A geocomputational analysis of Twitter activity around different world cities

Muhammad ADNAN*, Alistair LEAKb and Paul LONGLEYab

aDepartment of Geography, University College London, Gower Street, London WC1E 6BT, UK; bDepartment of Security and Crime Science, University College London, 35 Tavistock Square, London WC1H 9EZ, UK

(Received 19 November 2013; final version received 25 June 2014)

The penetration and use of social media services differs from city to city. This paper is aimed to provide a comparison of the use of Twitter between different cities of the world. We present a temporal analysis of activity on Twitter in 15 cities. Our study consists of two parts: First, we created temporal graphs of the activity in the 15 cities, through which hours of high and low activity could be identified. Second, we created heat map visualizations of the Twitter activities during the period of 19 September 2012–25 September 2013. The heat map visualizations make the periods of intense and sparse activity apparent and provide a snapshot of the activity during the whole year.

Keywords: Twitter; temporal analysis; heat map visualization; activity patterns

1. Introduction

Social media services generate huge quantities of data every day. The use of these data is increasing in the research and commercial sectors as recent years have shown an increase in the use of social media in service delivery. Users of the likes of Twitter, Facebook, Flickr, LinkedIn, Bebo, Sina Weibo, and Orkut are frequently mobile users of the developing range of smartphones and tablet devices.

This paper focuses on Twitter, a very popular social media service. This social-networking and micro-blogging service was launched in 2006 and within six years had accrued some 200 million active users, who typically send 340 million tweets every day (1). This is a huge data source and the analysis of these data can provide useful insights into the places and situations in which users avail themselves of this social media service.

The principal focus of this paper is to use geotagged tweets to analyze and compare the Twitter activity in different cities of the world. Geotagged tweets have location (latitude/longitude) information associated with them. Here, we provide the comparative analysis between the top 15 Twitter cities. These cities were called top Twitter cities with regard to the number of tweet messages users sent from each city.

There are a number of reasons for presenting this work. First, we believe that the temporal analysis of the activities on social media can reveal some interesting trends about a place or city, i.e. weekly patterns, hours of sleep vs. work, wake-up times, seasonal shifts of the activity, religious months, etc. Second, we believe that simple visualizations could be created from the terabytes of data that are collected. It is very difficult to process and make sense of data containing millions of records. However, transforming millions of records to a simple heat map visualization makes the understanding of the data much easier. Third, we believe that it is possible to get a basic understanding of a place or a city by the analysis of Twitter activities.

Previous research on Twitter data has emphasized different themes, such as the investigation of tweeting behavior (2), ethical issues in the analysis of Twitter data (3), tracking community happiness from tweets (4), geographic dissection of the Twitter network (5), community detection by using personal names (6), using machine learning techniques to infer political orientation, gender and ethnicity of Twitter users (7), and use of network metrics to compare the social dynamics of Twitter usage with those of physical communities (8). However, no large-scale geotemporal analysis of Twitter usage has yet been undertaken. Rios and Lin (9) presented a large-scale analysis of activity on Twitter in 50 major cities around the world throughout all of 2012, but did not attempt the analysis just on use of tweets which have location information associated with them i.e. geotagged tweets. In this paper, we concentrate on the use of geotagged tweets for a large-scale analysis of activity on Twitter in different cities.

In 2012, London, New York City, and Paris were the 3rd, 5th, and 7th ranked cities with regard to the amount of tweeting (10). This analysis was based on the use of all the tweets, i.e. both geotagged and non-geotagged. In this paper, we provide a list of the top Twitter cities with regard to the amount of geotagged tweets sent.

The structure of this paper is as follows:

- Section 2 describes the data used in our study.
- In Section 3, we define how the data were transformed to produce the temporal graphs and heat

*Corresponding author. Email: m.adnan@ucl.ac.uk

© 2014 Wuhan University
illustrates the pro-

lists the latitude, longi-

Buenos Aires, Madrid, Dubai, Seoul (Korea), and Osaka

City, London, and Los Angeles being the 5th, 6th, and

Finally, conclusions and directions for the future
research are presented in Section 6.

2. Data

Twitter provides the facility for programmers and develop-
ers to download live Twitter data using its public stream-
ing application program interface (API) [11]. The API provides a 1% sample of live geotagged tweets within a user-specified bounding rectangle, at any point in time.

For this paper, the Twitter Streaming API was used to download geotagged tweets for the whole world over the period of 19 September 2012–25 September 2013. We were not able to collect data throughout all of May, June, and July 2013. We have missing data for the following 41 days; May 16–29; June 26–30; and July 1–22. Thus, the study period is based on 331 days. A total of 800 million geotagged tweets (approximately) were downloaded during this period.

