Urban Scaling of Football Followership on Twitter

Attila Sóti[1], Eszter Bokányi[2], Gábor Vattay[2]

[1]Doctoral School of Regional Sciences and Business Administration at Széchenyi István University, Győr, Hungary, attila.soti@complex.elte.hu

[2]Department of Physics of Complex Systems, Eötvös Loránd University, Budapest Hungary, ebokanyi@ complex.elte.hu; vattay@complex.elte.hu

Abstract: Social sciences have an important challenge today to take advantage of new research opportunities provided by large amounts of data generated by online social networks. Because of its marketing value, sports clubs are also motivated in creating and maintaining a stable audience in social media. In this paper, we analyze followers of prominent footballs clubs on Twitter by obtaining their home locations. We then measure how city size is connected to the number of followers using the theory of urban scaling. The results show that the scaling exponents of club followers depend on the income of a country. These findings could be used to understand the structure and potential growth areas of global football audiences.

Keywords: urban scaling; Twitter; social media; football

1   Introduction

Today the online social network Twitter has more than 300 million monthly active users {Twitter2018}, with many of them actively following sports events, stars or clubs to exploit the possibilities of obtaining the latest news through instantaneous messaging {Bruns2014}. Large football clubs and football leagues invest money in establishing official social media channels to engage with their fan basis {Price2013} and seek to purchase players who bring them a massive number of Twitter followers {KpmgRonaldo}. Social media presence is especially important for clubs that rely more heavily on broadcasting and commercial revenues than on match day revenues, such as the global top 20 clubs from a recent analysis of the Deloitte Football Money League {Deloitte2018}. Because global fans have limited options to be present at matchday events, popularity on Facebook together with Twitter is a good indicator to judge the global follower success of a football club.

On the other hand, the geographic and socio-economic environment of a user still plays an important role in determining the probability of engaging with a globalized phenomenon. As such, complex spatial structures and the dynamics of changes in them have for some time been a focus of the scientific community as well as marketing experts. Recently, there has been a growing literature on the concept of urban scaling, which connects measurable outputs of cities to their size {Bettencourt2010d, Alves2015c, Arcaute2015a, Cottineau2017, Cottineau2018 DefiningEconomies, Bettencourt2013a, Bettencourt2013d, Gomez-Lievano2012,Yakubo2014SuperlinearCities}. Urban scaling laws have been detected for various quantities with respect to city size, such as GDP {Lobo2013}, urban economic diversification {Strumsky2016}, touristic attractiveness {Bojic2016ScalingStates}, crime concentration {Oliveira2017, Hanley2016a}, human interactions {Schlapfer2014b}, election data {Bokanyi2018UniversalResults} or even building heights {Schlapfer2015UrbanSize}. Some of these measures follow a super linear relationship with urban size, which means that the quantities are dis-proportionally over-represented in larger cities. These measures include GDP, number of patents or certain business types, where larger cities facilitate more the accumulation of wealth and resources needed for such phenomena. On the other hand, infrastructural-like quantities have sublinear scaling laws reflecting efficiency due to urban agglomeration effects.

In this paper, we investigate urban scaling laws for geolocated Twitter football club followers for three majors widely acknowledged clubs: Real Madrid, Manchester United and Bayern Munich. We calculate the scaling exponents for the number of followers of each club in the urban systems of five different countries. While the scaling exponents of clubs differ significantly within countries as well, the variations in the exponents across countries suggest that the wealthier a country is, the more sublinear its follower scaling exponent, and vice versa.

