



HAL
open science

Human diffusion and city influence

Maxime Lenormand, Bruno Goncalves, Antònia Tugores, José Javier Ramasco

► **To cite this version:**

Maxime Lenormand, Bruno Goncalves, Antònia Tugores, José Javier Ramasco. Human diffusion and city influence. 2015. hal-01111981v1

HAL Id: hal-01111981

<https://hal.science/hal-01111981v1>

Preprint submitted on 2 Feb 2015 (v1), last revised 17 Jul 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Human diffusion and city influence

Maxime Lenormand,¹ Bruno Gonçalves,² Antònia Tugores,¹ and José J. Ramasco¹

¹*Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), 07122 Palma de Mallorca, Spain*

²*Aix Marseille Université, Université de Toulon,
CNRS, CPT, UMR 7332, 13288 Marseille, France*

Cities are characterized by concentrating population, economic activity and services. However, not all cities are equal and hierarchy in terms of influence at local, regional or global scales naturally emerges. Traditionally, there have been important efforts to describe this hierarchy by indirect measures such the sharing of company headquarters, traffic by air, train or boats or economical exchanges. In this work, we take a different approach and introduce a method that uses geolocated Twitter information to quantify the impact of cities on rural or other urban areas. Since geolocated tweets are becoming a global phenomenon, the method can be applied at a world-wide scale. We focus on 58 cities and analyze the mobility patterns of people after visiting them for the first time. Cities such as Rome and Paris appear consistently as those with largest area covered by Twitter users after their visit and as those attracting visitors most diverse in origin. The study is also performed discerning users mobility by the contribution of locals and non-locals, which shows the relevance of the mixing ratio between them to have a global city. Finally, we focus on the mobility of users between cities and construct a network with the users flows between them. The network allows to analyze centrality defining it at a global and regional scale. The hierarchy of cities dramatically changes when referred only to urban users, with New York and London playing a predominant role.

Ever since Christaller proposed the central place theory in the 30's [1], researchers have worked to understand the relations and competition between cities leading to the emergence of a hierarchy. Christaller envisioned an exclusive area surrounding each city to which it provided services such as markets, hospitals, schools, universities, etc. The services display different levels of specialization, inducing thus a hierarchy among urban areas according to the type of services offered. In addition, this idea naturally brings an equidistant distribution of urban centers of similar category as long as no geographical constraints prevent it. Still, in the present globalized world the relations between cities go beyond the mere geographical distance. In order to take into account this fact, it was necessary to introduce the concept of world city [2]. These are cities that concentrate economic warehouses like the headquarters of large multinational companies or global financial districts, of knowledge and innovation as the cutting edge technological firms or universities, or political decision centers, and that play an eminent role of dominance over smaller, more local, counterparts. The concept of global cities is, nevertheless, vague and needs further mathematical formalization. This is attained by means of the world city networks, in which each pair of cities is linked whether they share a common resource or interchange goods or people [3–5]. For instance, a link can be established if two cities share headquarters of the same company [5–7], if both are part of good production chains [8], interchange finance services [9], internet data [10] or if direct commercial or cargo flights or boats connect them [3, 11–13]. Centrality measures are then applied to the networks and a ranking of the cities arises. New York and London are, for example, the top rankers in many of these studies [4, 7, 12]. There are, however, inconsistencies in terms of the meaning and stability of the results obtained from different networks or with different centrality measures [12, 14].

Here we use ICT technologies to approach the problem from a different perspective. Let us imagine an announcement with some information handed in the most central square of a given city. We would like to know which is the probability that a person reading that information is able to discuss about it personally in other location of the world a certain time later. The most *central* place at a global scale would simply be that with the greater potential to diffuse the message. This imagined experiment can be performed from a particular square or extending the origin to a full city. Likewise, the destination can be any other world location or another city. The first option, from city to any place, forms a hierarchy of cities according to their potential to spread information, while the latter (city to city) allows us to define a temporal network of influence.

In order to carry out this analysis, it is necessary to localize people while they travel worldwide. This can be done thanks to the introduction in the last years of geolocated ICT services. The use of devices such as mobile phones, credit or transport cards gets registered generating a large quantity of real time data on how people move [15–23]. This information has been used to study questions such as interactions in social networks [24–27], information propagation [28], city structure and land use [21, 29–35], or even road and long range train traffic [36]. It is bringing a new era in the so-called Science of Cities by providing a ground for a systematic comparison of the structure of urban areas of different sizes or in different countries [34, 35, 37–41]. Data coming from credit cards and mobile phones are usually constrained to a limited geographical area such as a city or a country, while those coming from online social media as Twitter, Flickr or Foursquare can refer to the whole globe. This is the reason why we focus here on geolocated tweets, which have already proven to be a useful tool to analyze mobility between countries [42].

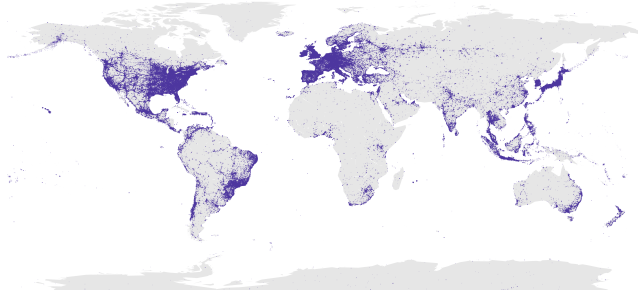


Figure 1: Positions of the geolocated tweets. Each tweet is represented as a point on the map location from which it was posted.

In particular, we select 58 out of the most populated cities of the world and analyze their influence in terms of the average radius traveled and the area covered by Twitter users visiting each of them. Differences in the mobility for local residents and external visitors are taken into account, in such a way that cities can be ranked according to the extension covered by the *diffusion* of visitors and residents, together and by separate, and by the attractiveness they exhibit towards visitors. Finally, we also consider the interaction between cities, forming a network that provide a framework to study urban communities and the role cities play within their own community (regional) versus a global perspective.

Materials and Methods

Twitter Dataset

Our database contains 21,017,892 tweets geolocated worldwide written by 571,893 Twitters users in the temporal period ranging from October 2010 to June 2013 (1000 days). There are on average 93 tweets per user. Non-human behaviors or collective accounts have been excluded from the data by filtering out users traveling faster than a plane (750 km/h). For this, we have computed the distance and the time spent between two successive geolocated tweets posted by the same user. The geographical distribution of tweets is plotted in Figure 1. The distribution matches population density in many countries, although it is important to note that some areas are under-represented as, for example, most of Africa and China.

For this work, we take as reference 58 cities around the world (see Table SI in Appendix for a detailed account). The criterion to choose them was to search a balance between being among the most populated cities worldwide and having enough (users of) geolocated tweets to produce significant results. In practice, the majority of these cities are part of the 100 largest cities in the world. Once the cities are selected, we need to define the extension covered by each urban area. This is not a trivial task as it has been

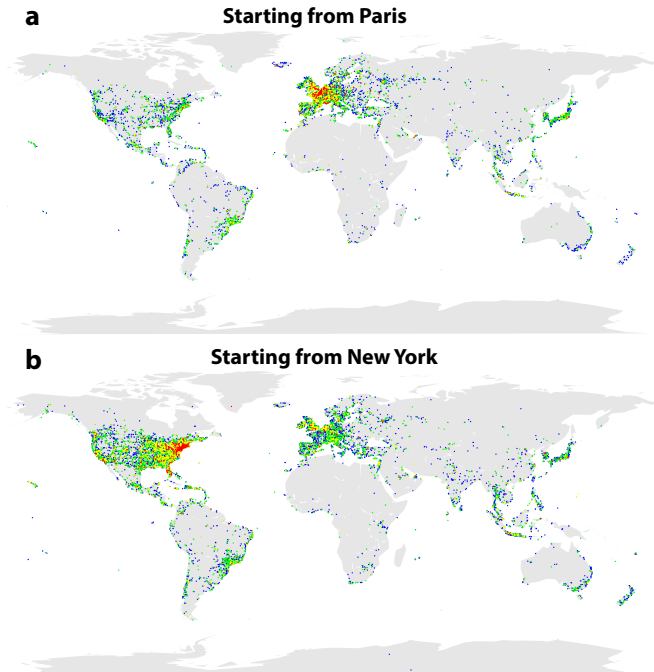


Figure 2: Geolocated tweets of users who have been at least once in Paris (a) and New York (b). The color changes according to the number of days since the first passage in the city. In red, one day; In yellow, between 1 and 10 days; In green, between 10 and 100 days; And in blue, more than 100 days.

shown that the choice can mediate the results if the objective is to analyze scaling relations [43]. Our aim here is different, not focusing on scaling laws, and the spatial scale to operate, global for us, should reduce the importance of this selection. To fairly compare cities and to avoid arbitrariness in the official definitions of urban areas in each country, we have defined the city area as a circle of radius 50 kilometers around the official city center (City Hall location).

