Supplementary Material: Human Diffusion and City Influence

Maxime Lenormand*[1], Bruno Gonçalves[2,3], Antònia Tugores[1], and José Javier Ramasco[1]

[1]IFISC, Campus Universitat de les Illes Baleares, 07122 Palma de Mallorca, Spain
[2]Aix Marseille Université, Université de Toulon, CNRS, CPT, UMR 7332, 13288 Marseille, France
[3]Université de Toulon, CNRS, CPT, UMR 7332, 83957 La Garde, France

As a first characterization of the data, we have computed the great circle distance Δ_r between successive positions of the same Twitter user living in one of the 58 cities (Figure S1). The distribution $P(\Delta_r)$ for each city is well approximated by a power law with an average exponent value of 1.5. These results are consistent with the exponent obtained in other studies [1, 2, 3]. It is interesting to note that the distributions are very similar for all the cities.

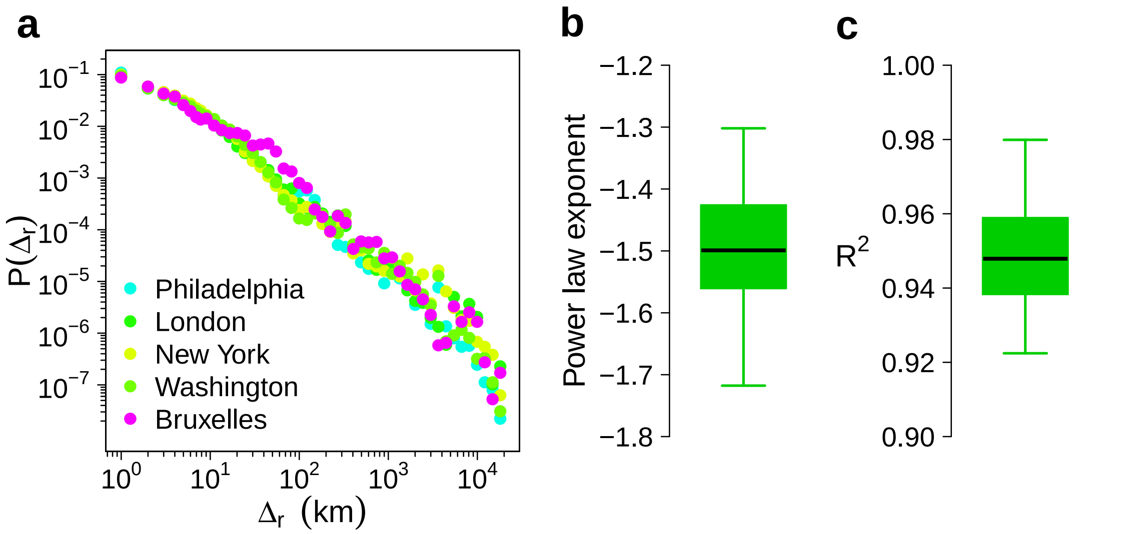

Figure S1: **Probablity density function of distance travelled by the local Twitter users.** (a) Probablity density function $P(\Delta_r)$ of the distance travelled by the local Twitter users for 5 cities drawn at random among the 58 case studies. Δ_r is the great circle distance between each successive position of the local Twitter users. (b) Boxplot of the 58 power-law exponent. (c) Boxplot of the R^2. The boxplot is composed of the minimum value, the lower hinge, the median, the upper hinge and the maximum value.