3. Data processing and analysis

After the data had been downloaded, the next step was to convert the data to a specific format and perform data analysis. There were three steps involved in this process which are described below.

3.1. Extraction of tweets for each city

Geotagged tweets have latitude and longitude information associated with each text message. Because all the geotagged tweet messages around the world were downloaded, the first step required the geographic attribution of tweets to individual cities of the world. Bounding box rectangle coordinates were used to query the database to extract the tweets for each individual city. Figure 1 shows the top 30 Twitter cities with regard to the number of downloaded geotagged tweets. Jakarta, Istanbul, and Paris are the top three Twitter cities, with New York City, London, and Los Angeles being the 5th, 6th, and 7th top cities correspondingly. Riyadh was the 10th city, Buenos Aires, Madrid, Dubai, Seoul (Korea), and Osaka (Japan) are also featured in the list of top 30 Twitter cities around the world.

The number of geotagged tweets is small beyond the top 15 or so cities, and so we used rank 15 as our cutoff point.

3.2. Date and time conversion

By default, the Twitter API provides the date and time of the tweet messages relative to where the data extraction program is running. Our program was run in London, so all the date and time values are recorded as GMT (Greenwich Mean Time) or British summer time. By way of illustration, Table 1 lists the latitude, longitude, and date and time fields of some of the tweet messages originating from Paris.

However, for the temporal analysis, it was necessary to convert the date and time field to the corresponding time zone of the cities where tweet messages originated. For the data listed in Table 1, Table 2 illustrates the procedure of transforming GMT to the corresponding time zone in Paris i.e. Central European Time Zone (UTC+01:00).

This process was performed for each of the top 15 Twitter cities. Next, temporal graphs of the aggregated temporal activity, around the top 15 Twitter cities, were created. These graphs are described in Section 4.

3.3. Creating intervals for heat map visualizations

The third step of the data analysis was to prepare the data for the production of heat map visualizations. For this, we divided each day into 30-min intervals and computed the number of tweets that were posted during each interval as a way of finding activity patterns. This resulted in 24 × 2 = 48 intervals for each day.

For each city, this resulted in 331 × 24 × 2 = 15,888 data points. This activity was performed for the top 15 Twitter cities. Afterwards, heat maps were created for each city. These heat map visualizations are described in Section 5.

4. Aggregated temporal activity

This section presents the aggregated temporal activity of the top 15 Twitter cities using a 24-h clock. Figure 2 shows a series of temporal graphs of the top 15 Twitter cities. The x-axes of the graphs show the 24-h clock and the y-axes show the relative frequency of the tweet messages. The temporal graphs of different cities have distinct patterns and overall Twitter activity trends could be identified from the graphs. The following is a city-by-city observation of each of the temporal graphs.

Jakarta: Jakarta is the capital and the largest city of Indonesia. The Twitter activity in Jakarta remains low between 1 am and 5 am. Activity starts increasing after 5 am and reaches a local peak at 9 am. During the day, Twitter usage keeps increasing and reaches the peak at 9 pm. After 9 pm, Twitter usage starts to decline. By looking at the temporal graph, we can say that residents in Jakarta wake up and go to sleep earlier when compared to the residents of other cities. The temporal graph of Twitter activity in Jakarta is very distinctive compared to the graphs of the other cities.
Istanbul: Istanbul is the largest city of Turkey. The Twitter activity in Istanbul remains low between 2 am and 7 am. The activity has two peaks, one at 12 pm noon and another at 9 pm. After 9 pm, the Twitter usage starts declining. We can infer that residents in Istanbul start sleeping after 9 pm, which is earlier than some of the other European cities.

Paris: Paris is the capital of France. The Twitter activity in Paris remains low between 2 am and 7 am. The activity has two peaks, one at 12 pm noon and another at 9 pm. After 10 pm, Twitter usage starts declining. By looking at the temporal graph, we can suppose that residents in Paris sleep late. However, Paris has a similar activity temporal graph to Istanbul and London.

Sao Paulo (Brazil): Sao Paulo is the largest city in Brazil. Twitter activity in Sao Paulo remains low between 3 am and 6 am. Twitter activity gradually increases after 6 am and reaches its first peak at 1 pm. Twitter activity reaches another peak during the night at 10 pm. Sao Paulo has a similar temporal graph to Rio de Janeiro.