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 (see Figure S1 in Appendix). 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 [15, 16, 42]. It is interesting to note that the distributions are very similar for all the cities.

In order to assess the influence of a city, we need to characterize how users travel after visiting it. To do so, for each city c and each user v who visited c for the first time at time t_0^v , we consider the tweets posted by v after t_0^v . A vector of spatio-temporal positions $X = ((x_i^v, t_i^v))_{i \geq 0}$ is then defined for every user v with the rest of locations i visited and the time at which this happened. For studying c , rather than on the absolute times t_i^v , we will focused on the time lapse since t_0^v and thus on $X = ((x_i^v, t_i^v - t_0^v))_{i \leq 0}$. In Figure 2, the

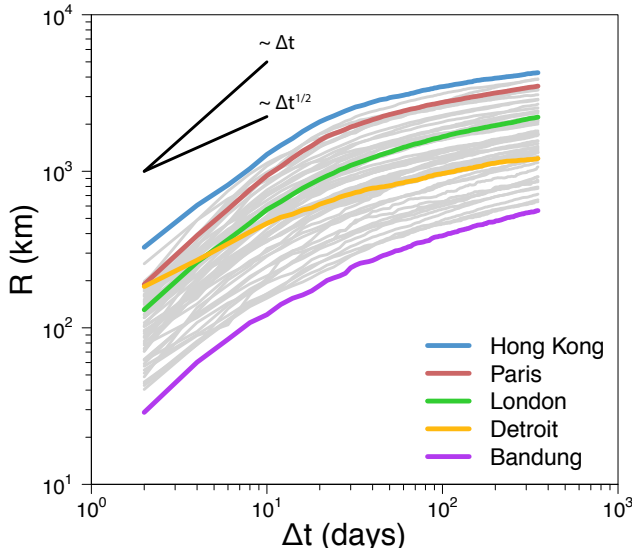


Figure 3: Evolution of the average radius. In order to show the general trend, each gray curve corresponds to a city. The evolution of the radius for several cities is highlighted, such as the top and bottom rankers or representatives of the two main detected behaviors. Curves with a linear and square root growth are also shown as a guide to the eye.

locations of geolocated tweets are plotted according to the number of days since the first visit in Paris and New York as an example. A large part of the tweets are concentrated around these cities but one can observe how users diffuse worldwide little by little.

Metrics to assess city influence

The number of geolocated users in each city and the time lapse that they remain active after visiting it varies. To compare fairly between cities, we have selected a fixed number of users u in each city at random and tracked their displacements in a given period of time Δt since their first tweet from it. The metrics defined next take this set of u users and the locations of their tweets as input. Note that every extraction of the users can generate a different outcome in the metrics, which induces stochasticity in the results. To reduce it, averages over 100 user extractions have been taken for all the results shown in this work. As shown in Figure S2 in Appendix, the longer Δt is, the lower is the population of users who remain active. For this reason, we needed to establish a trade off between high number of users (better statistic resolution) and long periods of time to analyze the system evolution. For most of the results below and unless otherwise stated, $u = 300$ and $\Delta t = 350$ days.

Average radius

The first indicator that we have used is the average radius traveled by Twitter users since their first tweet from a city c . We tracked for each user the positions from which he or she tweeted after visiting c , and compute the average distance from these locations to the center of c . The average radius, R , is then defined as the average over all the u users of their individual radii.

The average radius is informative but can be biased by the geography. Cities that are in relatively isolated positions such as islands may have a high average radius just because a long trip is the only option to travel to them. To avoid this effect, we define the normalized average radius \tilde{R} of a city c as the ratio between $R(c)$ and the average distance of all the Twitter users' places of residence to c (Figure S3 in Appendix). In order to identify the place of residence of Twitter user, the world surface has been divided in cells of 100×100 square kilometers using the cylindrical equal-area projection and the place of resident is identified as the cell most frequently visited. The inhabited cells is a fixed set containing around 5,000 cells. Note that \tilde{R} is a ratio and is, therefore, an adimensional variable.

Coverage

While the average radius measures how far the world exploration by the users starting from c has evolved, this second metric aims at measuring the geographical area covered by those users. Tracking the u users movements, we count the total number of cells from which at least a tweet has been posted and define coverage as this number. The coverage could be thus expressed as a ratio of active cells over the total, but we prefer to maintained it as a simple integer number to facilitate the distinction between nearby values.

Entropy

The u users traveling from a city c can distribute uniformly across the cells or they may prefer some particular locations. To quantify the heterogeneity in the user traveling patterns, we introduce an entropy measure. Let N be the total number of cells and n_i^t the number of tweets in the cell i which have been posted in a time period t . n_i^t is normalized by its total in order to obtain a probability $p_i^t = \frac{n_i^t}{\sum_{k=1}^N n_k^t}$. The entropy E at a time t is given by the following equation:

$$E(t) = \frac{-\sum_{i=1}^N p_i^t \log(p_i^t)}{\log(\mathcal{N}(t))}, \quad (1)$$

where $\mathcal{N}(t)$ is the number of cells with non-zero p_i^t (with tweets posted by the u users). The entropy index is normalized by its maximum value (i.e.

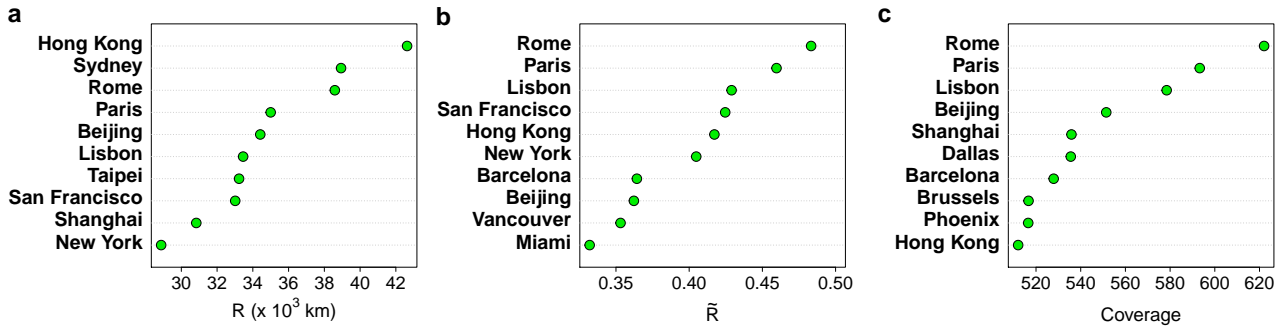


Figure 4: Rankings of the cities according to the average radius and the coverage. (a) Top 10 cities by the average radius after $\Delta t = 350$ days. (b) Top 10 cities by the normalized average radius \bar{R} . (c) The ranking for the coverage.

$\log(\mathcal{N}(t))$ to compare the cities even if the number of cells visited is different.

