Global Patterns of Synchronization in Human Communications

Alfredo J. Morales [1,2*], Vaibhav Vavilala [1], Rosa M. Benito [2], Yaneer Bar-Yam [1]
[1] New England Complex Systems Institute,
210 Broadway St., Cambridge MA 02139, USA.
[2] Grupo de Sistemas Complejos, Universidad Politécnica de Madrid,
ETSI Agronomos, Av. Comptlense S/N, 28040, Madrid, Spain.
* E-mail: alfredo@necsi.edu.

Supplementary Text

A) Urban Activity

In Figure S1 we show an average week of Twitter activity for 52 cities from around the world: The number of tweets per hour, averaged by weekday, after subtracting the mean and normalizing by the standard deviation.

We performed K-means clustering (*2*) to the time series vectors, starting from the minimum of activity. The quality of the partition optimizes at three clusters (*3*), colored accordingly in Figure S1. The blue cluster includes series that show a single peak. The green cluster includes series that show two similarly sized peaks. The yellow cluster includes series that show large afternoon peaks preceded by smaller morning peaks. Clusters share cultural and regional affinity. Most East-Asian cities are in the single-peaked group, European cities are in the two-peaked group and North and South American cities are in the third group.

We determined that peaks and valleys of activity emerge from social coordination and not solely due to daily light and dark cycles. For this purpose, we calculated the time difference between the series' morning valleys and sunrise times, as well as the time difference between the series' afternoon peaks and sunset times. We grouped these time differences in 10-day intervals and calculated their distributions. We found that the average largest and shortest time differences are significantly different ($p < 0.01$ for equatorial cities and $p < 0.001$ otherwise), indicating that variations in the sunset or sunrise times do not determine the times of peaks or valleys of activity.

B) Heartbeat

We correlated heartbeat ECG signatures with Twitter activity series, as plotted in Figure 2. The heartbeat signatures were obtained from individual heartbeat time series (*4, 5*). We determined the duration of an average heartbeat T, and divided the entire signal into time units 1/24th of the average heartbeat in duration. We averaged the ECG within each of these time units. For non-overlapping windows of 24 units for the entire time series, we identified the minimum value of the ECG in that window $x_{min,i}$. We defined a correlation windows w_i of length T, centered at the minimum value $x_{min,i}$, as $w_i : [x_{min,i} - T/2, x_{min,i} + T/2 - 1]$. For a single heartbeat time series, we obtained a set of 14 correlation windows during regular heartbeat rhythm. We correlated each of the set across the entire heartbeat time series and the average week of Twitter activity. Correlations with random time series were made on vectors of length $2T$, analogously defining correlation windows by determining the minimum within the middle 24 data values.

C) Spatial Patterns

In Figure S2, we show the spatial variation of an average day of Twitter activity of 40 urban areas. We quantified the activity in lattices of 20x20 patches in each area. We show patches' local activity after subtracting the average, normalizing by the standard deviation and coloring the activity above average. All cities have periods of high (colored) and low (black) activity. Activity in central areas during work hours is followed by activity in peripheral areas during rest and recreation hours. The average distance between tweets and the city center (calculated as the center of gravity of the spatial activity) significantly differs ($p < 0.001$ after bootstrapping) during the most contracted and expanded times between 9 am and midnight. In Table S1, we present the maximun and minimun radius of gyration of tweets expressed in kilometers for 50 major metropolitan areas during the same time period.

The spatial patterns of activity vary due to human mobility. We characterized individual users by their most frequently visited locations, identified by sets of tweets located very close to each other in space (within a radius of 100 m). In Figure S3 A, we show the probability density function (PDF) of the number of frequently visited locations per user for each city. In all cities, the distributions peak at two or three frequently visited locations. We considered the users with two or three locations and analyzed the times of the day when they are there. We counted the hourly tweets posted from each location throughout the entire observation period. We separated locations according to two dominant clusters (*2, 3*). In one, people usually tweet during work hours (Fig. S3 B), while in the other, people usually tweet during rest hours (Fig. S3 C). This suggests that locations are dominated by either home or work places (*6*). The average distance between residential locations with respect to the city center is significantly larger from that of work locations ($p < 0.001$ for half of the analyzed cities), indicating that residential locations are more widespread.