2   Materials and Methods

Twitter freely provides approximately 1\% of its data for download through its API. For those users that allow this option on their smartphones, the exact GPS coordinates are attached to their messages, the so-called tweets. By focusing the data collection on these geolocated tweets, we could determine the home location for a selected most active users in the database using the friend-of-friend algorithm clustering on their coordinated messages. This left us with a total of 26.3 million Twitter users that have home coordinates associated to them. We constructed a geographically indexed database of these users, permitting the efficient analysis of regional features {Dobos2013}. Using the Hierarchical

Triangular Mesh scheme for practical geographic indexing, we assigned cities to each user. City locations were obtained from http://geonames.org, city bounding boxes via the Google Places API. We downloaded the Twitter user identifiers of the followers of three selected football clubs: Real Madrid, Manchester United and Bayern Munich. Table 1 shows the number of followers (people who follow at least one of the three teams, later referred as overall follower count) that are also in our geolocated user database, which meant roughly 2-3\% of all followers in all three cases.

Table 1

Number of total followers for each football club on Twitter and number of followers from the geolocated user database used in our analysis

Team name	Total number of followers	Geolocated followers
Real Madrid	28.7M	808,427
Manchester United	17.3M	436,515
Bayern Munich	4.3M	119,056

The theory of urban scaling {Alves2015c} suggests that there is a power-law relationship between a socio-economic indicator measured in a city and its size. We can formulate this power-law relationship with the following equation:

$$Y = Y_0 * N^\beta \tag{1}$$

Where Y denotes the investigated quantity, N is the number of inhabitants in a city, Y_0 is a normalization constant, and ß is the so-called exponent that characterizes the behavior of the quantity in connection to changing city size. In the literature, it has been observed, that this ß parameter differs only slightly from 1. Most urban socio-economic indicators have super linear ß > 1 exponents, which is caused by larger cities being the centers of wealth, innovation and creative processes. Sublinear scaling ß < 1 characterizes material quantities associated with infrastructure, where the agglomeration into cities is more economic, which manifests in fewer overall road length, or overall cable need, etc. {Bettencourt2010d}.

If we take the logarithm of both sides, the equation becomes a linear relationship:

$$\log Y = \log \mathbf{Y_0} + \beta * \log N \tag{2}$$

It is then enough to fit a line onto the log Y -- log N pairs. We used binning of the data, where we took the mean of log N and log Y in each bin, and then fitted a line onto them using an OLS fit with weighting the bins by $1 / \sqrt{N}$. This error calculation assumes that higher follower numbers carry less error when fitting the scaling curves {Bettencourt2007}.

3   Results and Discussion

The geographical distribution of users that follow at least one of the three clubs can be seen in Figure 1.
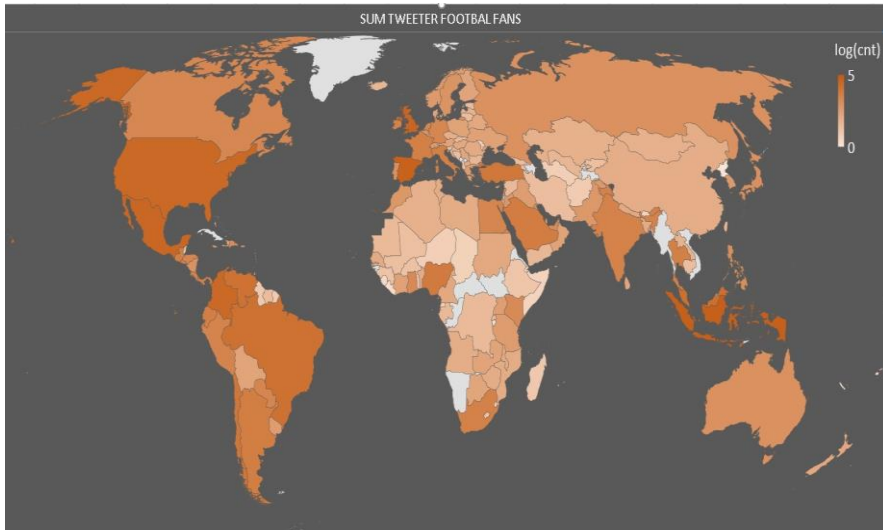

Figure 1

Distribution of geolocated Twitter users that follow at least one of the three selected clubs. Countries are colored according to the logarithm of the number of users.