Results

Comparing the influence of cities

We start by taking the perspective from the city to the world and compare how effective the cities are as starting points for the Twitter users' diffusion. The evolution of the average radius as a function of the time is plotted in Figure 3 for the 58 cities. The curves of the log-log plot show an initial fast increase followed by a much slower growth after approximately 15 – 20 days. The presence of these two regimes can be related to the mixing between residents and external visitors as will be discussed later. In the initial phase, the radius grows for all the cities at a rhythm faster than the square root of time, which is the classical prediction for 2D Wiener diffusion. This is not fully surprising since the users' mobility is better described by Levy flights than by a Wiener process. Still the differences between cities are remarkable. There are two main behaviors: the radius for cities such as Detroit grows slow, while some other cities as Paris show a quite fast increase (close to linear). After this initial transient, the average radius enters in a regime of slow growth for all the cities that is even slower than $\sqrt{\Delta t}$. This implies that the long displacements by the users are concentrated in the first month, after which the exploration becomes languid and probably more localized. Even though the curves of different cities may cross in the first regime, they reach a relatively stable configuration in the second one. We can see that the top ranker in terms of capacity of diffusion is Hong Kong for the whole time window considered and the bottom one is Bandung (West Java, Indonesia).

The top 10 cities according to the average radius are plotted in Figure 4a. It is interesting to realize that the hierarchy only includes New York at the last position, in contrast to previous rankings in the literature based on different data [4, 7, 12]. Many cities on the top are in the Pacific Basin (Hong-Kong, Sydney,

Beijing, Taipei, San Francisco and Shanghai), which can be an evidence for the impact of geography on R . To be fair in the comparison between cities, we have subtracted the effect of isolation by calculating the normalized radius \bar{R} . The corresponding ranking is displayed in Figure 4b. In this case, the top cities are Rome, Paris and Lisbon. These cities are located in a densely populated area as Europe but, taking into account this fact, are dispatching travelers further away than any other. This is a proof for their aptitude as sources for the thought information spreading experiment of the introduction. Actually, all the cities in the top 10 set are able to attract visitors at a worldwide scale, some are relatively far from other global cities and/or they may be the gate to extensive hinterlands (China). The same ranking for the coverage is shown in Figure 4c. Even though the two metrics are correlated (see Figure S4 in Appendix), both rankings do not necessarily agree in all positions. The top cities, however, are again the triplet Rome, Paris and Lisbon probably due to a combination of the factors explained above. One last word of caution, since the users extraction is stochastic, the rankings can variate from a realization to another (see Figure S5 in Appendix). However, when the ranking is performed on the average over several realizations it becomes stable and the results are reproducible (Figure S6 in Appendix).

Local versus non-local Twitter users

These results do not have into account that residents and visitors behave differently when traveling from a given city. An user is considered as a local for a city if most of his/her tweets are posted from it. Otherwise, he/she is seen as an external visitor. The first thing to note is that the average coverage for the 58 cities is higher for non-locals (about 260) than for locals (about 96). The entropy measured in the cities based only in local users is way lower than for the non-locals with a median entropy around 0.3 for the locals and around 0.7 for the non-locals (Figure S7 in Appendix). This means that the locals move toward more concentrated locations, in contrast to the com-

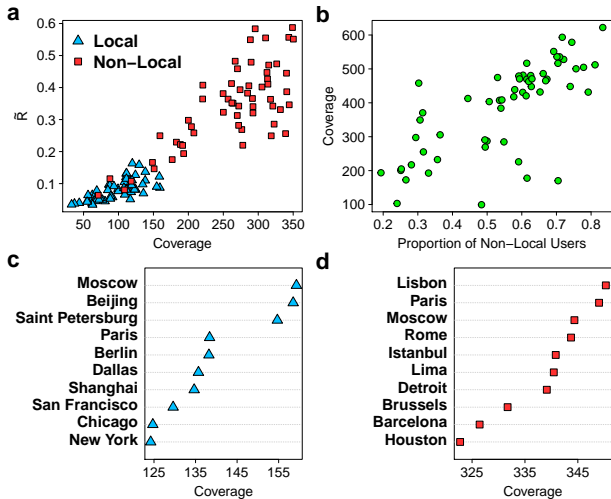


Figure 5: Relation between local and non-local users. (a) Scatter-plot of \tilde{R} as a function of the coverage for locals (blue triangles) and non-locals (red squares). (b) Coverage as a function of the proportion of non-local Twitter users. (c) Top 10 ranking cities based only on local users according to the coverage. (d) The same ranking but based only on the movements of non-local users. In all the cases, the number of local and non-local users extracted is $u = 100$ for every city.

paratively higher diversity of origins of the non-local visitors.

The difference between locals and non-locals is even more dramatic when \tilde{R} from each city is plotted as a function of the coverage for both types of users in Figure 5a. Two clusters clearly emerge, showing that the locals tend to move less than the visitors. Such difference between users is likely to be behind the change of behavior in the temporal evolution of the average radius detected in Figure 3, and introduces the ratio of visitors over local as a relevant parameter to describe the mobility from a city. Actually, the contribution of visitors is the most important factor for the city ranking based on the radius and the area covered (see Figure 5b for the coverage). The division between locals and non-locals allows as well to define new rankings for the cities based on the mobility of residents or visitors alone. Actually, the rankings based on residents bring fresh information on the relevance of cities from a more local perspective (Figure 5c for the coverage and Figure S8a in Appendix for \tilde{R}). The top rankers in this classification are Hong Kong and San Francisco in \tilde{R} and Moscow and Beijing in the coverage. All of them are cities that may act as gates for quite extense hinterlands. The rankings based on non-locals (Figures 5d for the coverage and Figure S8b in Appendix for \tilde{R}) get us back the more common top rankers such as Paris, New York and Lisbon.

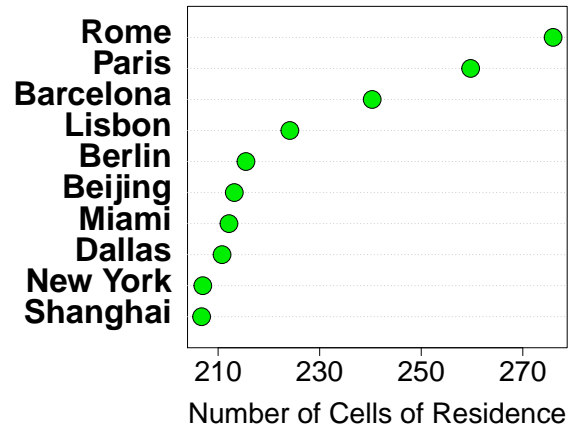


Figure 6: City attractiveness. Number of distinct cells of residence for $u = 1000$ Twitter users drawn at random for the top 10 cities.

City attractiveness

So far, we have considered a city as origin and analyzed how people visiting it diffuse across the planet. We change now the point of view to the opposite, from the world to the city. As before, the place of residence of every user is identified as the cell most frequently visited. With this information, we measure two metrics: the normalized average distance of the users' residences to the center of the considered city c and the number of different cells where these users come from. The normalization of the average distance residence-city is performed as with \tilde{R} to avoid biases due to the geographical distribution of the cities. These variables allow us to define new rankings, this time depicting the attractiveness of each city. The top 10 cities are shown in Figure 6 for the coverage (see Figure S9 in Appendix for the normalized average radius). Rome, Paris and Lisbon are also quite consistently the top rankers in terms of attractiveness to external visitors.