D) Global Synchrony

In Figure S4 we show 24 one-hour snapshots of a global network of the correlation of urban activity. Nodes represent cities and edges are present when the cities' time series are correlated above a threshold ($r > 0.9$). Highly synchronized cities will share more connections with each other than with the rest of the network. Correlations are calculated by using overlapping time windows of 12 hours across the cities' average day. The time windows provide snapshots of the network, which we later aggregate into a single graph by weighting edges as the number of times that each pair of cities have been correlated with each other. We colored nodes according to the community structure of the network (*7*), where each community is a group of nodes that are more connected to each other. Cities are often correlated by time zones. However, during some time windows (top row), cities from Europe, Africa and Asia show synchronized behavior.

Twitter interaction mechanisms also peak during the synchronization period. We study the evolution of both message exchange and topic identification mechanisms during an average day. The former is an active user-to-user interaction mechanism, called *mentions*, that people use to exchange pieces of

Table S1: Maximun and minimum radius of gyration (r_g) of tweets between 9 am and midnight expressed in kilometers for 50 major metropolitan areas.

City	Max. r_g	Min. r_g	City	Max. r_g	Min. r_g
Abuja	9.447 ± 0.307	8.801 ± 0.276	Amsterdam	4.832 ± 0.139	4.560 ± 0.146
Ankara	9.538 ± 0.226	8.158 ± 0.244	Athens	2.622 ± 0.054	2.513 ± 0.057
Bangalore	9.158 ± 0.323	8.735 ± 0.263	Bangkok	15.047 ± 0.382	13.251 ± 0.401
Berlin	6.700 ± 0.218	6.144 ± 0.232	Bogota	10.900 ± 0.589	10.403 ± 0.654
BuenosAires	7.048 ± 0.126	6.647 ± 0.122	Cairo	10.051 ± 0.209	9.818 ± 0.235
CapeTown	11.113 ± 0.256	9.799 ± 0.239	Caracas	9.007 ± 0.268	8.497 ± 0.245
Chicago	13.573 ± 0.272	12.965 ± 0.272	Delhi	13.295 ± 0.218	12.536 ± 0.226
Dubai	14.734 ± 0.318	14.837 ± 0.366	HongKong	12.032 ± 0.346	11.843 ± 0.352
Houston	25.036 ± 0.414	23.036 ± 0.447	Istanbul	15.578 ± 0.480	14.411 ± 0.486
Jakarta	13.671 ± 0.265	12.826 ± 0.270	Kingston	4.679 ± 0.093	4.256 ± 0.091
KualaLumpur	6.124 ± 0.103	5.340 ± 0.114	Lagos	13.609 ± 0.330	13.409 ± 0.349
London	13.545 ± 0.255	11.892 ± 0.320	LosAngeles	19.338 ± 0.402	18.336 ± 0.385
Madrid	9.617 ± 0.234	9.143 ± 0.229	Manila	3.391 ± 0.056	3.322 ± 0.055
Melbourne	19.508 ± 0.718	16.706 ± 0.732	Mexico	12.977 ± 0.368	11.135 ± 0.359
Moscow	11.157 ± 0.183	9.516 ± 0.219	Mumbai	10.410 ± 0.243	10.088 ± 0.240
Nagoya	8.736 ± 0.170	7.624 ± 0.193	Nairobi	9.214 ± 0.280	8.147 ± 0.299
NewYork	38.150 ± 1.119	35.572 ± 1.158	Osaka	7.924 ± 0.161	7.009 ± 0.170
Paris	4.066 ± 0.053	3.690 ± 0.061	Prague	4.328 ± 0.195	3.654 ± 0.181
Pretoria	10.809 ± 0.257	9.437 ± 0.272	PuertoRico	44.338 ± 1.418	44.240 ± 1.446
RioJaneiro	15.036 ± 0.472	15.030 ± 0.496	Riyadh	14.358 ± 0.420	13.472 ± 0.419
Rome	6.594 ± 0.177	5.834 ± 0.173	Santiago	19.045 ± 1.359	16.681 ± 1.184
SaoPaulo	21.490 ± 0.670	20.892 ± 0.681	Seoul	12.280 ± 0.263	10.755 ± 0.279
Shanghai	14.503 ± 0.785	14.296 ± 0.797	Stockholm	4.308 ± 0.102	3.447 ± 0.100
Surabaya	6.293 ± 0.148	5.857 ± 0.141	Sydney	18.134 ± 0.625	15.860 ± 0.599
Tokyo	23.855 ± 0.552	21.413 ± 0.612	Vienna	4.734 ± 0.124	4.054 ± 0.127

