



HAL
open science

Contribution of geotagged Twitter data in the study of a social group's activity space

Alexandre Cebeillac, Yves-Marie Rault

► To cite this version:

Alexandre Cebeillac, Yves-Marie Rault. Contribution of geotagged Twitter data in the study of a social group's activity space: The case of the upper middle class in Delhi, India.. NETCOM: Réseaux, communication et territoires / Networks and Communications Studies, 2016, 30-3/4, pp.231-248. 10.4000/netcom.2529 . hal-03952073

HAL Id: hal-03952073

<https://hal.science/hal-03952073>

Submitted on 23 Jan 2023

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.



Netcom

Réseaux, communication et territoires

30-3/4 | 2016

Réseaux sociaux et territoires

Contribution of geotagged Twitter data in the study of a social group's activity space

The case of the upper middle class in Delhi, India

Contribution à l'étude de l'espace d'activité d'un groupe social des données

Twitter géo-localisées. Le cas de la classe moyenne supérieure à Delhi, Inde

Alexandre Cebeillac and Yves-Marie Rault



Electronic version

URL: <http://netcom.revues.org/2529>

ISSN: 2431-210X

Publisher

Netcom Association

Printed version

Date of publication: 16 December 2016

Number of pages: 231-248

ISSN: 0987-6014

Brought to you by Bibliothèques de
l'Université Paris Diderot - Paris 7



Electronic reference

Alexandre Cebeillac and Yves-Marie Rault, « Contribution of geotagged Twitter data in the study of a social group's activity space », *Netcom* [Online], 30-3/4 | 2016, Online since 21 March 2017, connection on 27 March 2017. URL : <http://netcom.revues.org/2529>



Netcom – Réseaux, communication et territoires est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

**CONTRIBUTION OF GEOTAGGED TWITTER DATA IN THE
STUDY OF A SOCIAL GROUP'S ACTIVITY SPACE. THE CASE OF
THE UPPER MIDDLE CLASS IN DELHI, INDIA**

**“CONTRIBUTION A L'ETUDE DE L'ESPACE D'ACTIVITE D'UN
GROUPE SOCIAL DES DONNEES TWITTER GEO-LOCALISEES.
LE CAS DE LA CLASSE MOYENNE SUPERIEURE
A DELHI, INDE”**

CEBEILLAC ALEXANDRE¹, RAULT YVES-MARIE²

***Abstract** - Since 2009, the social media Twitter provides free access to localization data consisting of geotagged messages sent on the platform. We used this data to highlight their potential contributions to the study of social groups' activity space. Using a sample of luxury shoppers, we analyzed the daily activity and mobility patterns of the upper middle class in Delhi, India. We intersected these data with the results of an ethnographic study carried out in two luxury shopping places in Delhi and found that the lower middle class was overrepresented in the Twitter sample. We argue that the varying motivation to post on social media tend to skew the representativeness of social groups sampled on a spatial criterion. We eventually advocate the complementary use of geotagged social media data and qualitative data based on ethnographic research.*

***Keywords** - Activity space, Twitter, social networking service, social media, geotagged data, luxury shoppers, upper middle class, Delhi, India.*

***Résumé** - Depuis 2009, le réseau social en ligne Twitter garantit un accès libre à des données de localisation provenant des messages envoyés sur sa plateforme. Dans le cadre d'une étude sur la classe moyenne supérieure indienne, nous avons utilisé ces données pour analyser les espaces d'activité des consommateurs de luxe à Delhi. En croisant les données Twitter avec les résultats d'une étude ethnographique menée dans deux complexes commerciaux de luxe à Delhi, nous interroge les potentiels usages et les limites des données issues des réseaux sociaux dans l'étude des groupes sociaux. Nous*

¹ University of Rouen, alexandre.cebeillac@etu-univ-rouen.fr

² Paris Diderot University, yves.marie.rault@gmail.com

avançons que les différentes motivations sociales à tweeter biaisent la représentativité d'un échantillon sélectionné sur une base spatiale. Nous défendons alors l'usage de données qualitatives issues de recherches ethnographiques en complément des données géo-localisées.