A network of cities

Previously, we have considered how people move from a city to the world and from the world to a city. Here, the analysis is completed considering the travels between the 58 selected cities. For this, we have built a weighted directed network with the flows of movements of the local users from each city to the others. This network can be represented by a matrix $M = (m_{ij})_{1 \leq i, j \leq 58}$ where m_{ij} stands for the number of local Twitter users in the city i which have been at least one time in city j . Since the number of users is different according to the city, we have normalized the flows by dividing m_{ij} by $\sum_{k=1}^{58} m_{ik}$. This network shows stronger connections in the areas that interchange more travelers between them. It is, therefore, interesting to analyze the communities that may

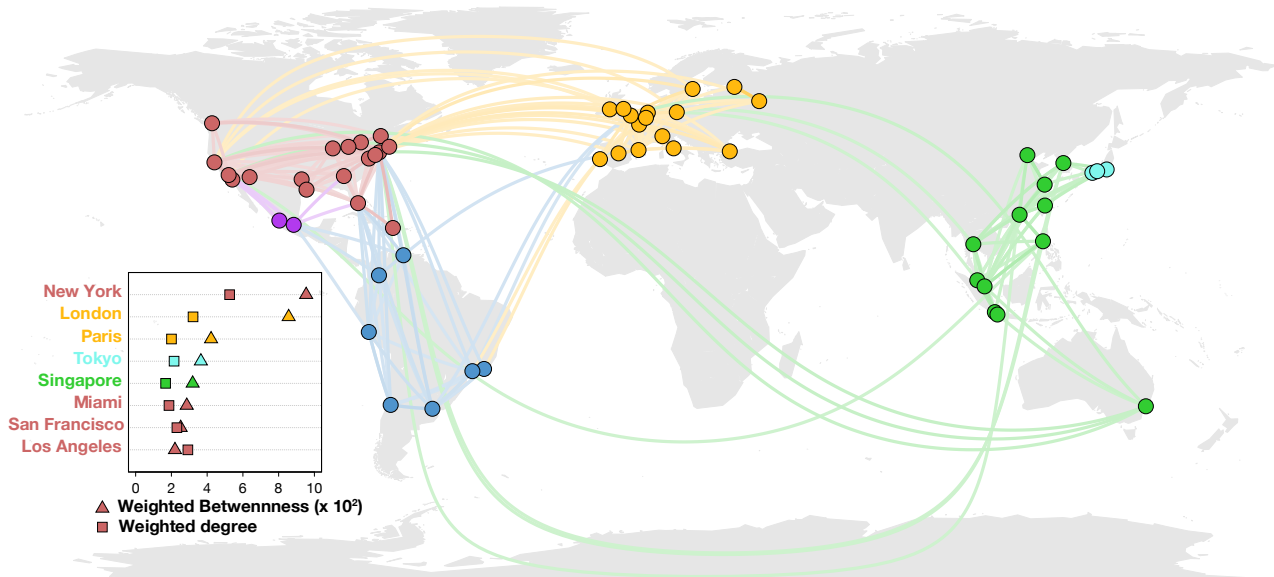


Figure 7: Mobility network. Local Twitter users mobility network between the 58 cities. Only the flows representing the top 95% of the total flow have been plotted. The flows are drawn from the least to the greatest. The inset shows the top 8 cities ranked by weighted betweenness and weighted betweenness.

surge from its structure. To this end, we have employed OSLOM as a clustering detection algorithm [44, 45]. Running OSLOM on the city network, we find 6 communities (see Figure 7). The clusters obtained follow approximately the natural boundaries between continents, two communities in North and Center America, one community in South America, another in Europe, two communities in Asia (Japan and rest of Asia plus Sydney). The community detection has also been performed using Infomap [46] and similar results were obtained.

The presence of network clusters allows to define regional vs global metrics when the measures are restricted to the cities in the cluster or extended to the full network. As centrality metrics, we have computed the weighted degree or strength of a node [47] and the weighted betweenness [48, 49] for all the cities. In the inset of the Figure 7, we identify New York and London as the most central nodes in terms of degree and betweenness and, particularly, New York for the weighted degree at a global scale. In the regional scene of each community, things can change and even the ranking of the most central cities can be modified. The rankings for the weighted degree are similar from the global and regional perspectives, the most central cities continue to be the same except for San Diego, which slipped three places down. On the contrary, the weighted betweenness rankings show some cities for which the rank has gone up or down significantly (see details in Table I and Table SII). For example, San Diego went down nine places meaning that this city has a global influence probably because San Diego is a communication hub between United State and Center America. We can also observed that Dallas went up six places, showing that the influence of Dallas is

North America	
Global Ranking	Regional Ranking
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

Europe	
Global Ranking	Regional Ranking
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

TABLE I: Comparison of the regional and the global betweenness rankings. In parenthesis the total global ranking position of each city.

higher at the regional scale than in the international arena.

Discussion

The study of competition and interactions between cities has a long history in fields such as Geography, Spatial Economics and Urbanism. This research has traditionally taken as basis information on finance exchanges, sharing of firm headquarters, number of passengers transported by air or tons of cargo dispatched from one city to another. One can define a network relying on these data and identify the so-called World Cities, those with a higher level of centrality as the global economic or logistic centers. Here, we have taken a radically different approach to measure quantitatively the influence of a city in the world. Nowadays, geolocalized devices generate a large quantity of real time and geolocated data permitting the characterization of people mobility. We have used Twitter data to track users and classify cities according to the mobility patterns of their visitors. Top cities as mobility sources or attraction points are identified as central places at a global scale for cultural and information interchanges. This definition of city influence makes possible its direct measurement instead of using indirect information such as firm headquarters or direct flights. Still, the quality of the results depends on the capacity of geolocated tweets to describe local and global mobility. Fortunately, several recent works have proven such capacity comparing different data sources as information collected from cell phone records, Twitter, traffic measure techniques and surveys [21, 22, 36].

More specifically and assuming data reliability, we consider the users' displacements after visiting each city. The urban areas are ranked according to the area covered and the radius traveled by these users as a function of time. These metrics are inspired by the framework developed for random walks and Levy flights, which allows us to characterize the evolution of the system with well defined mathematical tools and with a clear reference baseline in mind. Previous literature rankings usually find a hierarchy captured by New York and London as the most central world cities. The ranks dramatically change when one has into account users' mobility. A triplet formed by Rome, Paris and Lisbon consistently appear on the top of the ranking by extension of visitor's mobility but also by their attractiveness to travelers of very diverse origin. A combination of economic activity, appealing to tourism and diversity of links to other lands, in some cases product of recent history, can explain the presence of these cities on the top. These three cities are followed by others such as San Francisco that without being one of the most populated cities in the US extends its influence over the large Pa-

cific basin or Hong Kong, Beijing and Shanghai that replicates it on the other side of the Pacific region. These cities are in some cases gates to broad hinterlands. This is relevant since our metrics have into account the diversity in the visitors' origins.

These results rely on the full users population, discriminating only by the place of residence between locals and non locals to each city. The influence of cities measured in this way includes their impact in rural as well as in other urban areas. However, the analysis can be restricted to users residing in an urban area and to their displacements toward other cities. In this way, we obtain a weighted directed network between cities, whose links weights represent the (normalized) fluxes of users traveling from one city to another. This network provides the basis for a more traditional centrality analysis, in which we recover London and New York as the most central cities at a global scale. The network framework permits to run clustering techniques and divide the world city network in communities or areas of influence. When the centrality is studied only within each community, we obtain a regional perspective that induces a new ranking of cities. The comparison between the global and the regional ranking provides important insights in the change of roles of cities in the hierarchies when passing from global to regional.

Summarizing, we have introduced here a new method to estimate city influence using information and communication technologies. The method, despite some possible biases due to the population using online social media, allows for a direct measurement of a city influence in the world. This introduces the possibility of studying relations among cities and between cities and rural areas with unprecedented details and scale.

I. ACKNOWLEDGEMENTS

Partial financial support has been received from the Spanish Ministry of Economy (MINECO) and FEDER (EU) under projects MODASS (FIS2011-24785) and INTENSE@COSYP (FIS2012-30634), and from the EU Commission through projects EUNOIA, LASAGNE and INSIGHT. The work of ML has been funded under the PD/004/2013 project, from the Conselleria de Educacin, Cultura y Universidades of the Government of the Balearic Islands and from the European Social Fund through the Balearic Islands ESF operational program for 2013-2017. JJR from the Ramón y Cajal program of MINECO. BG was partially supported by the French ANR project HarMS-flu (ANR-12-MONU-0018).