information. The latter is a passive interaction mechanism, called *hashtags*, that people use to identify their information with ongoing trends. With these mechanisms, people spread pieces of information on the social network through a cascading effect (*8*). Cascades emerge and grow in social networks as people synchronize their behavior by paying attention to each other and talking about similar topics. We capture this synchrony by aggregating the hourly number of directed messages and shared hashtags between the European and Asian longitude ranges ([-30,30] and [90,180] respectively). Both interaction mechanisms significantly peak ($p < 0.001$) during the synchronized period (shadowed regions). At this time a larger number of directed messages are sent between these regions and more hashtags are shared in their messages. These results indicate that people tend to share more information about increasingly similar topics as they synchronize their activities.

E) Spectral Analysis

The spectral behavior (Fourier transform) of the Twitter activity from the 52 cities is shown in Fig. S5. All frequency spectra have three significant components at 24h, 12h and 8h (dashed lines). The first is due to variations associated to the daily cycle, the second to variations during 12 hours periods, night and day, the third corresponds to periodic variations within work, recreation and sleep 'shifts.'

F) Modeling Dynamics

The Twitter activity time series are modeled by adding three sinusoid signals of 24, 12 and 8 hours period respectively:

$$s(t) = a_{24} sin(\frac{2\pi t}{24} + \theta_{24}) + a_{12} sin(\frac{2\pi t}{12} + \theta_{12}) + a_8 sin(\frac{2\pi t}{8} + \theta_8) \tag{1}$$

where t is time in hourly resolution, θ represents the respective signal phase, and a is the signal amplitude in the range [0,1]. We respectively fit the parameters θ and a for each time series by minimizing the quadratic error between the $s(t)$ and the data points. The modeled curves remarkably fit the data ($p < 0.001$) as shown in Fig. S6.

G) Data

The data analyzed in this paper is available at:
http://necsi.edu/research/networks/globalsync/materials.html

References and Notes

1. Borg, I. & Groenen, P. J., 2005 *Modern Multidimensional Scaling: Theory and Applications*. New York, NY: Springer Science & Business Media.

2. MacQueen, J., 1967 Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pp. 281–297. Berkeley, CA: University of California Press.

3. Rousseeuw, P. J., 1987 Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53 – 65. ISSN 0377-0427. (doi: http://dx.doi.org/10.1016/0377-0427(87)90125-7).

4. Greenwald, S., 1986 Master's thesis, MIT.

5. Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P., Mark, R., Mietus, J., Moody, G., Peng, C.-K. & Stanley, E., 2000 Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* **101**, 215–220.

6. Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J. M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A. & Volinsky, C., 2013 Human mobility characterization from cellular network data. *Commun. ACM* **56**, 74–82. ISSN 0001-0782. (doi:10.1145/2398356.2398375).

7. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E., 2008 Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008. doi: 10.1088/1742-5468/2008/10/P10008.