Mots-clés - *Espaces d'activité, Twitter, réseaux sociaux, données géo-localisées, consommateurs du luxe, Delhi, Inde.*

INTRODUCTION

The number of single smartphone users in India has increased considerably in the past decade, particularly in urban areas. These devices facilitate the obtaining of data the geographic position of users, which several applications use to provide personalized services. In India, where data on mobility or activity spaces is often difficult to find, the increasing use of smartphones offer new possibilities for research in geography and social sciences; however, this data is often the property of private service providers, and public access to the data is limited. Twitter (since 2009) has been a significant exception, and provides the most comprehensive database of user locations. It thus allows new possibilities for research on activity space, defined in 1971 by Horton and Reynolds as the “*the subset of all urban locations with which the individual has direct contact as the result of day-to-day activities*” (Perkins et al., 2014).

While several studies have already used Twitter datasets, very few of them have dealt with spatiotemporal and human daily mobility analysis (Steiger et al., 2015). Nevertheless, some studies have highlighted the feasibility of using geotagged tweets to observe human mobility patterns at a global scale (Hawelka et al., 2014) at a country scale (Jurdak et al., 2015; Liu et al., 2015), and at a city scale (Cebeillac et al., 2017). However, the socio-spatial picture depicted by these studies does not provide an accurate picture of human mobility. Twitter users are never, at the global scale, or at the country scale, representative of socially diverse populations due to several reasons. First of all, Twitter users are unevenly distributed over space. In 2016, Twitter claimed 316 million active users worldwide, with a penetration rate more than ten times higher in South America than in Africa and in India (Hawelka et al., 2014). Secondly, they are unevenly distributed over the social spectrum, with notably certain professions and social classes having a more regular use than others (Preoțiuc-Pietro et al., 2015).

To our knowledge, only one study has used geotagged Twitter data to observe a particular social group in its urban setting. However, the sample used was selected on the basis of a common ethnic and racial background estimated through nicknames and messages analysis (Luo et al., 2016).

In this article, we identified a social group, the upper middle class, based on its presence in one of its places of leisure. We explore the possibilities offered by these datasets to study a specific social group's activity space. In India, the use of Twitter is

particularly well segmented along socio-economic lines, since Twitter requires a good command of the English language and the ownership of a smartphone with 3G or 4G network. Only a quarter of the population is able to buy consumer goods beyond basic necessities (Sridharan, 2011). While 350 million persons use the Internet at least once a year, 97 million of them are active social media users on their mobile phones³, among which only 46 million individuals have a Twitter account (3,7% of the total population), restricted mostly to urban areas.

As a consequence, we assumed that members of the urban Indian upper middle class, being the ablest to use Twitter, would be a relevant social group for our study. In order to find a representative sample of this class and to study their activity space, we chose to focus on the luxury shoppers in Delhi, one of the top Indian metropolitan city by its GDP per capita. We selected the shoppers of two luxury shopping places in South Delhi, located near the most expensive residential complexes of the city; the Khan Market, a high-street, and the DLF Emporio, a shopping mall. These are two of the most expensive retail destinations in India, with a monthly rental of 1 000 rupees per sq. ft. and 1 250 rupees per sq. ft., respectively⁴.

Our study was carried out in two stages. Firstly, in early 2015, we conducted on-site observations and surveys as well as off-site interviews with the users, inquiring about their socio-economic characteristics, place of residence, and places of leisure. The study notably revealed that the two luxury places did not only welcome ultra-high-net-worth individuals, but middle-income consumers as well. Simultaneously, we built a database of luxury shoppers' geotagged locations in Delhi based on a group of people who tweeted at least once in one of the two sites selected for the study.

Combining geotagged data and qualitative data, we explored the possibilities of using geotagged Twitter data in researching social groups. Our study also highlights the drawbacks of Twitter data, especially when not used along with ethnographic research. We will develop our argument in three parts. Firstly, we will thoroughly explain the construction of our database. Secondly, we will critically examine the possibilities it provides for the study of activity space and human mobility. Thirdly, we will demonstrate that the various social motivations to tweet create biases in social groups' representativeness, and eventually argue for the complementary use of geotagged social media and ethnographic data.

1. THE DATABASE: LOCATION, USER, AND DATE OF TWEETS

Twitter is a micro-blogging service on the Internet, where people can send short messages or "tweets". These messages can be private (shared with friends) or

³ Digital, Social & Mobile in India in 2015, WeAreSocial, 2015, <http://wearesocial.sg/?s=india>, consulted on 15.07.2016.