A major fan base is in Western Europe, North and Latin America as well as in the Pacific Region. Because Spanish and English teams are among the investigated clubs in Spain and in Great Britain the number of followers is high.

As analyzed countries, we chose the home countries of two of the teams, Spain and the UK, and we included traditional football supporter countries like Mexico. We chose Indonesia from the Pacific Region, and Columbia from South America. We also analyze the USA since it is a country with high Twitter penetration.
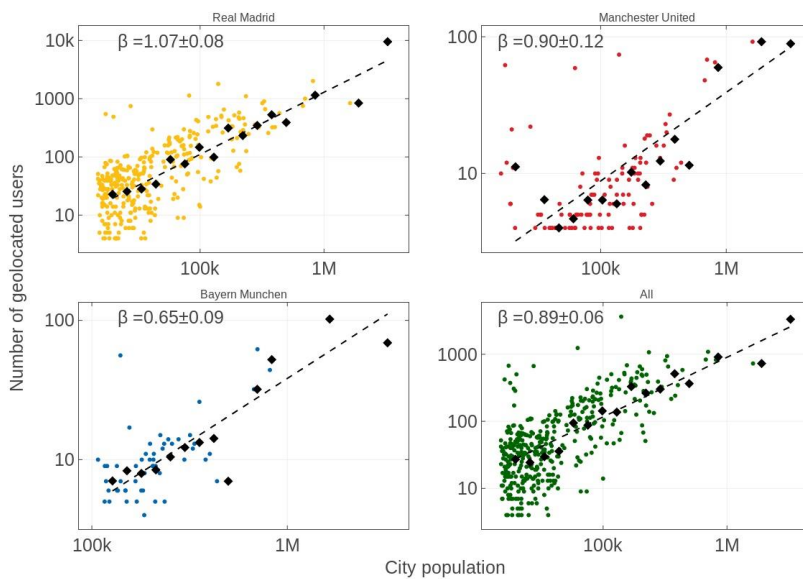
Figure 2

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Spain. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents ß$_{RM}$=1.07±0.08; ß$_{MU}$= 0.90±0.12; ß$_{BM}$ = 0.65±0.09 and ß$_{All}$ = 0.89\±0.06, respectively.

In the top left corner of Figure 2, we can see the urban scaling relationships of Spain for the three clubs (Real Madrid in the top left, Manchester United in the top right and Bayern Munich in the bottom left corner), and for the number of overall followers (bottom right corner). The exponent of Real Madrid, the "home" team is super linear ß$_{RM}$ = 1.07±0.08 while the exponent of the other two teams are sublinear with ß$_{MU}$ = 0.90 ± 0.12 for the Manchester United, and ß$_{BM}$ = 0.65 ± 0.09 for the Bayern Munich, respectively. It is spectacular how the second biggest city in Spain, Barcelona is a clear outlier in the Real Madrid urban scaling curve, with having much less followers than the size of the city would predict. The overall follower numbers in Spain also has a sublinear scaling.
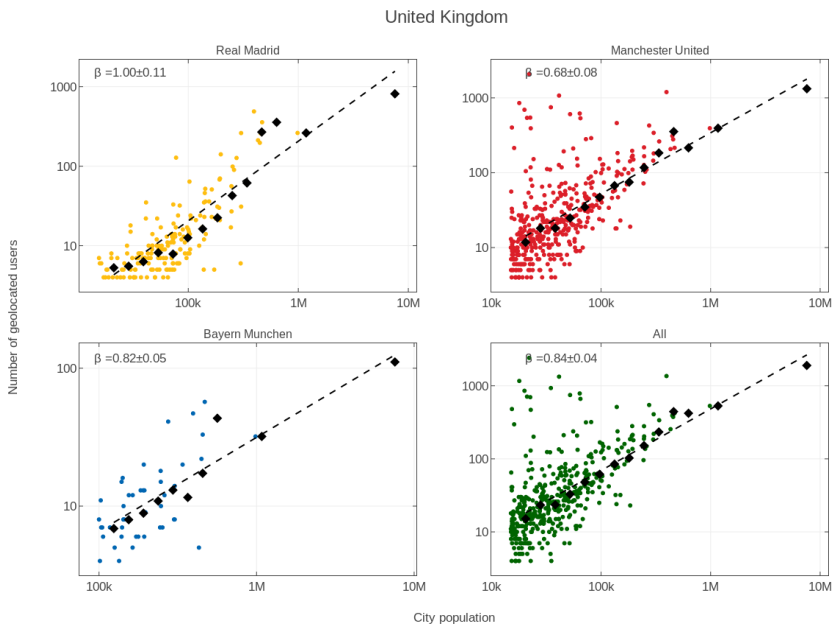
Figure 3