8. Morales, A., Borondo, J., Losada, J. C. & Benito, R. M., 2014 Efficiency of human activity on information spreading on Twitter. *Social Networks* **39**, 1–11. (doi:10.1016/j.socnet.2014.03.007).

Figure S1: Temporal variation of Twitter activity from 52 cities during an average week. Color indicates the result of a clustering classifier (see text). Vertical black lines show the time of synchronization discussed in the manuscript.

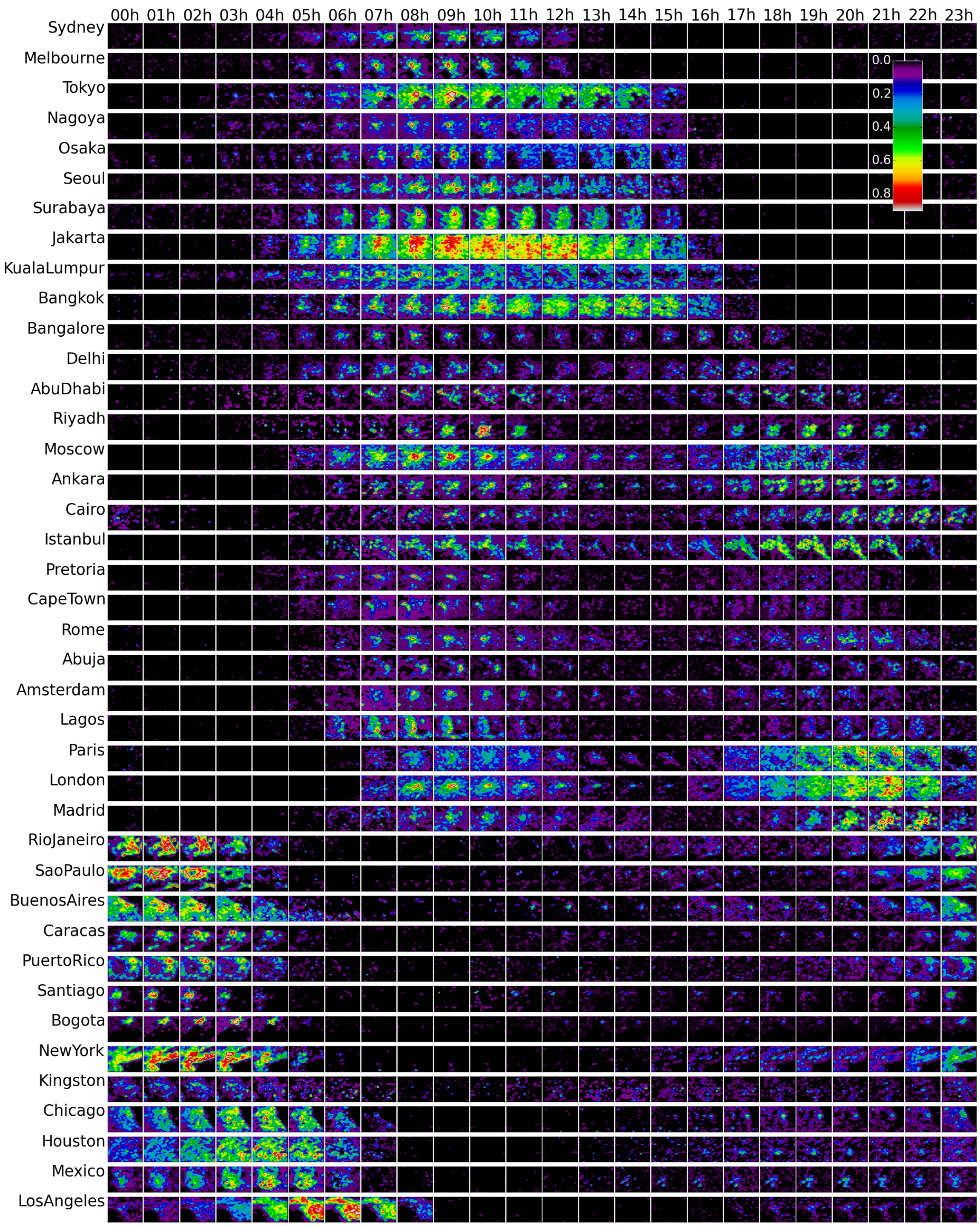

Figure S2: Spatio-temporal dynamics of Twitter activity in urban areas. Each row shows hourly activity during an average day according to UTC time. Colors indicate the normalized excess of activity from the average value at that location (scale shown in figure). Rows are ordered by longitude: Asia (top), Middle East, Europe and Africa (center), South and North America (bottom).

