Analysis of Twitter Usage in London, Paris, and New York City

Muhammad Adnan
Department of Geography, University College London, Pearson Building, Gower Street, London, United Kingdom. m.adnan@ucl.ac.uk

Paul Longley
Department of Geography, University College London, Pearson Building, Gower Street, London, United Kingdom. p.longley@ucl.ac.uk

Abstract

The penetration and use of social media services differs from city to city. This paper analyses the use of Twitter in London, Paris and New York. We develop a geographical analysis of Tweeting activity, and of the names, probable ethnicities and genders of Twitter users in these three cities. The results reveal that the majority of Twitter users are male and are of Anglo-Saxon (English speaking) extraction. The analyses also provide an insight into the ethnic diversity of populations in the three cities.

Keywords: Twitter, User generated data, Name and Ethnicity Analysis

1 Introduction

Social media services generate large quantities of data every day. The use of these data is increasing in the research and commercial sectors as recent years have seen an increase in the use of social media in service delivery. Users of the likes of Twitter, Facebook, Flickr, LinkedIn, Bebo, Sino Weibo, and Orkut are frequently mobile users of the developing range of smartphones and tablet devices. These users come from different social and ethnic backgrounds, though it is unclear how representative they are of the population at large.

This paper focuses upon 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 [9]. 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.

Previous research on Twitter data has emphasised different themes, such as the investigation of Tweeting behaviour [2], the dissemination of information across social networks [1], ethical issues in the analysis of Twitter data [4], tracking community happiness from Tweets [3], and geographic dissection of the Twitter network [6]. However, name, ethnicity, and gender analysis of Twitter usage remains largely unexplored. Previous research has shown that names provide indicators of the cultural, ethnic and linguistic characteristics of individuals ([8]). Naming and ethnicity analysis of georeferenced social media data can provide insight into the geographical structure of different ethnic groups who use a particular social media service. Gender analysis can ascertain whether male users use Twitter more than females or vice versa. Hence, this paper is aimed at analysing the name, ethnicity and gender of Twitter users.

In 2012, London, Paris, and New York City were amongst the top 10 cities with regard to the amount of Tweeting [7]. Hence, this paper undertakes an analysis of name, ethnicity and gender of Twitter users in these three cities. The structure of this paper is as follows.

Section 2 of this paper describes the data used in the analysis. The paper begins by analysing the areas of high Tweeting activity (Section 3), and goes on to analyse hourly and weekly Twitter usage (Section 4). Section 5 introduces a name and ethnicity analysis of Twitter users in the three cities. Section 6 presents a gender analysis of Twitter usage. Finally, conclusions and directions for future research are presented in Section 7.

2 Data

Twitter users produce a huge amount of data every day. Twitter provides the facility for programmers and developers to download live Twitter data using its public streaming API ([10]). The API provides a 1% sample of live geo-tagged Tweets within a user specified bounding rectangle, at any point in time.

For this paper, the Twitter Streaming API was used to download geo-tagged Tweets for London, Paris, and New York City over the period 10 September – 20 December 2012 (102 days). The numbers of geo-tagged Tweets for each city were:

London: 2, 412, 252
Paris: 740, 188
New York City: 646, 053

For this paper, the Twitter Streaming API was used to download geo-tagged Tweets for London, Paris, and New York City over the period 10 September – 20 December 2012 (102 days). The numbers of geo-tagged Tweets for each city were:

London: 2, 412, 252
Paris: 740, 188
New York City: 646, 053
The data downloaded from the API included the 'User Name', 'Default Profile Language', 'Latitude of the Tweet', 'Longitude of the Tweet', and 'Tweet Message'. For the analysis reported in this paper, we used 'User Name', 'Latitude of the Tweet', and 'Longitude of the Tweet' fields.

In every city, some users send more Tweets than others. The number of unique users who sent Tweets is:

- London: 140,919 users
- Paris: 42,729 users
- New York City: 59,272 users

3 Areas of high Tweeting Activity

Areas of high relative incidence of Tweeting activity shown in Figures 1 – 3 (given at the end of this paper).