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in the UK. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents β_{RM}=1.00±0.11, β_{MU}= 0.68±0. 08,; β_{BM} = 0.82±0.05 and β_{All} = 0.84±0.04, respectively.

In Figure 3 when we look at scaling curves in the UK, which has the longest football traditions of all of the other countries, we again see a similar picture of the exponents, with that of Real Madrid being higher than the other two, though it is only around the linear regime with β_{RM} =1.00 ± 0.11. However, Manchester United, apart from the outlier points of Manchester and its surroundings has an astoundingly low sublinear exponent β_{MU} = 0.68 ± 0.08 that suggests a strong relative decline of interest for this team with the city size. The overall follower trend is also strongly sublinear in the UK.

The case of the USA in Figure 4 is very similar to that of Spain, where Real Madrid followers scale super linearly, but the other two clubs have a sublinear relationship with city size.
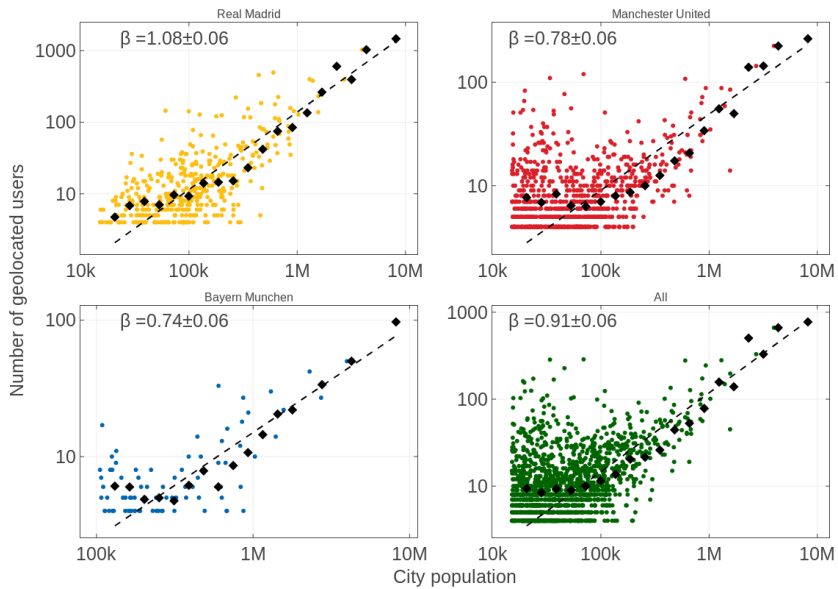

Figure 4
Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in the US. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $\beta_{RM} = 1.08\pm0.06$, $\beta_{MU} = 0.78\pm0.06$, $\beta_{BM} = 0.74\pm0.06$ and $\beta_{All} = 0.91\pm0.06$, respectively.