⁴ Cushman & Wakefield, *New Retail Frontiers: Emerging Main Streets in India*, 2014.

public (visible to everyone). Since 2009, it became possible to add the geolocation of the device via “geotagging” of the message, even if this option is blocked by default⁵. The Twitter Streaming API⁶ allows collecting in real-time of a sample of ~1 % of the total feed of public tweets for free (Morstatter et al., 2013; Wang et al., 2015). In this sample, around 10% of the tweets are geotagged, giving us information on the spatial location of the user, with an accuracy depending on the device - up to 30 metres for a smartphone outdoor (Zandbergen et Barbeau, 2011), and up to few kilometres for a computer far from the access point of the Internet service provider.

Between 25 June 2014 and 4 December 2015 (524 days), we have been able to collect a total of 13 672 354 geo-located and dated tweets (with a minute accuracy), for 507 482 unique users in India. However, Twitter users are not always “real persons”. The company admitted in 2014 that 8.5% of the active users connect through a tierce service application⁷. Since the platform includes a large worldwide audience, it is a good platform for promotion, advertisement or geo-marketing, and is actively used by organizations to bring forwards their products, ideas or causes. Most of the time, these organisations hire a community manager or create a “bot” (twitbot) that can animate the social network of the Twitter account owner (Ferrara et al., 2015). The geotagged tweets inform us on the ID of the user, but not on its characteristics. Consequently, we had to remove the non-human tweets from our database. We based our bot detection algorithm on the speed between two messages sent by a user so that they are slower than a plane speed (Hawelka et al., 2014; Jurdak et al., 2015). But sometimes, two tweets can be sent at the same time from different locations, probably when there are network problems. In that case, the speed is infinite. We considered this bug and chose that users with more than 2% of their tweets sent at a speed higher than 600km/h would be defined as bots. We also excluded users with an average speed higher than 70km/h in urban areas, as this speed is rarely reached in locations like Delhi. We removed all the IDs consisting of 12 hexadecimal digits because they can only be MAC addresses (computers), and all the IDs of 15 hexadecimal digits, which belong to beacons, devices used in geo-marketing. Eventually, we removed all the users with a very high tweet frequency, i.e. higher than 20 tweets a day on an average, which would mean around 2000 public tweets per day, as we could only collect 1% of the tweets.

After removing the robots, we also decided not to consider low activity Twitter users, i.e. with less than 15 active days of tweets, assuming that these users would provide enough locations to perform accurate analysis of the activity spaces. The cleaned database included 5 720 169 tweets for 48 637 unique users accross India – respectively 41,8% and 9,58% of the original database. Eventually, in order to keep our focus on Delhi, we selected the tweets that were sent within a 30 km radius of Delhi NCR, generating a final database of 967 379 tweets for 13 349 users.

⁵ Twitter Terms of Service (<https://twitter.com/tos?lang=en>).

⁶ <https://dev.twitter.com/streaming/overview>, consulted on 02.05.2016.

⁷ https://www.sec.gov/Archives/edgar/data/1418091/000156459014003474/twtr-10q_20140630.htm, consulted on 03.05.2016.

2. POTENTIALS OF GEOTAGGED TWITTER DATA

2.1. Study of influx

The geotagged tweets, including the date and time when the tweets were sent, provide us with accurate information on influx periods of Twitter users in select locations. In our study, this information proved helpful in finding out the preferred times and days individuals chose for visiting luxury shopping places. The Figure 1 shows the number of users in the two commercial complexes per hour of the week. In the Khan Market, we can see a moderate visiting peak around noon on the weekdays and a major peak in the evening. The data seems aligned to our observation that the high-street is a popular venue for lunch and dinner among professionals and government officials, due to the proximity to their offices. However, the Emporio, bordering two affluent districts in Delhi, does not attract such a clientele. Outside of lunch and dinner hours, the scarce population of users is mostly composed of women often accompanied by friends or relatives, sometimes for a “kitty party”, a meeting between housewives over meal or coffee. In both places, Friday evenings and weekends have the largest influx of visitors, mostly families and youngsters. The data broadly confirm what was observed through qualitative research.