The maps were created by using the following procedure:

a) In the first step, a grid map of 150 rows and 150 columns was created for every city. This resulted in 24,025 individual grid squares for each grid map.

b) Using the Twitter data ('latitude of the Tweet' and 'longitude of the Tweet') for every city, a point in polygon operation was performed to count the number of Tweets sent from the area of each grid cell.

c) Finally, the individual grid cells were given a colour based on the number of Tweets sent from them.

In London (Figure 1: given at the end of this paper), the users in the central part of the city sent more Tweets than the users in the rest of the city. This may be because of the high number of Twitter users travelling to the centre of the city for work, shopping, and visiting purposes. The outskirts of London have low Twitter usage. However, in London, there are few areas where users do not send any Tweets.

In New York City (Figure 2), the Twitter users in certain areas are more active than others. For example, there is high Twitter usage in lower and mid-town Manhattan, West Bronx, and in certain areas of Brooklyn. Tweeting activity in some areas (Queens and Staten Island) is very low.

In Paris (Figure 3), 25 or fewer Tweets were sent from many grid squares. The central area of Paris accounts for more Tweeting activity than rest of the areas. In Paris, like London, there are few areas from where users do not sent any Tweets.

4 Hourly and Daily Twitter Activity

The previous section has discussed the areas of high Tweeting activity in London, Paris, and New York City. In this section we break down this over-all picture by time of day and time in the week.

Figure 4 (given at the end of this paper) shows the hourly Twitter activity in London, Paris, and New York City. This figure was created by aggregating the number of Tweets, by hours of the day, sent during 10 September to 20 December, 2012.

Tweeting activity in London remains low between 2AM to 5AM. It starts to increase from 5AM which is earlier than in the other two cities. Tweeting activity in London reaches its peak between 10AM – 11AM. In the night time, Twitter usage is very high between 7PM – 11PM.

Tweeting activity in Paris remains low between 2AM to 6AM. It starts increasing from 6AM. There is high Tweeting activity between 10AM – 1PM, which is different from London. At night, Twitter usage is very high between 7PM – 11PM.

Similar to Paris, Tweeting activity in New York City remains low between 2AM to 6AM. It starts increasing from 6AM. There is high Tweeting activity during 10AM – 2PM. At night, Twitter usage is very high between 7PM – 11PM.

Figure 5 shows the daily Twitter activity patterns of London, Paris, and New York City. All three cities have higher numbers of Tweets on Wednesday and Thursday. The number of Tweets sent on Sunday is less than the number of Tweets sent on any other day of the week.

5 Name and Ethnicity Analysis of Twitter Users

A name is a statement of a bearer's cultural, ethnic, and linguistic identity [8]. Knowing the name of the person can thus reveal other useful information about that person. Name analysis of Twitter users can give us an insight into the ethnic identity of individuals, which could be useful in identifying the different types of users who use this social media service.

When registering to use Twitter, users are required to enter their name or other identifying data in the 'User Name' field. In many cases, tokens other than given and family names are entered, as in 'Justins_Home', 'What is Love', etc. However, many registrants fill this field with their real forename-surname pairs.

For the three cities, we developed our own software to divide the 'User Name' field into separate 'forename' and 'surname'
fields. The following table (1) shows the result of our text analytics work, where recognisable forename-surname pairs were found in the 'User Name' field.

Table 1: Result of the text analytics work

<table>
<thead>
<tr>
<th>City</th>
<th>Total number of users</th>
<th>Number of users where forename-surname pairs found</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>140,919</td>
<td>106,917</td>
</tr>
<tr>
<td>Paris</td>
<td>42,729</td>
<td>28,775</td>
</tr>
<tr>
<td>New York City</td>
<td>59,272</td>
<td>42,674</td>
</tr>
</tbody>
</table>

The next step describes the Ethnicity Analysis performed on the users where forename-surname pairs were found.

5.1 Ethnicity Analysis

Onomap [8] is software which may be used to assign a predicted ethnic group to a forename-surname pairing. Onomap was run on the forename-surname pairs of Twitter users. The following table shows the number of users for each city, where an ethnic group was successfully assigned to the forename-surname pairs.