*maxime@ifisc.uib.es

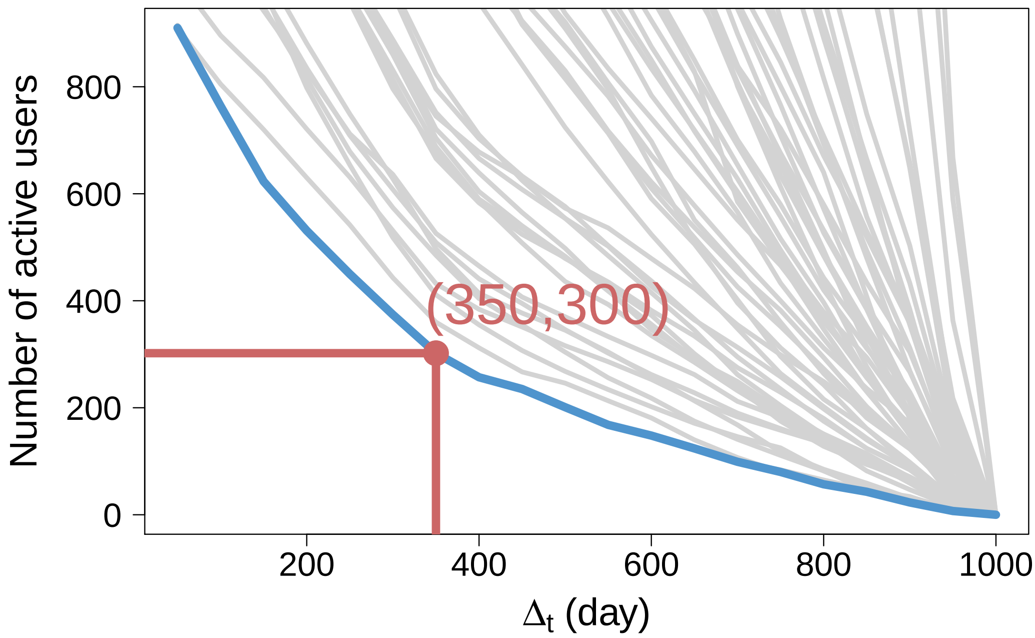

Figure S2: **Minimum number of active users as a function of Δ_t (blue line).** The gray lines represent the number of active users as a function of Δ_t for the 58 cities.

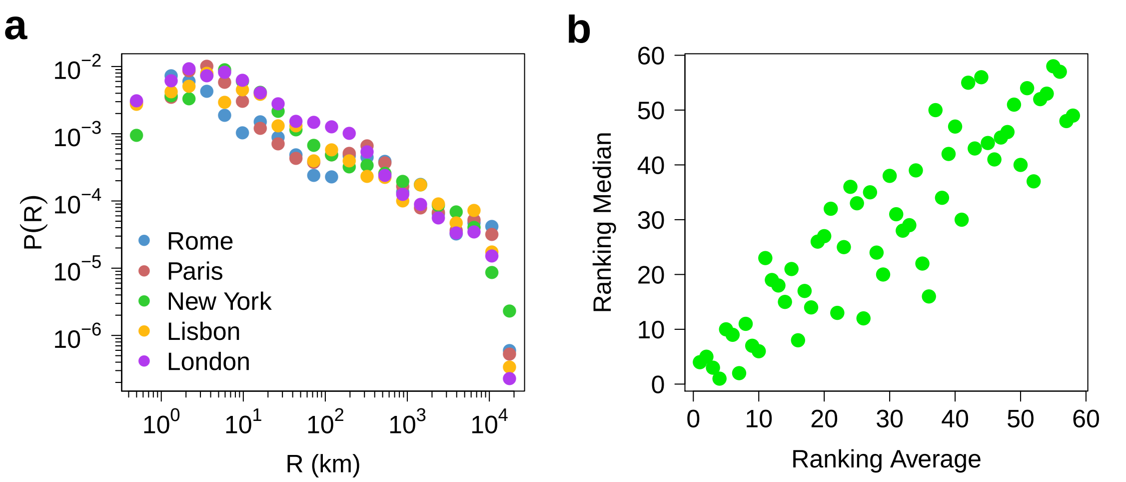

Figure S3: **Radius.** (a) Probablity density function of the radius per Twitter users for 5 cities. (b) Ranking by median radius as a function of the ranking by average radius. The rankings are based on an average of the two statistics over 100 independent extractions of a set of $u = 300$ users.

