirmed from different datasets and various levels of city definition, this finding seems to be quite robust. In this section we propose a possible explanatory mechanism for it – a simple discrete choice model based on very basic

assumptions, including linear dependence of the number of potential attractions on the city size, with no superlinearity being directly or indirectly assumed. The model is able to produce superlinear scaling of city attractiveness with various exponents depending on one single model parameter.

Assume that every foreign visitor, which is model as an agent, makes a decision about which city to visit within the selected country following a simple criterion. Suppose that each city  $c$  has a number of Points of Interest (PoI)  $p(c)$  that is directly proportional to the city size, while the individual attractiveness of each of those PoI for the particular agent is a random variable following a certain probability distribution (the same for each point of interest) with the unit mean and variance  $\sigma^2$ . If a standard deviation  $\sigma$  is a parameter of the model, then the cumulative attractiveness utility score  $U(c)$  of each city  $c$  for the particular agent is estimated as a sum of those individual attractiveness scores for all the PoI within the considered city. Finally, the agent makes an obvious choice of visiting the city that has the highest cumulative utility score  $U(c)$  compared to all other cities.

Although the model can be easily implemented using agent-based simulation, we also provide a probabilistic analytic framework for it. According to the central limit theorem, no matter what the original distribution for the individual POI attractiveness is,  $U(c)$  can be considered to follow an approximately normal distribution with the mean  $p(c)$  and variance  $\sigma^2 p(c)$ . Finally, the probability  $P(c)$  of each city to be chosen, which serves as a direct proxy for total city attractiveness  $A(c)$  when multiplied by the total number of agents in the model, is defined as following:

$$P(c) = \int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{p(c)}} \phi\left(\frac{x - p(c)}{\sigma \sqrt{p(c)}}\right) \prod_{c^* \neq c} \Phi\left(\frac{x - p(c^*)}{\sigma \sqrt{p(c^*)}}\right) dx, \quad (1)$$

where  $\phi$  and  $\Phi$  are probability density and cumulative distribution functions of the standard normal distribution, respectively. Once the size estimates  $p(c)$  are given for each city  $c$ , the above probabilities  $P(c)$  depend only on the model parameter  $\sigma$ . A numerical experiment shows that the distribution of  $P(c)$  values asymptotically follows a power-law, while its slope depends on the parameter  $\sigma$  as reported in Figure 4a. In this experiment we assume city population  $p$  to follow a Zipf law  $p(r) \sim 1/r$  as it has already been shown in some of the previous studies [48].

In order to see how well our model performs in describing the actual empirical data, we applied it to the actual LUZ data fitting the parameter  $\sigma$  to match the model prediction with the observed LUZ attractiveness. The optimal value of  $\sigma = 294$  was found to minimize the log-scale square deviation  $\sum (\log(P_{\text{model}}(c)) - \log(P_{\text{data}}(c)))^2$ . Figure 4b shows the fitted model prediction for LUZs.

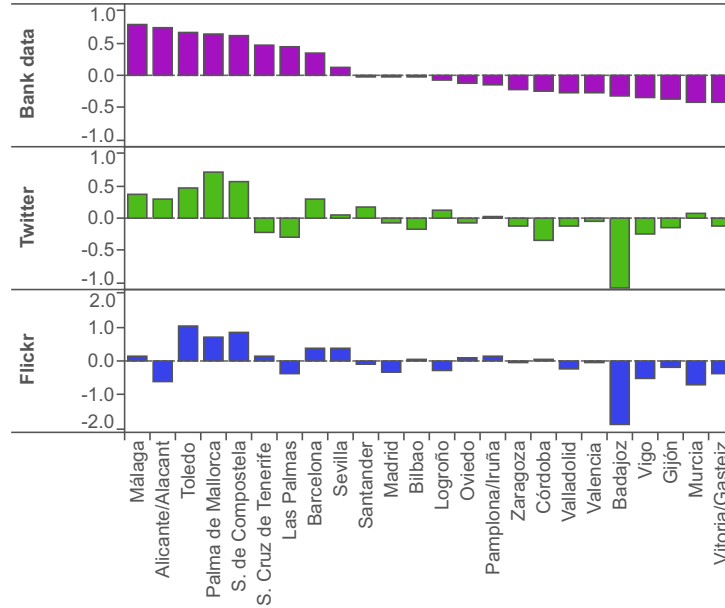


Figure 3. LUZ scale-independent attractiveness through three data sets.