[1] Christaller, W. Die Zentralen Orte in Süddeutschland: eine Ökonomisch-Geographische Untersuchung Über die Gesetz Massigkeit der Verbreitung und Entwick-

lung der Siedlungen mit Städtischen Funktionen, Fischer Verlag, Jena (1933). (English translation: Christaller, W. & Baskin, C.W. Central places in

- Southern Germany, Prentice Hall, Englewood Cliffs, N.J. (1966))
- [2] Friedmann, J. & Wolff, G. World city formation: an agenda for research and action. *International Journal of Urban and Regional Research* **6**, 309–344 (1982).
 - [3] Knox, P.L. & Taylor, P.J. (eds.). World cities in a world-system. Cambridge University Press, 1995.
 - [4] Rimmer, P. Transport and Telecommunications among world cities. In Lo, F.C & Yeung, Y.M. (eds.). Globalization and the world of large cities, Tokyo: United Nations University Press, pp. 433–470 (1998).
 - [5] Taylor, J.P. Specification of the World City Network. *Geographical Analysis* **33**, 181–194 (2001).
 - [6] Derudder, B., Taylor, P.J., Witlox F. & Catalano, G. Hierarchical tendencies and regional patterns in the world city network: A global urban analysis of 234 cities. *Regional Studies* **37**, 875–886 (2003).
 - [7] Derudder, B. & Witlox, F. Assessing central places in a global age: on the networked localization strategies of advances producer services. *J. Retailing and Consumer Services* **11** 171–180 (2004).
 - [8] Brown, E. et al. World City Networks and Global Commodity Chains: towards a world-systems' integration. *Global Networks* **10**, 1470–2266 (2010).
 - [9] Bassens, D., Derudder, B. & Witlox, F. Searching for the Mecca of finance: Islamic financial services and the world city network. *Area* **42**, 35–46 (2010).
 - [10] Neal, Z. Differentiating centrality and power in the world city network. *Urban Studies* **48**, 2733–2748 (2011).
 - [11] Zook, M.A. & Brunn, S.D. Hierarchies, Regions and Legacies: European cities and global commercial passenger air travel. *J. Contemporary European Studies* **13**, 203–220 (2005).
 - [12] Derudder, B. & Witlox, F. On the use of inadequate airline data in mappings of a global urban system. *J. Air Transport Management* **11**, 231–237 (2005).
 - [13] Derudder, B. & Witlox, F. Mapping world city networks through airline flows: context, relevance, an problems. *J. Transport Geography* **16**, 305–312 ((2008).
 - [14] Allen, J. Powerful city networks: More than connections, less than domination and control. *Urban Studies* **47** 2895–2911 (2010).
 - [15] Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).
 - [16] Gonzalez, M.C., Hidalgo, C.A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
 - [17] Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J. & Vespignani, V. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* **106**, 21484–21489 (2009).
 - [18] Noulas, A. , Scellato, S., Lambiotte, R., Pontil, M. & Mascolo, C. A tale of many cities: Universal patterns in human urban mobility. *PLoS one* **7**, e37027 (2012).
 - [19] Bagrow J.P. & Lin, Y.-R. Mesoscopic structure and social aspects of human mobility. *PLoS ONE* **7**, e37676 (2012).
 - [20] Grabowicz, P.A., Ramasco, J.J., Goncalves, B. & Eguiluz, V.M. Entangling mobility and interactions in social media. *PLoS ONE* **9**, e92196 (2014).
 - [21] Lenormand, M, Picornell, M, Garcia Cantú, O, Tugores, A, Louail, T, Herranz, R, Barthélemy, M, Frías-Martínez, E, & Ramasco, J.J. Cross-checking different source of mobility information. *PLoS ONE* **9**, e105184 (2014).
 - [22] Tizzoni, M. et al. On the use of human mobility proxy for the modeling of epidemics. *PLoS Computational Biology* **10**, e1003716 (2014).
 - [23] Jurdak, R. et al. Understanding human mobility from Twitter. Available online at <http://arxiv.org/abs/1412.2154>.
 - [24] Java, A., Song, X., Finin, T. & Tseng, B. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th ACM WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 56–65 (2007).
 - [25] Krishnamurthy, B., Gill, P. & Arlitt, M. A few chirps about twitter. In *WOSP '08: ACM Proceedings of the first workshop on Online social networks*, pages 19–24, New York (2008).
 - [26] Huberman, B.A., Romero, D.M. & Wu, F. Social networks that matter: Twitter under the microscope. *First Monday* **14** (2008).
 - [27] Grabowicz, P.A., Ramasco, J.J., Moro, E., Pujol, J.M. & Eguiluz, V.M. Social features of online networks: The strength of intermediary ties in online social media. *PLoS ONE* **7**, e29358 (2012).
 - [28] Ferrara, E., Varol, O., Menczer, F. & Flammini, A. Traveling trends: social butterflies or frequent fliers? In *Proceedings of the first ACM Conference on Online Social Networks COSN '13*, pages 213–222, New York (2013).
 - [29] Reades, J, Calabrese, F, Sevtsuk, A, & Ratti, C. Cellular census: Explorations in urban data collection. *Pervasive Computing* **6**, 30–38 (2007).
 - [30] Reades, J, Calabrese, F, & Ratti, C. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design* **36**, 824–836 (2009).
 - [31] Soto, V & Frías-Martínez, E. Automated land use identification using cell-phone records, *Procs. of the ACM conference HotPlanet'11* pp. 17–22, New York, (2011).
 - [32] Frías-Martínez, V., Soto, V.,Hohwald, H. & Frías-Martínez, E. Characterizing urban landscapes using geolocated tweets. In *IEEE SocialCom/PASSAT*, pages 239–248 (2012).
 - [33] Noulas, A, Mascolo, C, & Frías-Martínez, E. Exploiting Foursquare and Cellular Data to Infer User Activity in Urban Environments, *Procs. MDM '13*, pp. 167–176 (2013).
 - [34] Louail, T, Lenormand, M, Garcia Cantú, O, Picornell, M, Herranz, R, Frías-Martínez, E, Ramasco, J. J, & Barthelemy, M. From mobile phone data to the spatial structure of cities. *Scientific Reports* **4**, 5276 (2014).
 - [35] Grauwín, S, Sobolevsky, S, Moritz, S, Godor, I, & Ratti, C. Towards a comparative science of cities: using mobile traffic records in new york, london and hong kong. Available online at <http://arxiv.org/abs/1406.4400>.
 - [36] Lenormand, M, Tugores, A., Colet, P. & Ramasco, J.J. Tweets on the road. *PLoS ONE* **9**, e105407 (2014).
 - [37] Bettencourt, L, Lobo, J, Helbing, D, Kuhnert, C, & West, G. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. USA* **104**, 7301–7306 (2007).
 - [38] Batty, M. The Size, Scale, and Shape of Cities. *Science* **319**, 769–771 (2008).
 - [39] Bettencourt, L & West, G. A unified theory of urban

- living. *Nature* **467**, 912–913 (2010).
- [40] Bettencourt, L. The Origins of Scaling in Cities. *Science* **340**, 1438–1441 (2013).
- [41] Batty, M. *The New Science of Cities*. The MIT Press (2013).
- [42] Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P. & Ratti, C. Geo-located twitter as a proxy for global mobility patterns. *Cartography and Geographic Information Science* **41**, 260–271 (2014).
- [43] Arcaute, E., Hatna, E., Ferguson, P., Youn, H., Johansson, A. & Batty, M. Constructing cities, deconstructing scaling laws. Available online at <http://arxiv.org/abs/1301.1674>
- [44] Lancichinetti, A., Radicchi, F. & Ramasco, J.J. Statistical significance of communities in networks. *Phys. Rev. E* **81**, 046110 (2010).
- [45] Lancichinetti, A., Radicchi, F., Ramasco, J.J. & Fortunato, S. Finding Statistically Significant Communities in Networks. *PLoS ONE* **6**, e18961 (2011).
- [46] Rosvall, M. & Bergstrom, C.T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA* **105**, 1118–1123 (2008).
- [47] Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **101**, 3747–3752 (2004).
- [48] Newman M.E.J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* **64**, 016132 (2001).
- [49] Brandes, U. A faster algorithm for betweenness centrality. *J. Math. Soc.* **25**, 163–177 (2001).