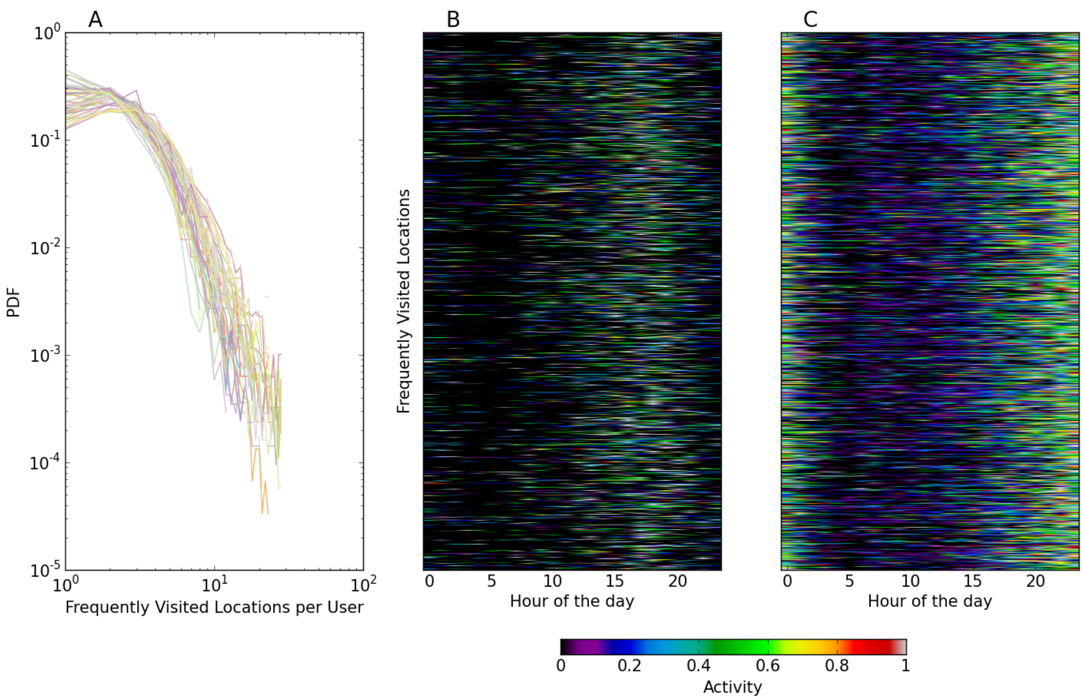

Figure S3: Clusters of frequently visited locations according to the time of the day of individual activity. Dominant location clusters have primary activity during either conventional work (9-5) or rest hours according to local time. A. Probability density function (PDF) of the number of frequently visited locations per user for each city. B. Total activity in all cities at clusters whose primary activity is during conventional work hours. C. Like B but for clusters whose primary activity is not during work hours (scale is shown in the figure).