A very different effect takes place in Indonesia according to Figure 5. Here, all four scaling relationships are in the highly super linear range, which means that club followership is a measure that is driven by urban factors.
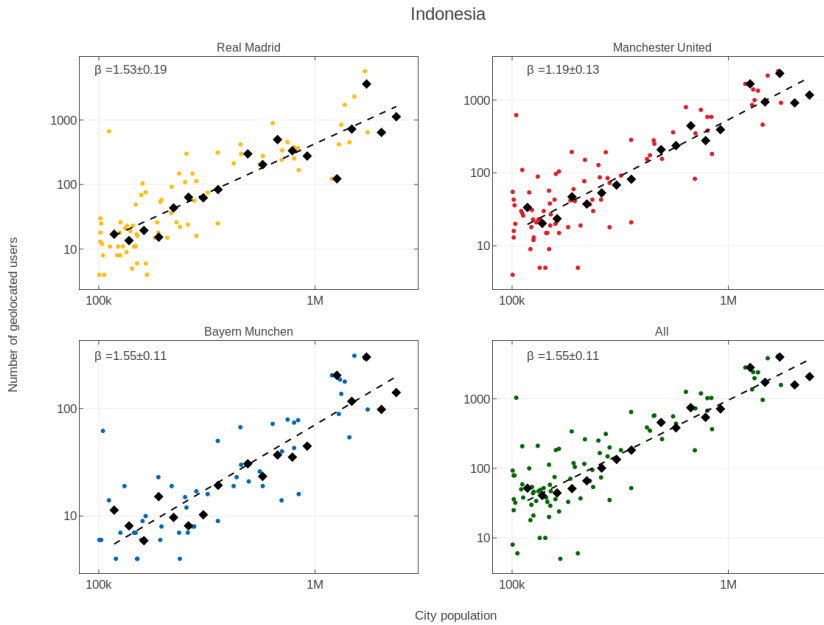
Figure 5

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Indonesia. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $\beta_{RM} = 1.53 \pm 0.19$, $\beta_{RM} = 1.19 \pm 0.13$, $\beta_{BM} = 1.55 \pm 0.11$ and $\beta_{All} = 1.55 \pm 0.11$, respectively.

Though less pronounced because of slightly smaller, but still super linear exponents, this is also the case for Columbia in Figure 6.
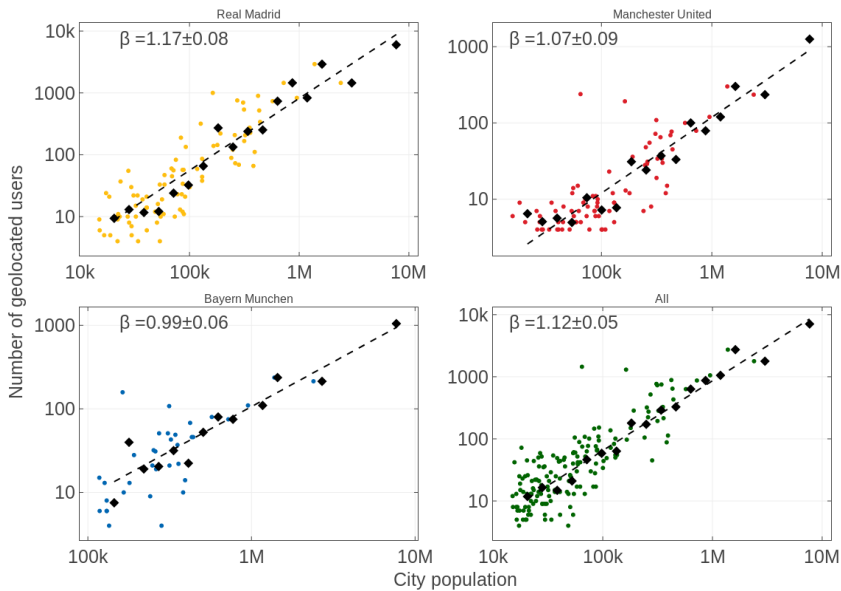
Figure 6

Number of followers for the three selected clubs (A-C), and combined follower number (D) as a function of city size in Columbia. Black diamonds correspond to bin averages, dashed lines represent the OLS fits with exponents $\text{ß}_{RM} = 1.17\pm0.08$, $\text{ß}_{MU}= 1.07\pm0.09$, $\text{ß}_{BM} = 0.99\pm0.06$ and $\text{ß}_{All} =1.12\pm0.05$, respectively.