Appendix

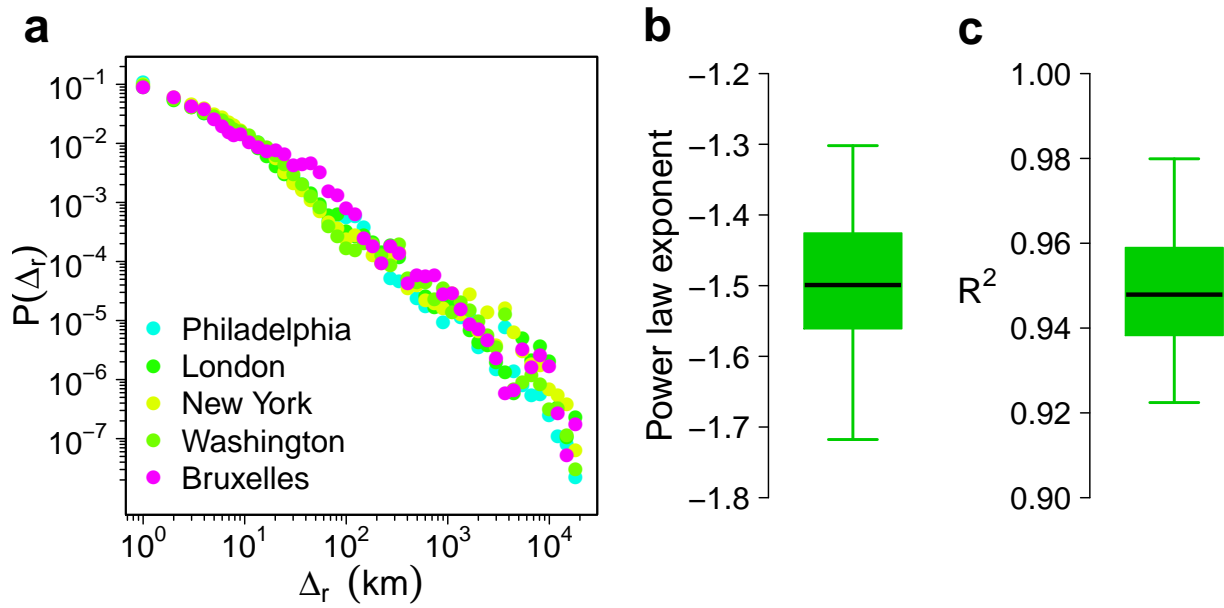


Figure S1: Probability density function of distance traveled by the local Twitter users. An user is considered as living in a city if most of his/her tweets are posted from it. (a) Probability density function $P(\Delta_r)$ of the distance traveled by the local Twitter users for 5 cities drawn at random 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.

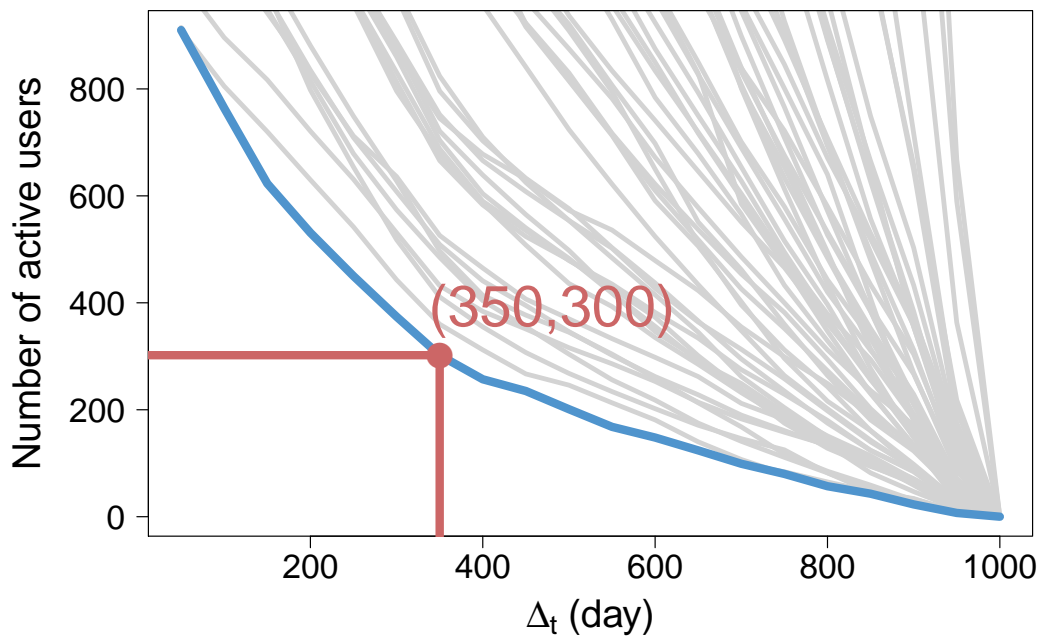


Figure S2: Minimum number of active users as a function of Δ_t (blue line). The gray lines represent the number of active users as a function of Δ_t for the 58 cities.

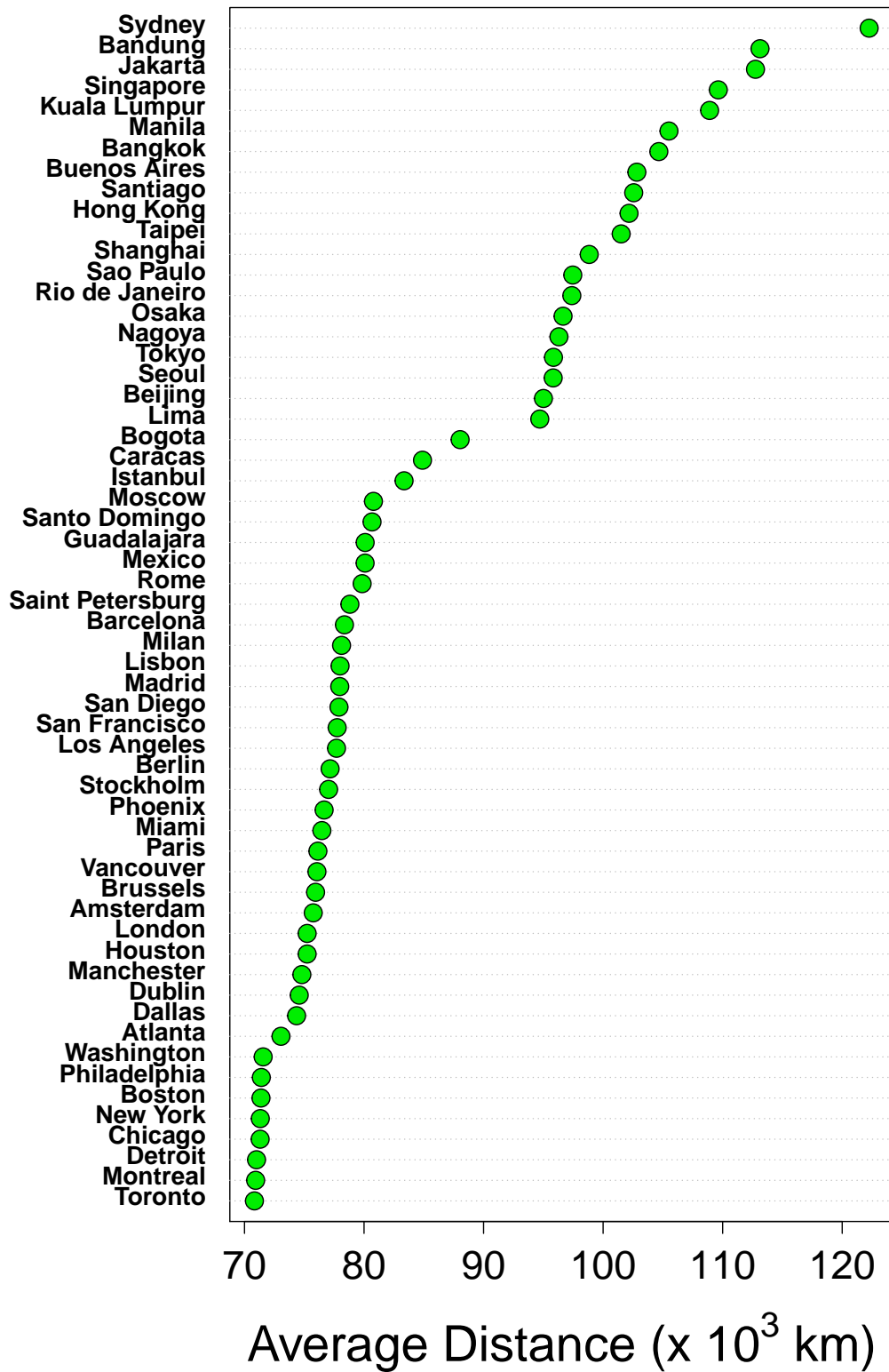


Figure S3: Ranking of the cities according to the average distance between the center of the city and all the Twitter users' place of residence.