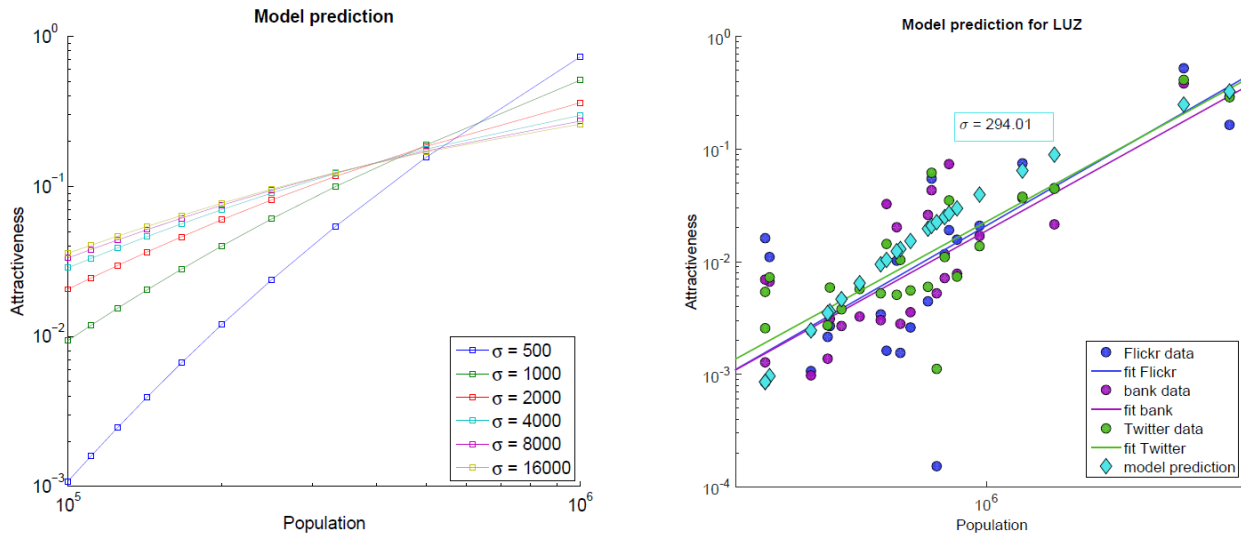


Figure 4. City attractiveness scaling predicted by the model for a) artificially generated city population sizes following Zipf law for different  $\sigma$ , b) actual population of 24 LUZs with the optimal  $\sigma = 294$ .

However, those observations by themselves cannot prove that people actually follow the suggested decision making mechanism, which one would expect to be much more complicated in real life. Obviously, more experiments are needed in order to validate that the suggested framework could be actually useful for partial prediction of human real behavior. Nevertheless, the framework serves as a good illustration of how the empirically observed superlinear scaling of urban attraction could be derived from simple reasonable assumptions and basic probabilistic framework not involving any superlinearity by themselves.

## CONCLUSIONS

In this study we used three types of big data created by human activity to quantify how cities in Spain can attract foreign visitors. In general, city attractiveness was found to demonstrate a strong superlinear scaling law with the city size. A high consistency of the scaling exponents found across different ways of defining cities, as well as across all three datasets used in the study, provides an evidence for the robustness of our finding and also serves as an indirect proof of the applicability of selected datasets for that purpose.



Moreover, we analyzed the temporal variation of the scaling exponent during a year, which was found to reveal a very intuitive pattern quantified by a noticeable drop of the exponent value over the summer – visitor activity seems to be more spread across different smaller destinations within the country over more touristic summer time, while more concentrated at major destinations in autumn and spring, presumably because of having more business visitors in the country. Once again the pattern appears to be pretty robust and consistent as being confirmed from all datasets for all different city definitions.

Finally, we propose a very simple possible explanatory mechanism that is able to demonstrate how the observed superlinear scaling of city attractiveness could be derived through basic intuitive assumptions about individual decision making and straightforward probabilistic framework. In our future work we will test the performance of our model on a larger set of data, as well as look at reasons why people find cities attractive in the first place.

#### ACKNOWLEDGMENTS

The authors would like to thank Banco Bilbao Vizcaya Argentaria (BBVA) for providing the dataset for this research. Special thanks to Assaf Biderman, Marco Bressan, Elena Alfaro Martinez and María Hernández Rubio for organizational support of the project and stimulating discussions. We further thank BBVA, MIT SMART Program, Center for Complex Engineering Systems (CCES) at KACST and MIT, Accenture, Air Liquide, The Coca Cola Company, Emirates Integrated Telecommunications Company, The ENEL foundation, Ericsson, Expo 2015, Ferrovial, Liberty Mutual, The Regional Municipality of Wood Buffalo, Volkswagen Electronics Research Lab, UBER, and all the members of the MIT Senseable City Lab Consortium for supporting the research. Part of this research was also funded by the Austrian Science Fund (FWF) through the Doctoral College GIScience (DK W 1237-N23), Department of Geoinformatics - Z\_GIS, University of Salzburg, Austria. Finally, the authors also acknowledge support of the research project "Managing Trust and Coordinating Interactions in Smart Networks of People, Machines and Organizations", funded by the Croatian Science Foundation.