New York City: New York City (NYC) is the most populous city in United States. Twitter activity in NYC remains low between 3 am and 6 am. The activity has two peaks, one at 1 pm and another at 9 pm. After 8 pm, Twitter usage starts declining. NYC has a similar temporal graph to Los Angeles, Chicago, and Dallas.

London: Twitter activity in London remains very low between 2 am and 7 am. The activity has two peaks, one at 1 pm and another at 9 pm. After 8 pm, Twitter usage starts declining. Twitter activity starts increasing after 6 am, which indicates that residents in London wake-up earlier than many cities. During the day, Twitter activity has a peak at 10 am which is very different from other cities where the first peak of activity is normally at lunch time. Activity in London has another peak during the night between 9 pm and 10 pm.

Los Angeles: Los Angeles is the most populous city in the US State of California. Twitter activity in Los Angeles remains very low between 2 am and 7 am. Activity starts increasing after 7 am and reaches a peak around lunch time i.e. 12 pm noon. Activity in Los Angeles has another peak during the night at 10 pm. Los Angeles has similar activity temporal graphs to New York City, Chicago, and Dallas.

Rio de Janeiro: Rio de Janeiro is the second largest city of Brazil. There is low Twitter activity in Rio de
Janeiro between 3 am and 6 am. Twitter activity gradually increases after 6 am and reaches a first peak at 1 pm. Activity reaches another peak during the night at 10 pm. Rio de Janeiro has a similar temporal graph to Sao Paulo.

*Mexico City*: Mexico City is the capital of Mexico. There is low Twitter activity in Rio de Janeiro between 3 am and 6 am. Twitter activity gradually increases after 6 am and reaches a first peak at 1 pm. Activity reaches another peak during the night at 10 pm. Mexico City has a similar activity temporal graph to Buenos Aires.

*Riyadh*: Riyadh is the capital and largest city of Saudi Arabia. Twitter activity in Riyadh is very different from...
other cities. There is a constant Twitter activity during the whole day. The Twitter activity is low between 3 am and 6 am; however, it is not as low as we can find in the temporal graphs of other cities. Twitter activity in Riyadh has two peaks, one at 1 pm and another at 10 pm. However, more tweets were sent at 1 pm than 10 pm, which is different to the temporal pattern observed in other cities. This shows that, in Riyadh, there is more Twitter activity during lunch time than at night.

Tokyo: Tokyo is the capital of Japan. Twitter activity in Tokyo is very different from other cities. The activity remains low between 1 am and 7 am, which is the longest period of low Twitter activity observed in any of the cities. During the day, temporal graphs show three different peaks of Twitter activity. The first peak is at 8 am i.e. breakfast time, the second peak is at 12 pm i.e. lunch time, and the third peak is at 6 pm. The last peak of Twitter activity is very early compared to the other cities. Twitter activity gradually decreases after 6 pm. This shows that residents in Tokyo sleep very early as compared to other cities.

Chicago: Chicago is the third most populous city in the United States. Twitter activity in Chicago remains low between 3 am and 6 am. Activity has two peaks, one at 12 pm noon and another at 10 pm. Chicago has similar activity temporal graphs to New York City, Los Angeles, and Dallas.

Buenos Aires (Argentina): Buenos Aires is the capital and largest city of Argentina. There is low Twitter activity in Buenos Aires between 4 am and 7 am, which indicates that residents in Buenos Aires wake up relatively late. Twitter activity gradually increases after 7 am and reaches a first peak at 2 pm. Activity reaches another peak during the night between 10 pm and 11 pm. Buenos Aires has a similar temporal profile to Mexico City.

Madrid: Madrid is the capital of Spain and its largest city. Twitter activity in Madrid is also very different to the other cities. There is a constant Twitter activity during the whole day, i.e. the frequency of tweets is always higher than 0.03. There is a first peak of activity at 7 am, which is not found on the temporal graphs of other cities. The activity remains low after the breakfast time and a low number of tweets are sent at 12 pm noon i.e. around lunch time. The activity has two more peaks after lunch time, one at 3 pm and another at 11 pm. The temporal graph also suggests that residents in Madrid sleep relatively late.

Dallas: Dallas is the ninth largest city in United States. Twitter activity in Dallas remains low between 3 am and 7 am. During the rest of the day, the activity has two peaks, one at 12 pm noon and another between 10 pm and 11 pm. Dallas has a similar temporal profile to New York City, Los Angeles, and Chicago.

5. Heat map visualizations of Twitter activity

This section presents the heat map visualizations of the top 15 Twitter cities around the world. Figure 3 shows a series of heat map visualizations for the 15 cities. These heat maps show the intensity of Twitter activity in terms of time of the day and day of the year. The x-axes show day of year and the y-axes show time of day. Darker shades indicate higher levels of Twitter activity and lighter shades indicate a low level of Twitter activity.