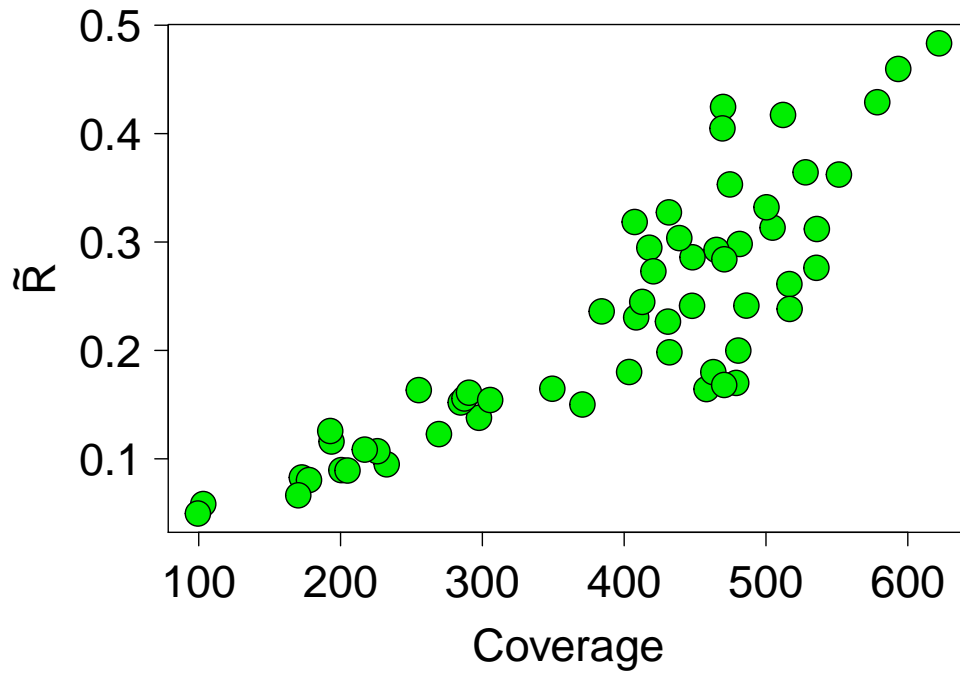


Figure S4: \tilde{R} as a function of the coverage for the 58 cities. A certain level of correlation can be observed between both metrics.

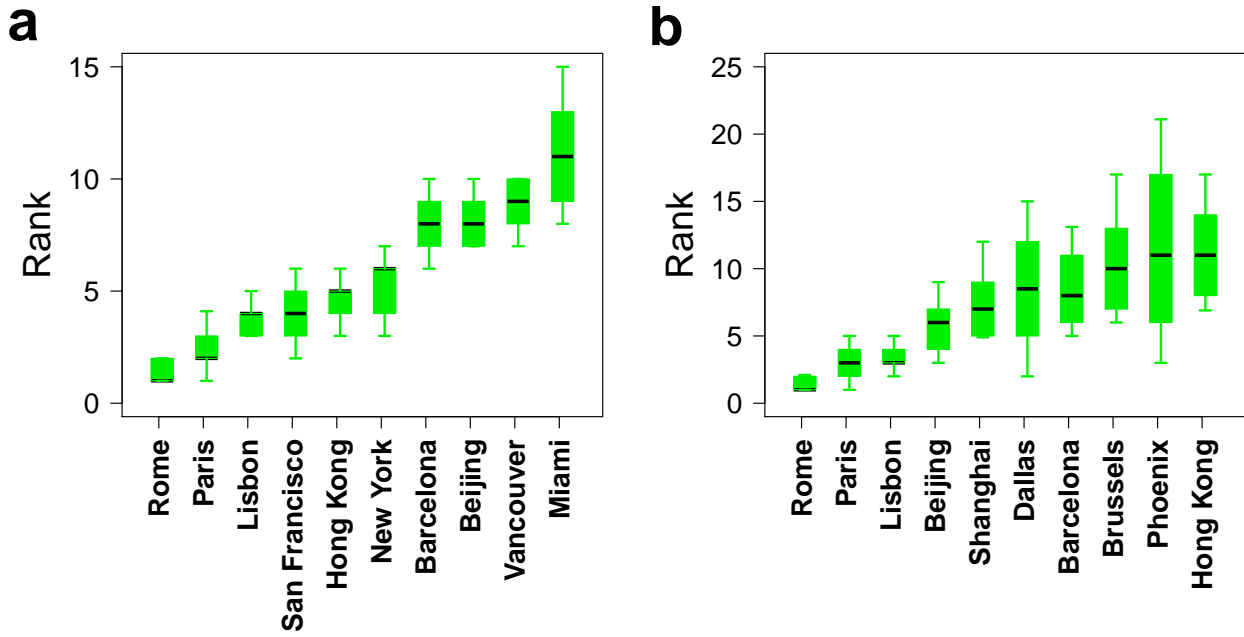


Figure S5: Variations of the rankings over 100 realizations. (a) Ranking for the normalized average radius. (b) Ranking for the coverage. The boxplot is composed of the minimum value, the lower hinge, the median, the upper hinge and the maximum value.

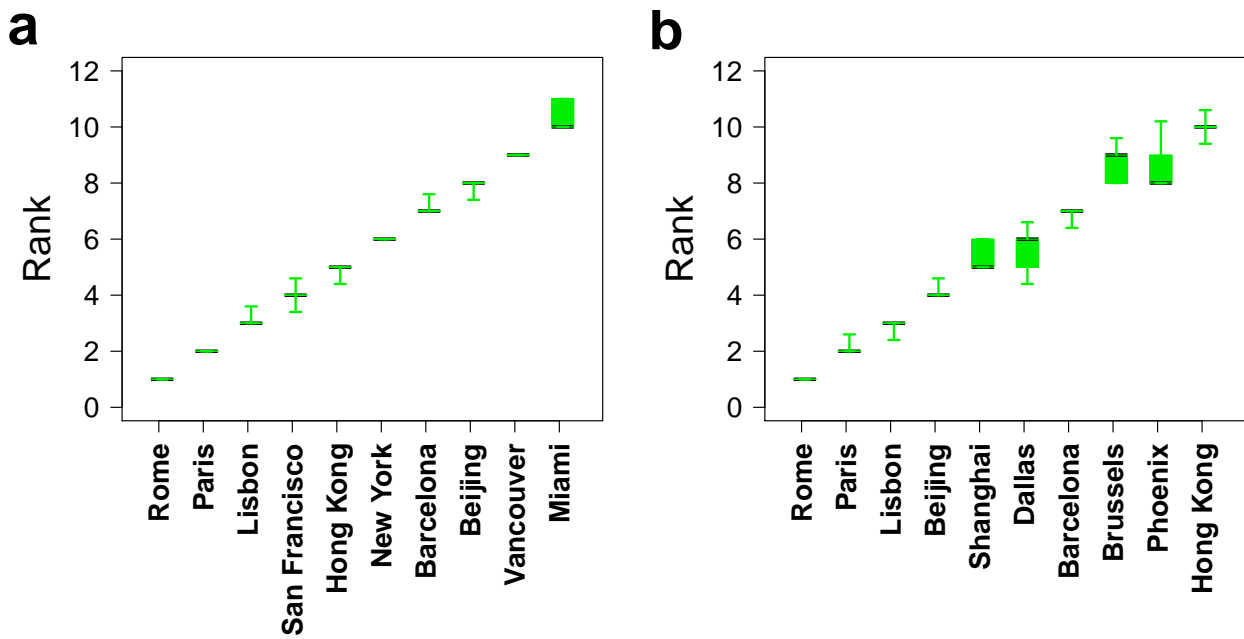


Figure S6: Variations of the rankings over 10 realizations performed on the average over 10 realizations. (a) Ranking for the normalized average radius. (b) Ranking for the coverage. The boxplot is composed of the minimum value, the lower hinge, the median, the upper hinge and the maximum value.