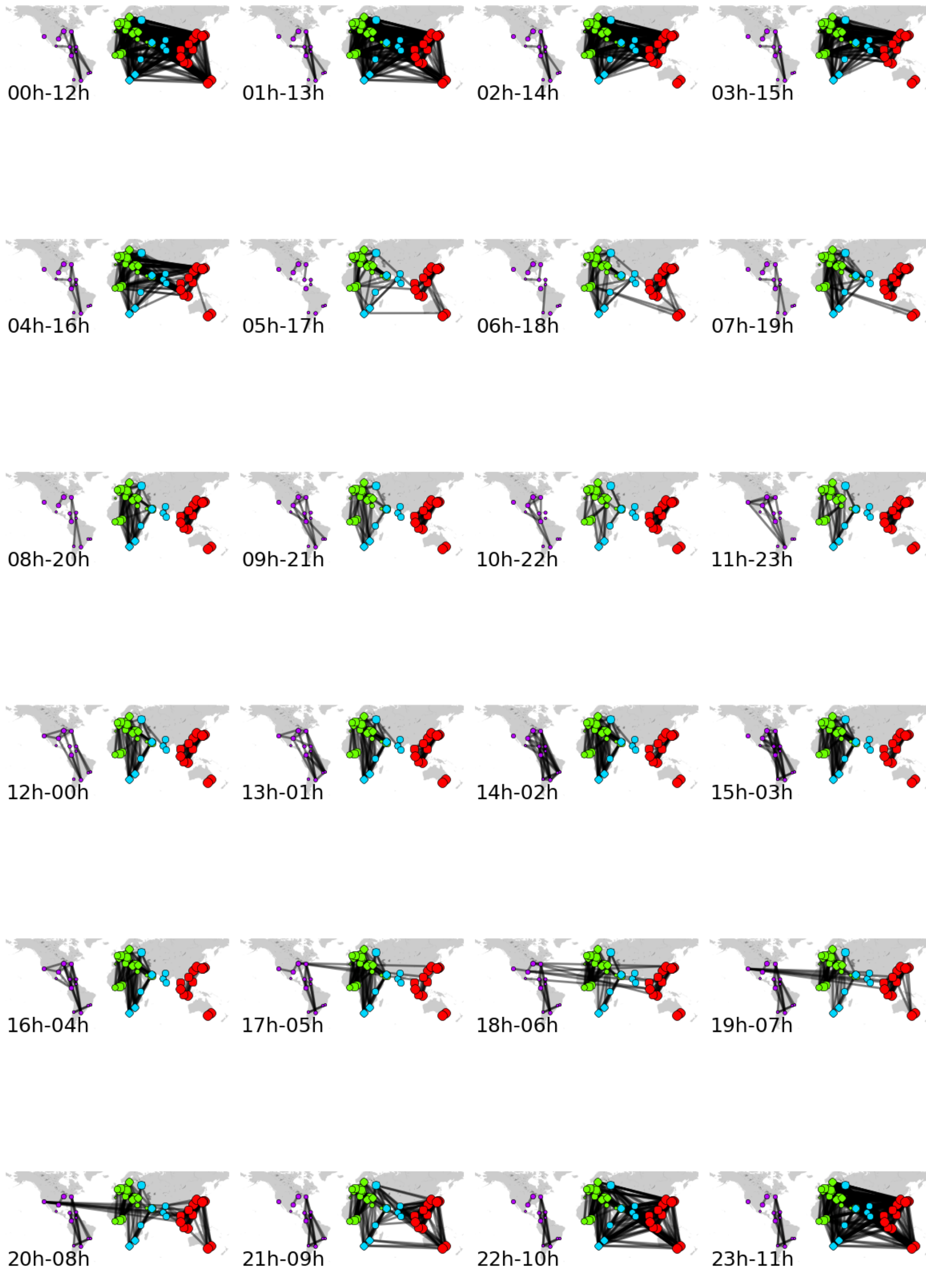

Figure S4: Urban correlation network over an average day. Nodes represent cities and edges are present when 12-hour activity series (labels) are correlated above 0.9. Color indicates the community structure of the aggregated network (see text).