The following is a city-by-city discussion of each of the heat map visualizations. We start our observations with Riyadh, Istanbul, and Jakarta because of their interesting resemblance of activity patterns during the month of Ramadan (the ninth month of the Muslim year, during which fasting is observed between dawn and sunset).

Riyadh: This was the most surprising visualization of all the cities. It appears that users mostly tweet between 12 pm and 3 pm. This trend appeared in the temporal activity graph (in Section 4) as well. We also find a dramatic shift in user activity at the end of July and during the initial days of August. The highest tweeting activity during this period is between 3 am and 7 am, and this corresponds to the month of Ramadan. During this period, there is no high tweeting activity at 12 pm. Therefore, users might sleep late during the night-time and may not wake up until midday. Muslims observe fasting during Ramadan from sunrise to sunset. Sunset is the time when they take the first opportunity to eat. Hence, at 7 pm (during the end of July and the start of August) tweeting activity is very low with an observable strip of lighter shades.

Istanbul: Turkey is a secular country, although the majority of the population is Muslim. Hence, we observe the same shift of high tweeting activity at 4 am during the month of Ramadan. Sunrise is earlier in Istanbul than Riyadh because it is at a higher latitude. However, the rest of the tweeting activity in Istanbul resembles Paris and London where we find a peak of activity during the day and another peak during the night-time.

Jakarta: Indonesia is another Muslim state. It is considered more secular than Saudi Arabia and less secular than Turkey. Jakarta is also the top Twitter city as far as number of posted geotagged tweets is concerned. There are two interesting patterns which could be observed from the heat map visualization of Jakarta. First is the time of the highest tweeting activity, which is between 4 pm and 8 pm. Second is the effect of Ramadan in Jakarta. There is a stripe of darker shades at about 4 am during the month of Ramadan. Also, tweeting activity starts later during the month of Ramadan.

Paris: The heat map visualization of Paris shows that the Twitter activity remains low from 2 am to 7 am. Late night Twitter activity increases dramatically during the summer time when users tweet until 3 am or 4 am. During the summer time, the activity starts late during November, January, March, and May. During these periods, we can observe tweeting activity until 3 am. Also, a late start in the activity during the daytime is also apparent.

Sao Paulo: The heat map visualization of Sao Paulo shows more late night activity between the months of
December and March, when users keep tweeting until 4 am. During these months, tweeting activity starts late in the daytime. This is also the case for the month of August where tweeting activity starts very late during the day. There is a peak of tweeting activity between 10 pm and 11 pm throughout the year.
The heat map visualization for NYC shows vertical stripe patterning which corresponds to weekly cycles of activity. Focusing on the morning hours, between 4 am and 8 am, Twitter activity is low. The pattern changes later in the day, and the lighter shades of the heat map become progressively darker. During the summer time, there is consistently high tweeting activity between 9 pm and 12 am throughout the year. This activity pattern is similar to that observed in other cities. During these months, tweeting activity starts late in the daytime. There is a major peak of tweeting activity which is during the day and night-time.

For example, the activity patterns in Tokyo are completely different to those in Riyadh and Madrid. For example, the activity patterns in Tokyo are completely different to those in Riyadh and Madrid.

The heat map visualization for Los Angeles shows a low activity between 2 pm and 7 am. The activity starts increasing after 7 am and remains high until late night.

Rio de Janeiro: Similar to the activity pattern of Buenos Aires and Sao Paulo, the activity pattern of Rio de Janeiro shows more late night activity between the months of December and March, when users keep tweeting until 4 am. During these months, tweeting activity starts late in the daytime hours. This is also the case for the month of August. There is consistently high tweeting activity between 10 pm and 12 am midnight throughout the year.

Mexico City: There is low Twitter activity in Mexico City between 3 am and 6 am. Twitter activity gradually increases after 6 am. An interesting pattern of high activity can be observed in December, when users tweet relatively late at night compared to the rest of the year. There is also high activity between 10 pm and 12 am.

Tokyo: Patterns of activity in Tokyo appear very different from other cities. It appears that users mostly tweet in the evenings and do not tweet a lot late at night. There is no substantial shift to late night usage in response to seasonal changes. Overall, user behavior is consistent throughout the year.

Chicago: The heat map visualization of Chicago shows a consistent pattern of late night activity throughout the year with some intervals of high activity. An interesting observation is the late night high activity at the end of March and during the summer time. In general, the activity is very high during the night-time, i.e., between 10 pm and 12 pm.