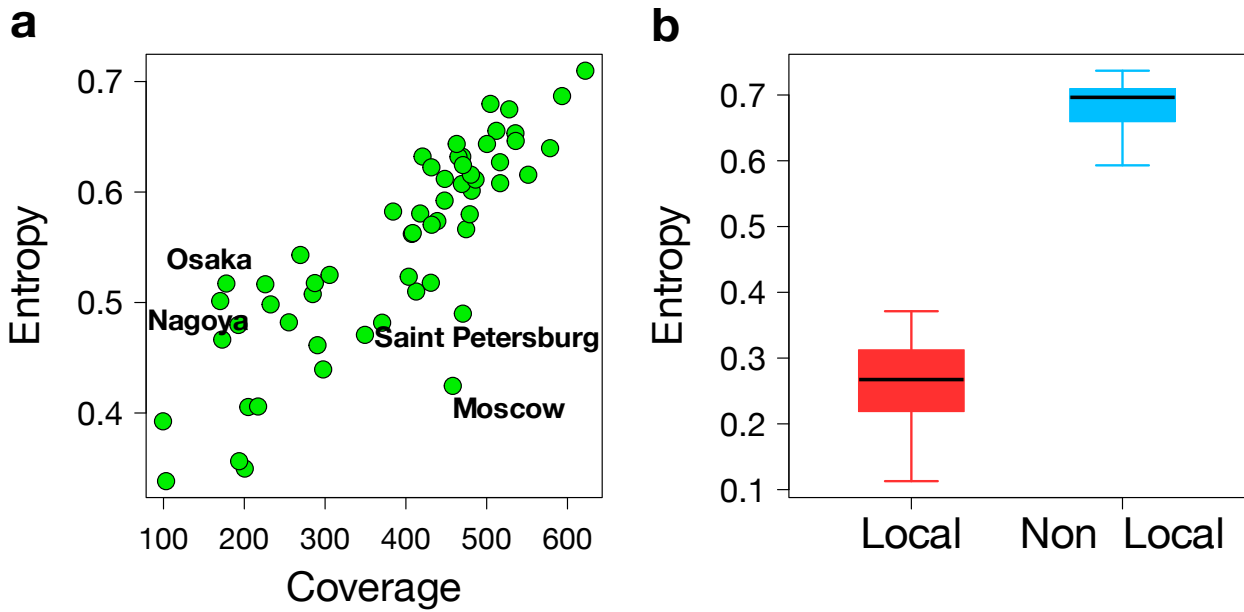


Figure S7: Entropy index according to the Twitter user type. (a) Entropy index as a function of the number of cells visited by $u = 300$ Twitter users drawn at random. The entropy enhances with the coverage despite the normalization, which implies that the tweets tend to distribute more uniformly for 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 area 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. (b) Box plot with the entropy measured for the different cities separating the users as locals and non-locals. The number of users $u = 100$ in this case. 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.

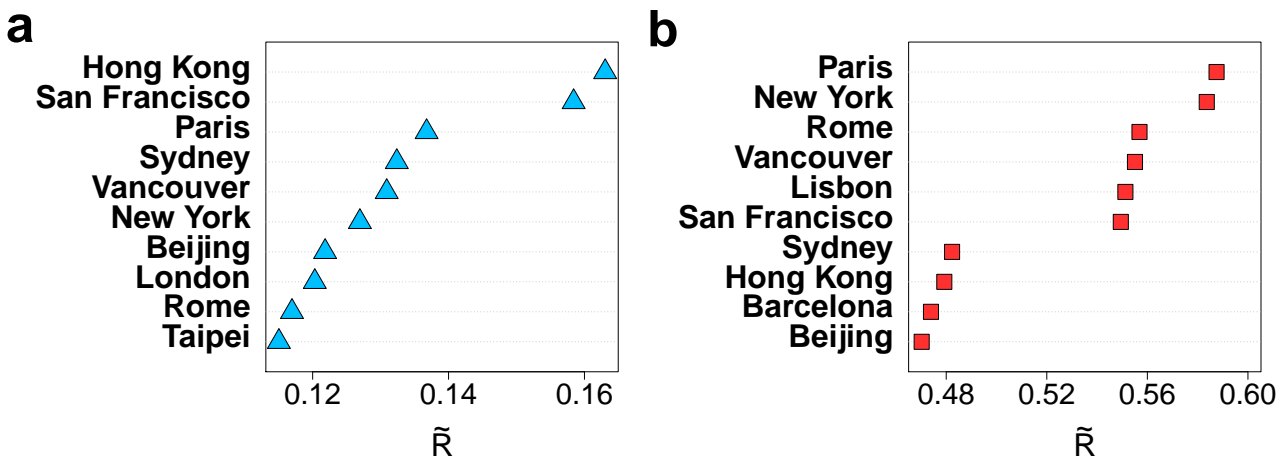


Figure S8: Relation between local and non-local users. (a) Top 10 ranking cities based only on local users according to the average radius. (b) Top 10 ranking cities based only on non-local users according to the average radius. In all the cases, the number of local and non-local users extracted is $u = 100$ for every city.

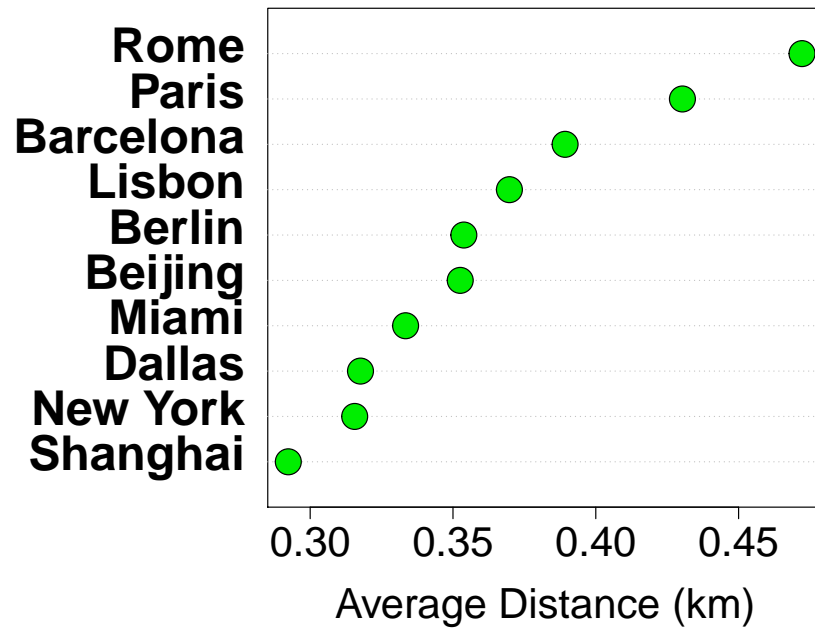


Figure S9: City attractiveness. Normalized average distance between the Twitter users' residences and the city center for the 10 top cities.

TABLE SI: 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 II: Comparison of the regional and the global betweenness rankings.

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