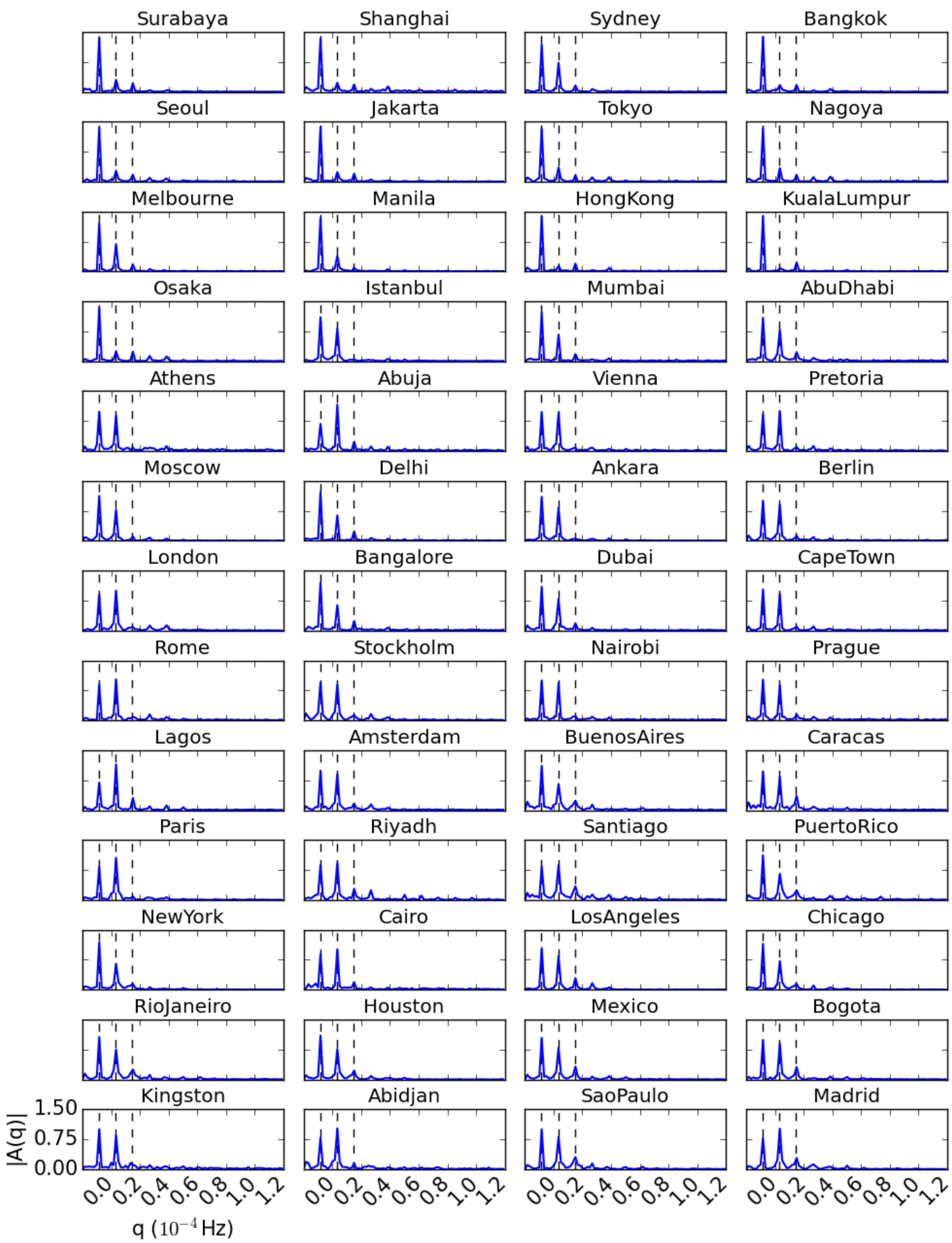

Figure S5: Spectral analysis of the Twitter activity (amplitude of the Fourier Transform) from urban areas. The dashed lines indicate (from left to right) the frequencies (q) corresponding to the periods of 24 hours, 12 hours and 8 hours respectively.