Conclusions

The summary Figure 7 shows that Columbia, Indonesia and Mexico, are the countries whose exponents for the overall supporter count are super linear. This means that in these countries that globalized football tracking is an increasingly urban phenomenon. In countries where football culture is older, and/or general income is higher, sublinear exponents may signal a relative attention shift for football to smaller settlements, and a change in the composition of consumers of football-related content. This may be an important message for marketers trying to increase social media attention and responsiveness because people from different environments may need quite different targeting messages.
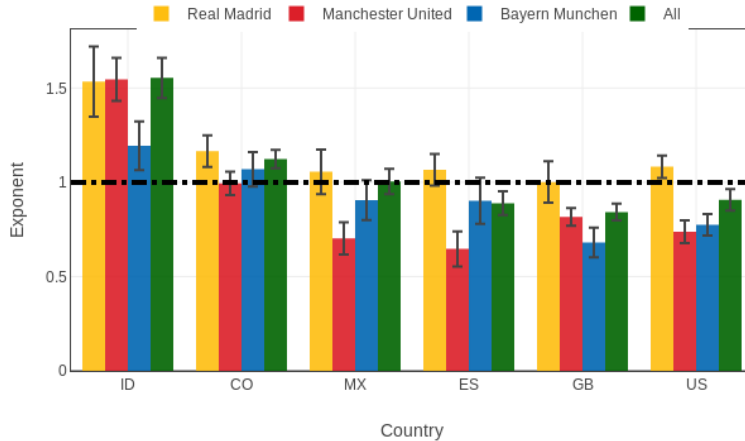
Figure 7

Summary Figure showing the exponents per team according to countries. The vertical line at ß=1 corresponds to linear scaling.

The difference between the club exponents in the same country suggests that Real Madrid followers are relatively more abundant in bigger cities, and the other two teams have the same exponents. This suggests that even within a country, different clubs may have different follower audiences and fans from different cultural backgrounds.

In this paper, we analyzed urban scaling in the follower numbers of three football clubs, Real Madrid, Manchester United and Bayern Munich. We determined user geolocation from Twitter messages that had GPS coordinates attached to them and fitted scaling relationships using population data for cities of six different countries. While for higher-income countries, urban scaling exponents tended to be in the sublinear, linear or in a few cases, a slightly super linear range, exponents for lower-income countries are almost exclusively super linear. This suggests that in a globalized football fandom, very different factors drive followership. Exponents also exhibited variations between clubs, which suggests that the followers of different football clubs are embedded into different socio-economical environments that are related to the degree of urbanization as well.

Acknowledgment

References

[1]   Twitter, "Selected Company Metrics and Financials Monthly Active Users," Tech. Rep., 2018

[2]   Bruns, K. Weller, and S. Harrington, "Twitter and Sports," in Twitter and Society. Peter Lang, New York, 2014, pp. 263-280

[3]   J. Price, N. Farrington, and L. Hall, "Changing the Game? The Impact of Twitter on Relationships between Football Clubs, Supporters and the Sports Media," Soccer and Society, Vol. 14, No. 4, pp. 446-461, 2013

[4]   "KPMG Football Benchmark Ronaldo Economics," Tech. Rep., 2018

[5]   "Deloitte Football Money League 2018," Tech. Rep., 2018

[6]   L. Bettencourt and G. West, "A Unified Theory of Urban Living," Nature, Vol. 467, No. 7318, pp. 912-913, 102010

[7]   L. G. A. Alves, R. S. Mendes, E. K. Lenzi, and H. V. Ribeiro, "Scale-Adjusted Metrics for Predicting the Evolution of Urban Indicators and Quantifying the Performance of Cities," PLOS ONE, Vol. 10, No. 9, p. e0134862, 9 2015