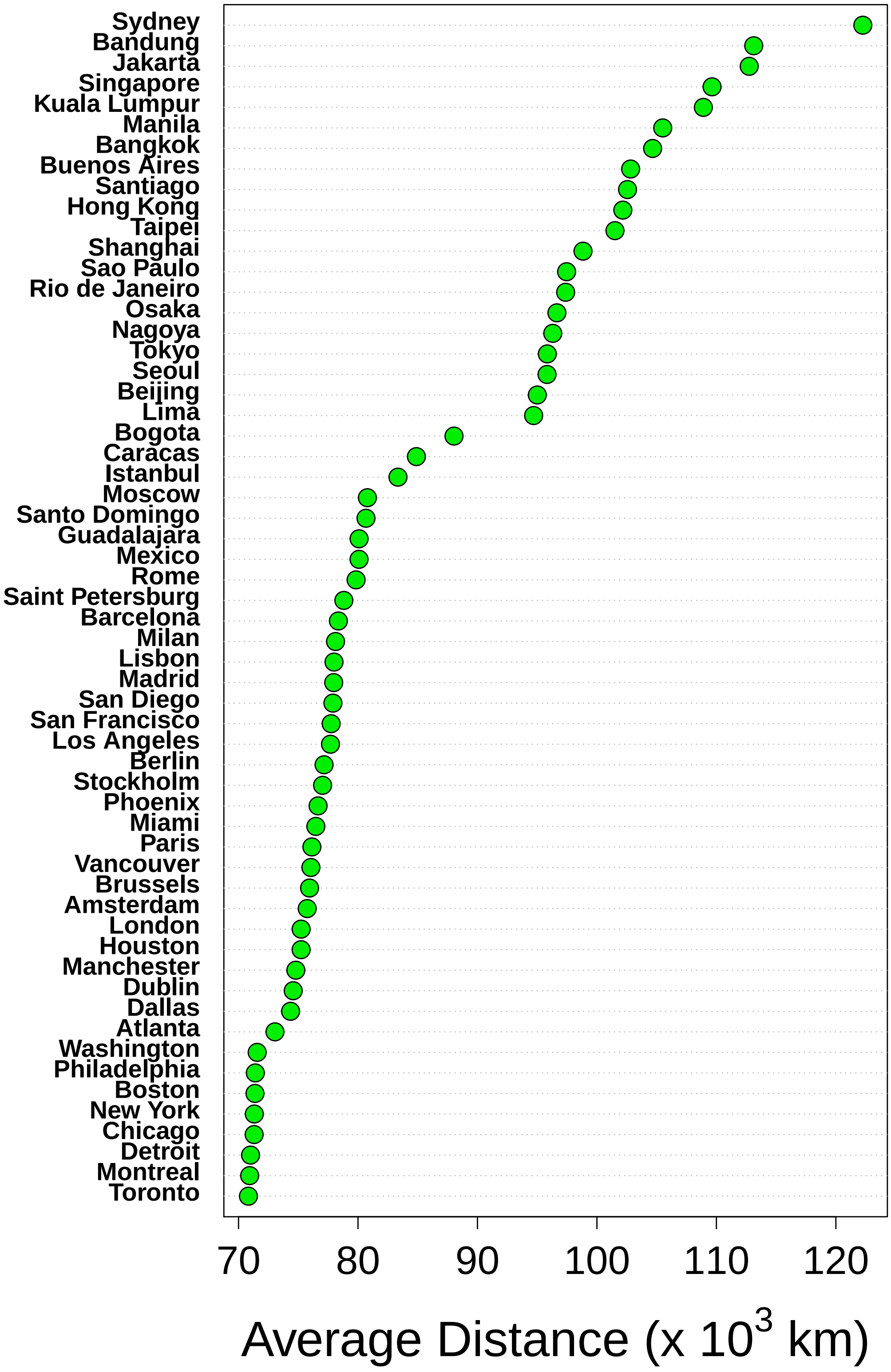

Figure S4: Ranking of the cities according the the average distance between the center of the city and all the Twitter users' place of residence (represented by the centroid of the cell of residence).