#### REFERENCES

- [1] C. Ratti, S. Williams, D. Frenchman, and R. Pulselli, "Mobile landscapes: Using location data from cell phones for urban analysis," *Environment and Planning B*, vol. 33, no. 5, pp. 727–748, 2006.
- [2] F. Calabrese and C. Ratti, "Real time rome," *Networks and Communication Studies*, vol. 20, no. 3-4, pp. 247–258, 2006.
- [3] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat, "Digital footprinting: Uncovering tourists with user-generated content," *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 36–43, 2008.
- [4] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft, "Recommending social events from mobile phone location data," in *Proceedings of the 10th IEEE International Conference on Data Mining*, 2010, pp. 971–976.
- [5] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. Strogatz, and C. Ratti, "Quantifying the benefits of vehicle pooling with shareability networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 37, pp. 13 290–13 294, 2014.
- [6] C. Kang, S. Sobolevsky, Y. Liu, and C. Ratti, "Exploring human movements in Singapore: A comparative analysis based on mobile phone and taxicab usages," in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, 2013, pp. 1–8.
- [7] M. Bagchi and P. White, "The potential of public transport smart card data," *Transport Policy*, vol. 12, no. 5, pp. 464–474, 2005.
- [8] N. Lathia, D. Quercia, and J. Crowcroft, "The hidden image of the city: Sensing community well-being from urban mobility," in *Pervasive Computing*, ser. Lecture Notes in Computer Science, 2012, vol. 7319, pp. 91–98.
- [9] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: Understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis*, 2007, pp. 56–65.
- [10] M. Szell, S. Grauwin, and C. Ratti, "Contraction of online response to major events," *PLoS One*, vol. 9, no. 2, pp. 1–9, 2014.
- [11] M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth, "Happiness and the patterns of life: A study of geolocated Tweets," *Scientific Reports*, pp. 1–9, 2013.
- [12] S. Sobolevsky, I. Sitko, S. Grauwin, R. Tachet des Combes, B. Hawelka, J. M. Arias, and C. Ratti, "Mining urban performance: Scale-independent classification of cities based on individual economic transactions," *Proceedings of the 2nd ASE International Conference on Big data Science and Computing*, pp. 1–10, 2014.
- [13] S. Sobolevsky, I. Sitko, R. Tachet des Combes, B. Hawelka, J. M. Arias, and C. Ratti, "Money on the move: Big data of bank card transactions as the new proxy for human mobility patterns and regional delineation. The case of residents and foreign visitors in Spain," in *Proceedings of the IEEE International Congress on Big Data*, 2014, pp. 136–143.
- [14] S. Sobolevsky, I. Sitko, R. T. d. Combes, B. Hawelka, J. M. Arias, and C. Ratti, "Cities through the prism of people's spending behavior," *arXiv preprint arXiv:1505.03854*, pp. 1–21, 2015.
- [15] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz, "Redrawing the map of Great Britain from a network of human interactions," *PLoS One*, vol. 5, no. 12, pp. 1–6, 2010.
- [16] S. Sobolevsky, M. Szell, R. Campari, T. Couronné, Z. Smoreda, and C. Ratti, "Delineating geographical regions with networks of human interactions in an extensive set of countries," *PLoS One*, vol. 8, no. 12, pp. 1–10, 2013.