Table 2: Ethnicity analysis of twitter users in the three cities

<table>
<thead>
<tr>
<th>City</th>
<th>Number of Users</th>
<th>Number of users where an ethnic group was found</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>106,917</td>
<td>99,974</td>
</tr>
<tr>
<td>Paris</td>
<td>28,775</td>
<td>23,450</td>
</tr>
<tr>
<td>New York City</td>
<td>42,674</td>
<td>39,110</td>
</tr>
</tbody>
</table>

For London, 69 ethnic groups were found. The following figure (6) shows the top-10 ethnic groups of Twitter users in London.

In London, ‘ENGLISH’ Twitter users are in very high numbers (51%). The users of second most common group (‘IRISH’) account for 4% of Tweets. The graph gives us an indication of long established and recent immigrants living in London, as indicated by the considerable number of Twitter users from the ‘SPANISH’ (2.25%), ‘PAKISTANI’ (1.9%), ‘INDIAN’ (1.56%), ‘PORTUGUESE’ (1.34%), and ‘TURKISH’ (1.024%) ethnic groups.

For New York City, 67 ethnic groups were found. The following figure (7) shows the top-10 ethnic groups of Twitter users in New York City.

5.2 Top Surnames

This sub-section is aimed at identifying the top-10 surnames of Twitter users and their ethnic groups in London, New York City and Paris. The following tables (3-5) list the top-10 surnames on Twitter in the three cities.
Gender Analysis

Forename can also be used to determine the gender of a person. GenderChecker [5] is a database which allows checking whether a forename is male, female, or unisex. Forenames of Twitter users, retrieved in section 5, were assigned a gender by using the GenderChecker database. The users were assigned 'Male', 'Female', or 'Unisex' genders. 'Not Found' was assigned when a forename was not found in the database. The result of the analysis is shown in the following figure (9).

Following table (6) lists the absolute numbers of different genders used in the analysis.

Table 6: Gender Analysis of the three cities

<table>
<thead>
<tr>
<th>Gender</th>
<th>London</th>
<th>Paris</th>
<th>New York</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>53,872</td>
<td>12,779</td>
<td>20,116</td>
</tr>
<tr>
<td>Female</td>
<td>36,180</td>
<td>9,175</td>
<td>14,541</td>
</tr>
<tr>
<td>Unisex</td>
<td>10,101</td>
<td>1,939</td>
<td>4,435</td>
</tr>
<tr>
<td>Not Found</td>
<td>6,764</td>
<td>4,882</td>
<td>3,582</td>
</tr>
<tr>
<td>Total Users</td>
<td>106,917</td>
<td>28,775</td>
<td>42,674</td>
</tr>
</tbody>
</table>

The graph, in figure 9, shows that almost 50% of the users are Male, and 33% of the users are Female in the three cities. This shows a dominance of Male users on Twitter. In all the three cities, less than 10% of users were assigned to the 'Unisex' category.

7 Conclusion 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 between different cities and countries. This paper has presented a preliminary descriptive analysis of Tweeting activity, specifically according to location, name, ethnicity and gender of Twitter users in the cities of London, Paris, and New York City. The analysis suggests that the majority of Twitter users are male users and have Anglo-Saxon roots (i.e. English speaking). The analysis also provides an insight into the ethnic diversity of population in the three cities. However, it must be remembered that the sources and operation of bias in our (very large) dataset are unknown: by no means everybody Tweets, and there is no a priori reason to assume that
Tweeters who disclose their locations are representative of those that do not.

This is preliminary research in a promising research area and could be extended in different ways. One possible extension is a detailed geo-spatial analysis of different ethnic groups in the three cities by the identification of their residential and work place areas. The second possible extension is to conduct a country wide analysis of Twitter users in United Kingdom, United States of America, and France. This kind of analysis can give us a detailed insight into the identity of Twitter users and how they use this particular social media service.

8 Acknowledgements

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).

References


Figure 1: Tweet Density Map of London

Figure 2: Tweet Density Map of New York City
Figure 3: Tweet density map of Paris.

Legend

- 0
- 1 - 25
- 26 - 50
- 51 - 200
- 201 - 300
- 301 - 7181

Figure 4: Hourly Twitter usage in London, Paris and New York City