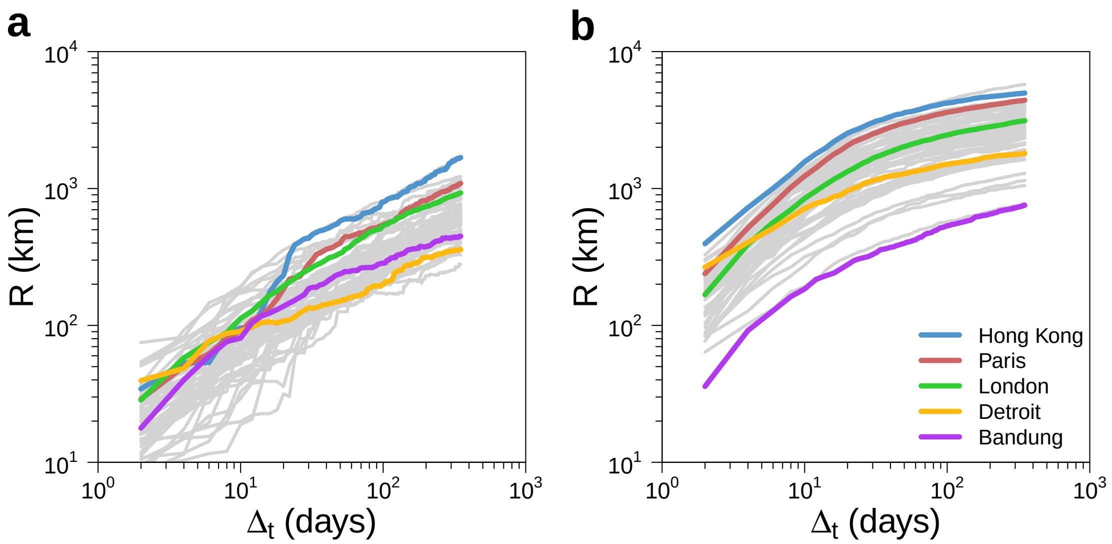

Figure S5: **Evolution of the average radius for the local users (a) and for the non-local users (b).** Each curve represents the evolution of the average radius R averaged over 100 independent extractions of a set of $u = 100$ users as a function of the number of days Δ_t since the first passage in the city. In order to show the general trend, each gray curve corresponds to a city. The evolution of the radius for several cities is highlighted, such as the top and bottom rankers or representatives of the two main detected behaviors.

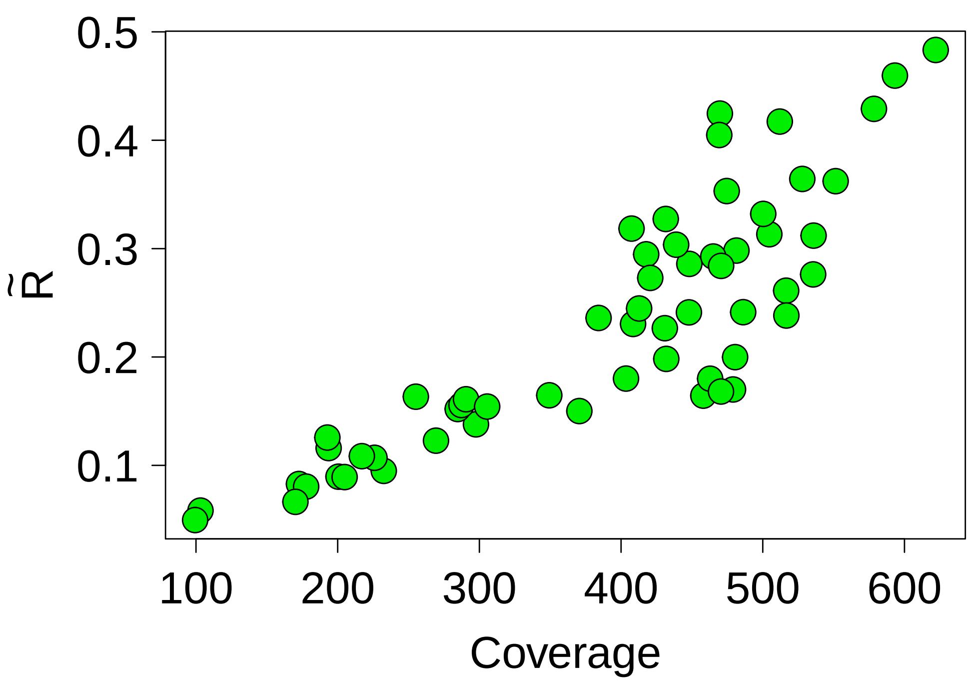

Figure S6: **Coverage as a function of $\tilde{R}$ for the 58 cities.** A certain level of correlation can be observed between both metrics. Both metrics are averaged over 100 independent extractions of a set of $u = 300$ users.