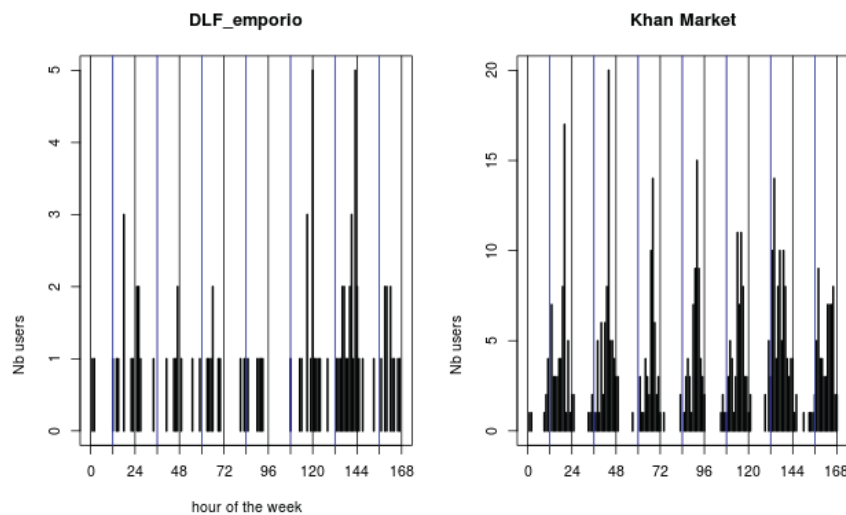


Figure 1: Visiting frequency of Twitter users in two luxury shopping places.

2.2. Study of activity space

Place, time, and frequency of visits are the three most important attributes of the activity space of an individual, defined as the set of daily places of living, working, and recreation. To locate such places, most studies adopt methodologies involving questionnaires with open or multiple choice questions. However, self-reported responses to such questionnaires often turn out to be inaccurate. For instance, in the case of our initial study including face-to-face interviews, the users were asked about the places where they usually go for shopping in Delhi. The interviewees gave

incomplete answers and could only recall a few major places. Usually, they would give a handful of major locations, from the scale of a shop to the scale of a shopping mall, without providing precise information on the frequency of the visits. On the contrary, the Twitter dataset allows to precisely know about the frequency of visits to places whose functions can be identified using OpenStreetMap (OSM), a mapping community project that contributes to the mapping of roads, railways, and all kind of places in the world (www.openstreetmap.org). The OSM database notably helped us to highlight the functions of each user location, which we obtained by aggregating tweets within a spatial radius using the same method as Jurdak et al. (2015) a dbscan approach introduced by Ester et al., 1996 (described in Figure 2).

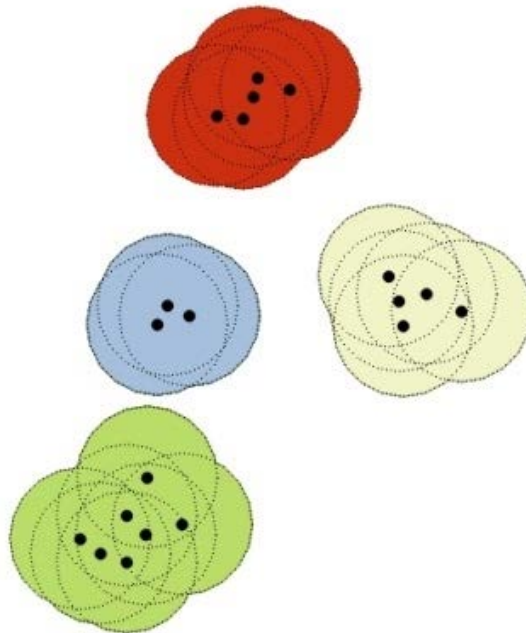


Figure 2: *A theoretical activity space of a user, where each location is aggregated into a cluster using the dbscan method. In this case, 20 locations are merged in 4 clusters (or activity centers).*

We created clusters of spatial points (tweets) that are all located within a certain distance, 50 meters for this study, implying that in every cluster location, all points should be at an inferior or equal distance to 50 meters. Figure 3 shows the percentage of users having a specific number of locations in their activity space (using a log/log plot). Almost the same ratio has from 1 to 7 locations, but it decreases rapidly after 10, suggesting that most users visit few different places, with a little amount of them having tweeted in many different locations. It is noteworthy that most Twitter users have few places in their activity space.