- [17] T. Pei, S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou, "A new insight into land use classification based on aggregated mobile phone data," *International Journal of Geographical Information Science*, vol. 28, no. 9, pp. 1988–2007, 2014.
- [18] S. Grauwin, S. Sobolevsky, S. Moritz, I. Gódor, and C. Ratti, "Towards a comparative science of cities: Using mobile traffic records in New York, London and Hong Kong," in *Computational Approaches for Urban Environments*, ser. Geotechnologies and the Environment, 2015, vol. 13, pp. 363–387.
- [19] M. González, C. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, 2008.
- [20] K. Kung, K. Greco, S. Sobolevsky, and C. Ratti, "Exploring universal patterns in human home/work commuting from mobile phone data," *PLoS One*, vol. 9, no. 6, pp. 1–15, 2014.
- [21] S. Hoteit, S. Secci, S. Sobolevsky, C. Ratti, and G. Pujolle, "Estimating human trajectories and hotspots through mobile phone data," *Computer Networks*, vol. 64, pp. 296–307, 2014.
- [22] A. Amini, K. Kung, C. Kang, S. Sobolevsky, and C. Ratti, "The impact of social segregation on human mobility in developing and industrialized regions," *EPJ Data Science*, vol. 3, no. 1, pp. 1–20, 2014.
- [23] B. Hawelka, I. Sitko, E. Beinart, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, "Geo-located Twitter as proxy for global mobility pattern," *Cartography and Geographic Information Science*, vol. 41, no. 3, pp. 260–271, 2014.
- [24] S. Paldino, I. Bojic, S. Sobolevsky, C. Ratti, and M. C. González, "Urban magnetism through the lens of geo-tagged photography," *arXiv preprint arXiv: 1503.05502*, pp. 1–17, 2015.
- [25] J. Sinkienė and S. Kromalca, "Concept, directions and practice of city attractiveness improvement," *Public Policy and Administration*, vol. 31, no. 1, pp. 147–154, 2010.
- [26] A. W. Spirn, *The language of landscape*. Yale University Press, 1998.
- [27] F. Girardin, A. Vaccari, A. Gerber, A. Biderman, and C. Ratti, "Quantifying urban attractiveness from the distribution and density of digital footprints," *International Journal of Spatial Data Infrastructures Research*, vol. 4, pp. 175–200, 2009.
- [28] X. Y. Lehto, L. A. Cai, J. T. O'Leary, and T.-C. Huan, "Tourist shopping preferences and expenditure behaviours: The case of the Taiwanese outbound market," *Journal of Vacation Marketing*, vol. 10, no. 4, pp. 320–332, 2004.
- [29] J. Y.-J. Oh, C.-K. Cheng, X. Y. Lehto, and J. T. O'Leary, "Predictors of tourists' shopping behaviour: Examination of socio-demographic characteristics and trip typologies," *Journal of Vacation Marketing*, vol. 10, no. 4, pp. 308–319, 2004.
- [30] A. Yuksel, "Shopping experience evaluation: A case of domestic and international visitors," *Tourism Management*, vol. 25, no. 6, pp. 751–759, 2004.
- [31] M. S. Rosenbaum and D. L. Spears, "Who buys that? Who does what? Analysis of cross-cultural consumption behaviours among tourists in Hawaii," *Journal of Vacation Marketing*, vol. 11, no. 3, pp. 235–247, 2005.
- [32] M. Uncles, A. Ehrenberg, and K. Hammond, "Patterns of buyer behavior: Regularities, models, and extensions," *Marketing Science*, vol. 14, no. 3, pp. G71–G78, 1995.
- [33] Flickr dataset. [Online]. Available: <http://sfgeo.org/data/tourist-local>
- [34] Yahoo! Webscope dataset YFCC-100M. [Online]. Available: <http://labs.yahoo.com/Academic-Relations>
- [35] Twitter Streaming API. [Online]. Available: <https://dev.twitter.com/docs/streaming-apis>
- [36] M. Schläpfer, L. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West, and C. Ratti, "The scaling of human interactions with city size," *Journal of the Royal Society Interface*, vol. 11, no. 98, pp. 1–9, 2014.
- [37] L. M. Bettencourt, J. Lobo, D. Strumsky, and G. B. West, "Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities," *PLoS One*, vol. 5, no. 11, pp. 1–9, 2010.
- [38] L. M. Bettencourt, "The origins of scaling in cities," *Science*, vol. 340, no. 6139, pp. 1438–1441, 2013.
- [39] L. M. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, and G. B. West, "Growth, innovation, scaling, and the pace of life in cities," *Proceedings of the National Academy of Sciences*, vol. 104, no. 17, pp. 7301–7306, 2007.
- [40] M. Batty, "The size, scale, and shape of cities," *Science*, vol. 319, no. 5864, pp. 769–771, 2008.
- [41] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.
- [42] C. Roth, S. M. Kang, M. Batty, and M. Barthélemy, "Structure of urban movements: Polycentric activity and entangled hierarchical flows," *PLoS One*, vol. 6, no. 1, pp. 1–8, 2011.
- [43] City Statistics - Urban Audit. [Online]. Available: [http://epp.eurostat.ec.europa.eu/portal/page/portal/region\\_cities/city\\_urban](http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/city_urban)
- [44] ESPON project 1.4.3: Study on urban functions: Final report. [Online]. Available: <http://www.espon.eu/export/sites/default/Documents/Projects/ESPON2006Projects/StudiesScientificSupportProjects/UrbanFunctions/fr-1.4.3-April2007-final.pdf>
- [45] Afeas Urbanas de Españã. [Online]. Available: <http://alarcos.esi.uclm.es/per/fruiz/audes>
- [46] Eurostat. [Online]. Available: <http://ec.europa.eu/eurostat>
- [47] Instituto Nacional Estadística. [Online]. Available: [www.ine.es](http://www.ine.es)
- [48] G. K. Zipf, *Human behaviour and the principle of least-effort*. Addison-Wesley, 1949.