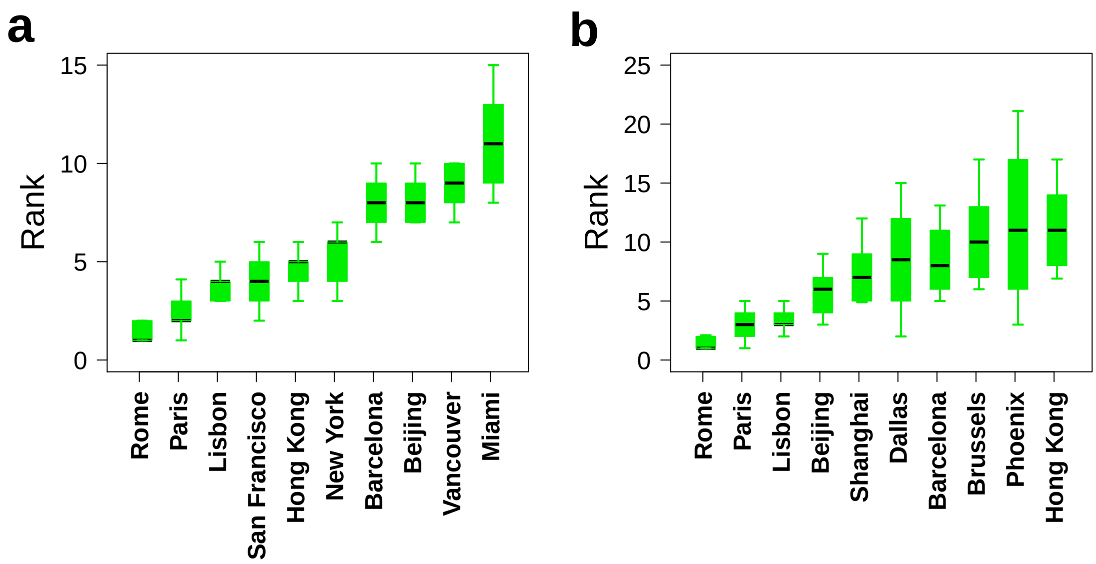

Figure S7: **Variations of the rankings over 100 realizations.** (a) Ranking for the normalized average radius. (b) Ranking for the coverage. The boxplot is composed of the minimum value, the lower hinge, the median, the upper hinge and the maximum value. The rankings are averaged over 100 independent extractions of a set of $u = 300$ users.

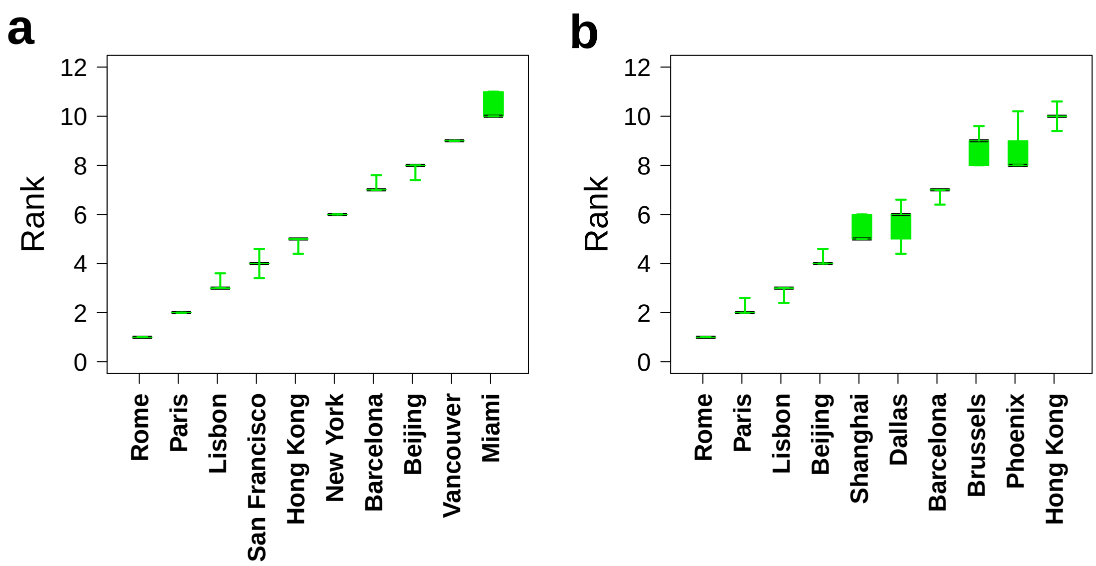

Figure S8: **Variations of the rankings over 10 realizations performed on the average over 10 realizations.** (a) Ranking for the normalized average radius. (b) Ranking for the coverage. The boxplot is composed of the minimum value, the lower hinge, the median, the upper hinge and the maximum value. The rankings are averaged over 100 independent extractions of a set of $u = 300$ users.