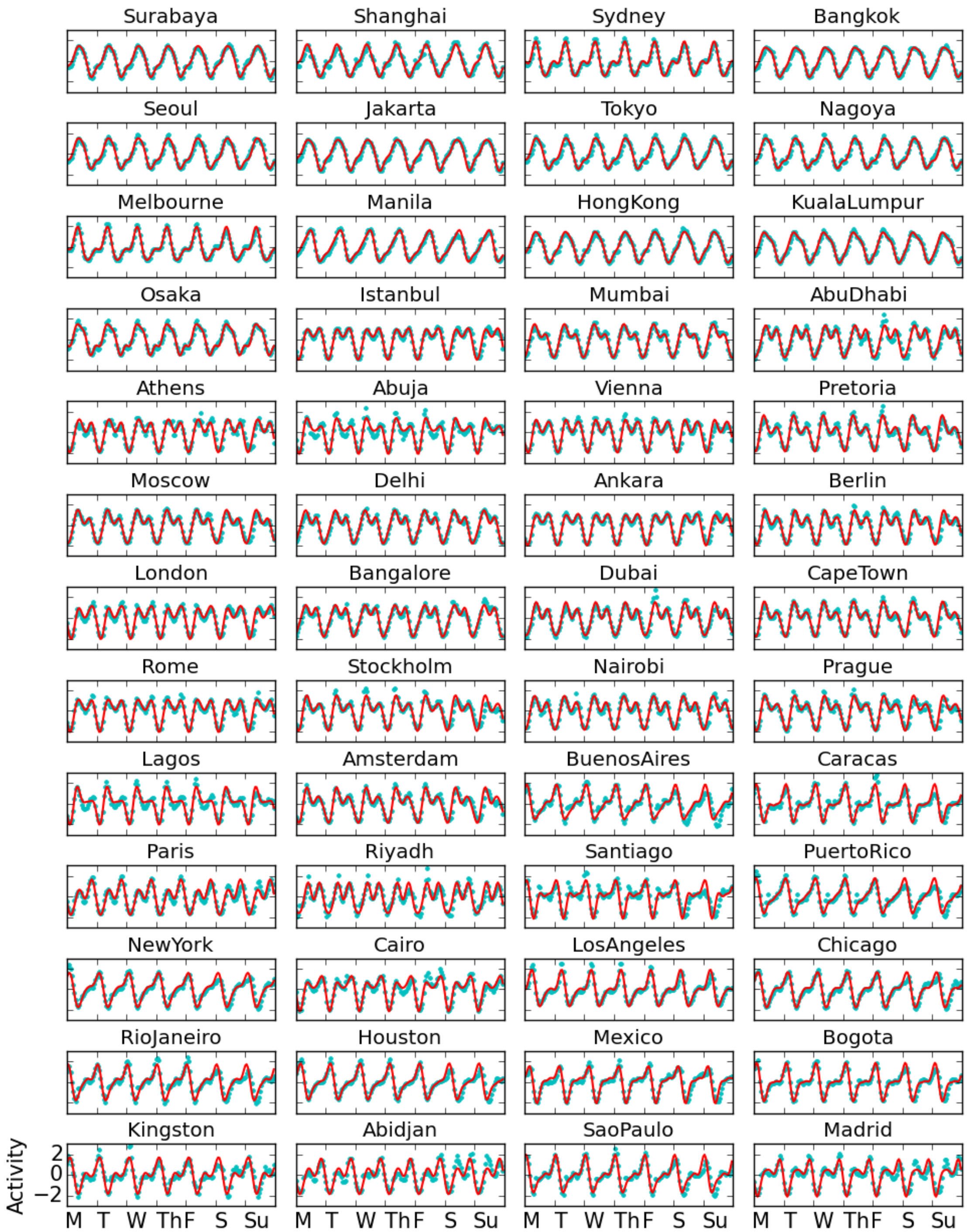

Figure S6: Modeling the Twitter activity time series from urban areas. Blue dots represent the average hourly number of tweets during an average week at each city. The red curve represents the model results. The modeled curves fit the data with $p < 0.001$.