[8]   E. Arcaute, E. Hatna, P. Ferguson, H. Youn, A. Johansson, and M. Batty, "Constructing Cities, Deconstructing Scaling Laws," Journal of The Royal Society Interface, Vol. 12, No. i, pp. 3-6, 2015

[9]   C. Cottineau, E. Hatna, E. Arcaute, and M. Batty, "Diverse Cities or the Systematic Paradox of Urban Scaling Laws," Computers, Environment and Urban Systems, Vol. 63, No. July, pp. 80-94, May 2017

[10]   C. Cottineau, O. Finance, E. Hatna, E. Arcaute, and M. Batty, "Defining Urban Clusters to Detect Agglomeration Economies," 2018

[11]   L. M. A. Bettencourt, "The Origins of Scaling in Cities." Science, Vol. 340, No. 6139, pp. 1438-1441, 6 2013

[12]   L. M. A. Bettencourt, J. Lobo, and H. Youn, "The Hypothesis of Urban Scaling: Formalization, Implications and Challenges," SFI Working Paper, Vol. 2013-01-00, p. 37, 1 2013

[13]   Gomez-Lievano, H. Youn, and L. M. A. Bettencourt, "The Statistics of Urban Scaling and Their Connection to Zipf's Law," PLoS ONE, Vol. 7, No. 7, p. e40393, jul 2012

[14]   K. Yakubo, Y. Saijo, and D. Korosak, "Superlinear and Sublinear Urban Scaling in Geographical Networks Modeling Cities," Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, Vol. 90, No. 2, pp. 1-10, 2014

[15]   J. Lobo, L. M. A. Bettencourt, D. Strumsky, and G. B. West, "Urban Scaling and the Production Function for Cities," PLoS ONE, Vol. 8, No. 3, p. e58407, 3 2013

[16]   H. Youn, L. M. A. Bettencourt, J. Lobo, D. Strumsky, H. Samaniego, and G. B. West, "Scaling and Universality in Urban Economic Diversification," Journal of The Royal Society Interface, Vol. 13, No. 114, p. 20150937, Jan 2016

I.   Bojic, A. Belyi, C. Ratti, and S. Sobolevsky, "Scaling offoreignattractivenessforcountriesandstates," Applied Geography, Vol. 73, pp. 47-52, 2016

[17]   M. Oliveira, C. Bastos-Filho, and R. Menezes, "The Scaling of Crime Concentration in Cities," PLoS ONE, Vol. 12, No. 8, 2017

[18]   Q. S. Hanley, D. Lewis, and H. V. Ribeiro, "Rural to Urban Population Density Scaling of Crime and Property Transactions in English and Welsh Parliamentary Constituencies," PLoS ONE, Vol. 11, No. 2, pp. 25-27, 2016

[19]   M. Schläpfer, L. M. A. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West, C. Ratti, M. A. Bettencourt, and M. Schla, "The Scaling of Human Interactions with City Size," Journal of the Royal Society, Interface / the Royal Society, Vol. 11, No. July, pp. 20130789–, 2014

[20]   E. Bokányi, Z. Szüllási, and G. Vattay, "Universal Scaling Laws in Metro Area Election Results," PLoS ONE, Vol. 13, No. 2, 2018

[21]   M. Schläpfer, J. Lee, and L. M. A. Bettencourt, "Urban Skylines: Building Heights and Shapes as Measures of City Size," pp. 1-17, 12 2015

[22]   L. Dobos, J. Szüle, T. Bodnár, T. Hanyecz, T. Sebők, D. Kondor, Z. Kallus, J. Stéger, I. Csabai, and G. Vattay, "A Multi-Terabyte Relational Database for Geo-tagged Social Network Data," in 4th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2013 Proceedings, 2013, pp. 289-294

[23]   L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kuhnert, 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, Apr 2007