Entropy index

The natural way of taking the heterogeneity of visiting frequencies into consideration is to introduce an entropy measure. If we define the probability p_i^t than an individual tweet originating from the users we are considering originated in a cell i, then the entropy S for a given time interval Δt is given by:

$$S(t) = -\frac{\sum_{i=1}^{N} p_i^t \log(p_i^t)}{\log(\mathcal{N}(t))} \tag{1}$$

where the normalizing factor $\mathcal{N}(t)$, the number of cells with non-zero number of tweets, corresponds to the uniform case where each tweet has the same probability of being produced within each cell. With this normalization, the entropy is defined to vary just between 0 and 1, regardless of the number of cells and tweets we might consider in each case.

The entropy as a function of the number of visited cells is plotted in Figure 7a. The entropy enhances with the number of visited cells despite the normalization, which implies that the tweets tend to distribute more uniformly for those cities with larger areas covered and therefore with a larger global projection. Besides the general trend, there are some interesting outliers such as Moscow and Saint Petersburg, with a high area covered given the size of Russia but low entropy meaning that the travels concentrate toward a few cells (likely the cities in a vast territory). On the other extreme, we find Osaka and Nagoya with a low are covered but high entropy. A possible reason is that the travels can be mostly within Japan but since the population in the country is well distributed, the trip destinations are well mixed.

As can be seen in Figure 7b, the entropy measured in the cities based only in local users is way lower than for the non-locals. This means that the locals move toward more concentrated locations, in contrast to the comparatively higher diversity of origins of the non-local visitors.

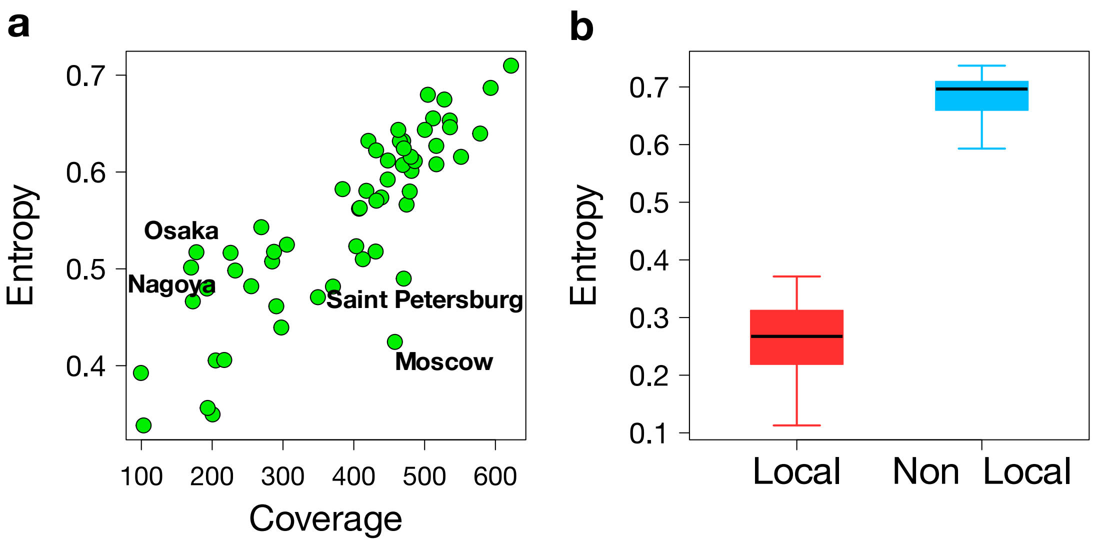

Figure S9: **Entropy index according to the Twitter user type.** (a) Entropy index as a function of the number of cells visited by $u = 300$ Twitter users drawn at random. (b) Box plot with the entropy measured for the different cities separating the users as locals and non-locals. The number of users is $u = 100$ in this case.