Dallas: Similar to Chicago, the heat map visualization of Dallas shows a consistent late night activity throughout the year with a further increase during the summer time. The tweeting activity in Dallas reaches a peak between 9 pm and 11 pm.

6. Conclusions and future work
Social media services generate large quantities of user-generated content, and the exploitation of this resource is increasing in the research and commercial sectors. The use of the Twitter social media services varies with different cities and countries. This paper has presented a case study where millions of records of raw data are transformed into simple visualizations that reveal a great deal about differences in Twitter user behavior in different cities.

The temporal graphs are the footprints of the activity in different cities. Heat map visualizations show the intensity of Twitter activity in terms of time of day and day of the year. From temporal activity graphs and heat map visualizations, we can suggest different overall patterns of the behaviors of the users in different cities. These include: weekly patterns of activity (sleeping hours vs. waking hours vs. work time); high and low usage times for the social media services; seasonal shifts of the activity; and the impacts of major behavior patterns such as Ramadan. We also observe a lot of differences between the activity patterns of different cities. For example, the activity patterns in Tokyo are completely different to those in Riyadh and Madrid.

There are a number of ways in which this work might be further developed in the future. One shortcoming of our analysis is the missing data between the periods of May, June, and July 2013. The current analysis could be improved by filling in the missing data. This could result in a new version of heat map visualizations which could show the activity patterns of different cities for a complete year. Another possible improvement is to compare this analysis of geotagged Twitter activity with that of the non-geotagged one to find the potential differences in the activity patterns.

Acknowledgments
This work was completed as part of the EPSRC research Grant “The Uncertainty of Identity: Linking Spatiotemporal Information in the Real and Virtual Worlds” (EP/J005266/1).

Notes on contributors
Muhammad Adnan is a postdoctoral research associate at the Department of Geography, University College London. His research focus is on Big Data analytics, data mining, and visualization of large spatio-temporal databases.
Alistair Leak is a research student at the Department of Security and Crime Science, University College London. His research relates to the analysis of large social media data-sets and their applications within geography.

Paul Longley is a professor of Geographic Information Science at University College London. His publications include 14 books and more than 125 refereed journal articles and book chapters. He is a co-editor of the journal “Environment and Planning B” and a member of five other editorial boards.

References
(1) Twitter. *What is Twitter?* [http://business.twitter.com/basics/what-is-twitter/](http://business.twitter.com/basics/what-is-twitter/) (accessed Dec 31, 2012).
(2) Huberman, B.A.; Romero, D.M.; Wu, F. Social Networks that Matter: Twitter Under the Microscope. *First Monday* 2009, 14 (1), 1–9.
(3) Neuhaus, F.; Webmoor, T. Agile Ethics for Massified Research and Visualization. *Information Commun. Soc*. 2012, 15 (1), 43–65.
(4) Quercia, D.; Ellis, J.; Capra, L.; Crowcroft, J. Tracking “Gross Community Happiness” from Tweets. *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work (CSCW)*, San Antonio, TX, 2012; pp 965–968.
(5) Kulshrestha, J.; Kooti, F.; Nikravesh, A.; Gummadi, K.P. Geographic Dissection of the Twitter Network. *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, Dublin, 2012; pp 1–8.
(6) Adnan, M.; Lansley, G.; Longley, P.A. A Geodemographic Analysis of the Ethnicity and Identity of Twitter users in Greater London. *Proceedings of the 21st Conference on GIS Research UK (GISRUK)*, Liverpool, 2013; pp 1–6.
(7) Pennacchiotti, M.; Popescu, A.M. A Machine Learning Approach to Twitter User Classification. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, 2011; pp 281–288.
(8) Quercia, D.; Seaghdha, D.O.; Crowcroft, J. Talk of the City: Our Tweets, Our Community Happiness. *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, Dublin, 2012; pp 555–558.
(9) Rios, M.; Lin, J. Visualizing the Pulse of World Cities on Twitter. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, Cambridge, 2013; pp 717–720.
(10) Bennett, S. Revealed: The Top 20 Countries and Cities of Twitter [STATS]. [http://www.mediabistro.com/alltwitter/twitter-top-countries_b26726](http://www.mediabistro.com/alltwitter/twitter-top-countries_b26726) (accessed Dec 31, 2012).
(11) Twitter. *The Streaming APIs*. [https://dev.twitter.com/docs/streaming-apis](https://dev.twitter.com/docs/streaming-apis) (accessed Jan 22, 2012).