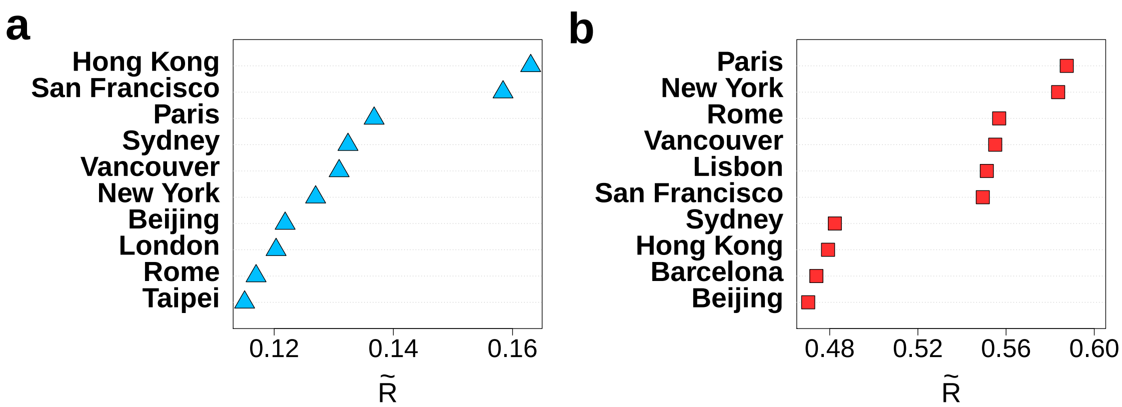

Figure S10: **Relation between local and non-local users.** (a) Top 10 ranking cities based only on local users according to the average radius. (b) Top 10 ranking cities based only on non-local users according to the average radius. In all the cases, the number of local and non-local users extracted is $u = 100$ for every city and all the metrics are averaged over 100 independent extractions.

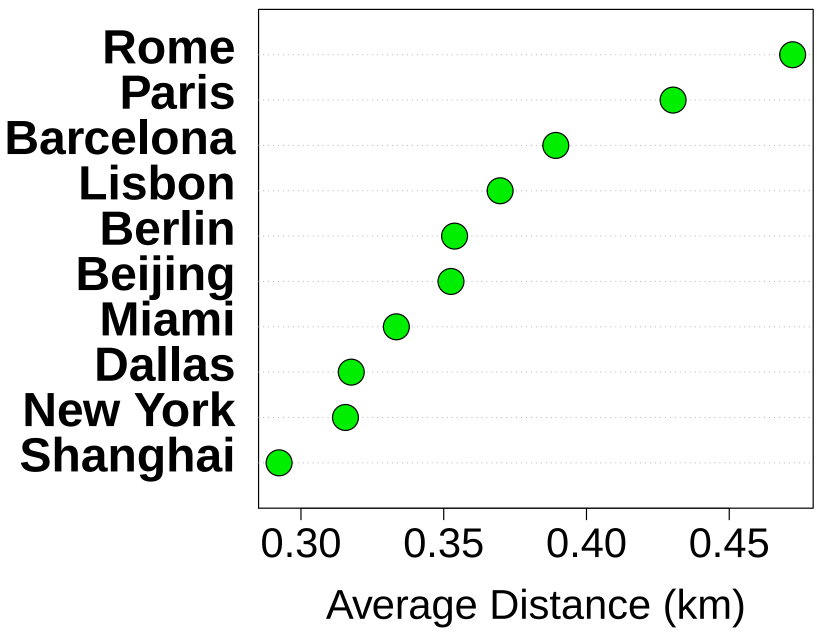

Figure S11: **City attractiveness.** Top 10 cities ranked by the average distance between the Twitter users' residences (represented by the centroid of the cell of residence) and the city center for $u = 1000$ Twitter users drawn at random. The metric is averaged over 100 independent extractions.

Table S1: **Description of the case studies**

City	Number of users	Number of Tweets	Number of Tweets per user
Amsterdam	2661	305363	114.75
Atlanta	2863	296390	103.52
Bandung	5620	405241	72.11
Bangkok	2604	239514	91.98
Barcelona	1713	165934	96.87
Beijing	1299	131922	101.56
Berlin	678	45238	66.72
Bogota	2226	213739	96.02
Boston	752	73561	97.82
Brussels	1243	97688	78.59
Buenos Aires	411	28500	69.34
Caracas	3625	375933	103.71
Chicago	2191	257572	117.56
Dallas	1214	128834	106.12
Detroit	13608	938524	68.97
Dublin	704	78434	111.41
Guadalajara	721	57031	79.10
Hong Kong	1098	108203	98.55
Houston	1582	186830	118.10
Istanbul	1321	103117	78.06
Jakarta	1919	196188	102.23
Kuala Lumpur	509	42665	83.82
Lima	360	42186	117.18
Lisbon	6782	698998	103.07
London	6392	580084	90.75
Los Angeles	1760	159781	90.78
Madrid	1566	202650	129.41
Manchester	1792	163090	91.01
Manila	4118	293015	71.15
Mexico	2534	247486	97.67
Miami	688	84544	122.88
Milan	666	61175	91.85
Montreal	1239	133461	107.72
Moscow	2334	263132	112.74
Nagoya	9668	892442	92.31
New York	4044	398769	98.61
Osaka	2567	247449	96.40
Paris	432	43301	100.23
Philadelphia	2206	247159	112.04
Phoenix	1380	150468	109.03
Rio de Janeiro	3292	352777	107.16
Rome	824	88402	107.28
Saint Petersburg	497	51601	103.82
San Diego	1810	182035	100.57
San Francisco	4628	419032	90.54
Santiago	2471	250639	101.43
Santo Domingo	302	20245	67.04
Sao Paulo	6479	653909	100.93
Seoul	1898	152666	80.44
Shanghai	526	49282	93.69
Singapore	3501	288267	82.34
Stockholm	745	106366	142.77
Sydney	1176	121426	103.25
Taipei	485	40259	83.01
Tokyo	10333	844602	81.74
Toronto	1476	135914	92.08
Vancouver	796	70018	87.96
Washington	3755	421374	112.22

Table S2: **Comparison of the regional and the global betweenness rankings.**

Community	Global Ranking	Regional Ranking
North America	1. New York (1)	1. New York
	2. Miami (6)	2. Los Angeles
	3. San Francisco (8)	3. Chicago
	4. Los Angeles (9)	4. Toronto
	5. Chicago (18)	5. Detroit
	6. Toronto (19)	6. Miami
	7. San Diego (23)	7. Dallas
	8. Detroit (25)	8. San Francisco
	9. Montreal (26)	9. Washington
	10. Atlanta (27)	10. Atlanta
	11. Washington (29)	11. Phoenix
	12. Vancouver (35)	12. Vancouver
	13. Dallas (36)	13. Montreal
	14. Phoenix (46)	14. Boston
	15. Boston (47)	15. Houston
	16. Houston (48)	16. San Diego
	17. Philadelphia (50)	17. Philadelphia
	18. Santo Domingo (58)	18. Santo Domingo
Europe	1. London (2)	1. London
	2. Paris (3)	2. Paris
	3. Madrid (10)	3. Moscow
	4. Barcelona (11)	4. Barcelona
	5. Moscow (16)	5. Berlin
	6. Berlin (20)	6. Rome
	7. Rome (21)	7. Madrid
	8. Amsterdam (24)	8. Lisbon
	9. Lisbon (38)	9. Amsterdam
	10. Milan (40)	10. Saint Petersburg
	11. Brussels (41)	11. Dublin
	12. Istanbul (42)	12. Istanbul
	13. Saint Petersburg (45)	13. Manchester
	14. Dublin (49)	14. Brussels
	15. Manchester (51)	15. Milan
	16. Stockholm (57)	16. Stockholm
Asia	1. Singapore (5)	1. Singapore
	2. Hong Kong (7)	2. Hong Kong
	3. Taipei (13)	3. Jakarta
	4. Jakarta (15)	4. Bangkok
	5. Kuala Lumpur (22)	5. Shanghai
	6. Seoul (30)	6. Taipei
	7. Bangkok (31)	7. Sydney
	8. Shanghai (32)	8. Kuala Lumpur
	9. Beijing (33)	9. Seoul
	10. Sydney (34)	10. Manila
	11. Manila (43)	11. Bandung
	12. Bandung (56)	12. Beijing
South America	1. Buenos Aires (12)	1. Buenos Aires
	2. Sao Paulo (14)	2. Sao Paulo
	3. Bogota (28)	3. Bogota
	4. Santiago (37)	4. Rio de Janeiro
	5. Rio de Janeiro (39)	5. Santiago
	6. Lima (44)	6. Caracas
	7. Caracas (55)	7. Lima
Japan	1. Tokyo (4)	1. Tokyo
	2. Osaka (53)	2. Osaka
	3. Nagoya (54)	3. Nagoya
Mexico	1. Mexico (17)	1. Guadalajara
	2. Guadalajara (52)	2. Mexico

References

[1] Brockmann D, Hufnagel L, Geisel T. 2006 The scaling laws of human travel. *Nature* **439**, 462–465. (doi:10.1038/nature04292)

[2] Gonzalez MC, Hidalgo CA, Barabasi A-L. 2008 Understanding individual human mobility patterns. *Nature* **453**, 779–782. (doi:10.1038/nature06958)

[3] Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C. 2014 Geo-located twitter as a proxy for global mobility patterns. *Cartography and Geographic Information Science* **41**, 260–271. (doi:10.1080/15230406.2014.890072)