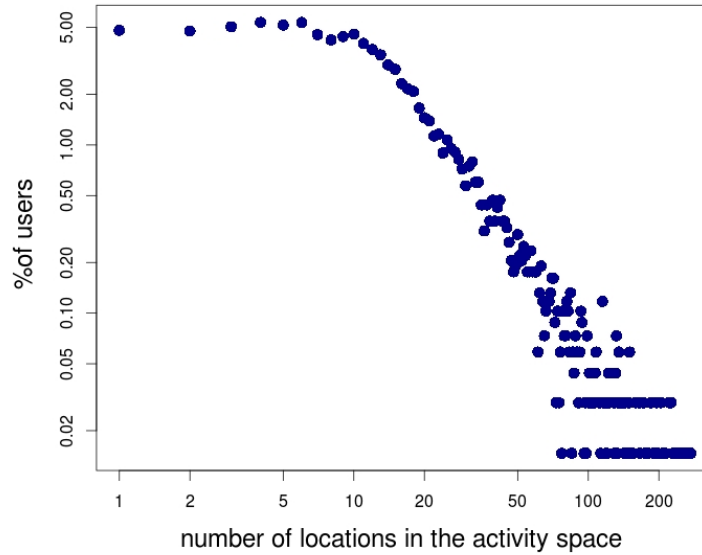


Figure 3: Number of locations in the activity space per user (log/ log).

Among all the tweets sent in Delhi, 11% were located in an OSM spatial polygon (55% on a major road, 45% in another kind of place). Due to differences in size, function and public of each place, and the absence of several places in the OSM database, we found significant differences in the amount of geotagged tweets in each location.

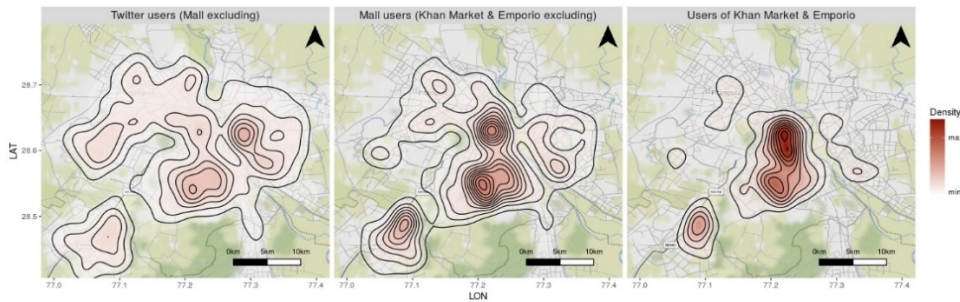


Figure 4: Twitter users density between 7 and 8pm.

From left to right: users who never visit shopping malls, users who visit shopping malls (Khan Market and Emporio excluded); users who visit Khan Market and/or DLF Emporio.

Sources: Twitter dataset; Made with R using ggplot2 and ggmap.

In order to compare the activity spaces of the luxury shoppers with other Twitter users, we created three kinds of datasets: (i) Twitter users that don't visit malls, (ii) Twitter users visiting shopping malls in Delhi (except Khan Market and DLF Emporio) and (iii) Twitter users visiting Khan Market and DLF Emporio. When gathering the clusters maps of Twitter users at different time periods of the day, we can

observe the “urban pulsations”, showing mobility and activity patterns in Delhi. If we compare the three snapshots of geotagged tweets sent from 7 to 8 pm, we observe that people have different mobility and activity patterns (Figure 4). For instance, while the map with all Twitter users shows that they are wide-spread across the city, they tend to be more narrowly clustered for the shopping mall and luxury shoppers. The latter are less present in East Delhi and Dwarka (middle-income suburbs), while they are more present in South Delhi and Gurgaon, the two most expensive residential areas of Delhi region.

In each Twitter user category, people exhibit different preferences for shopping locations. Drawing on our sample of people who had tweeted at least once in the two luxury commercial places, we were able to identify four shopping places that these users often visit too (see Figure 5, Figure 6), two high-end shopping mall and two high streets.

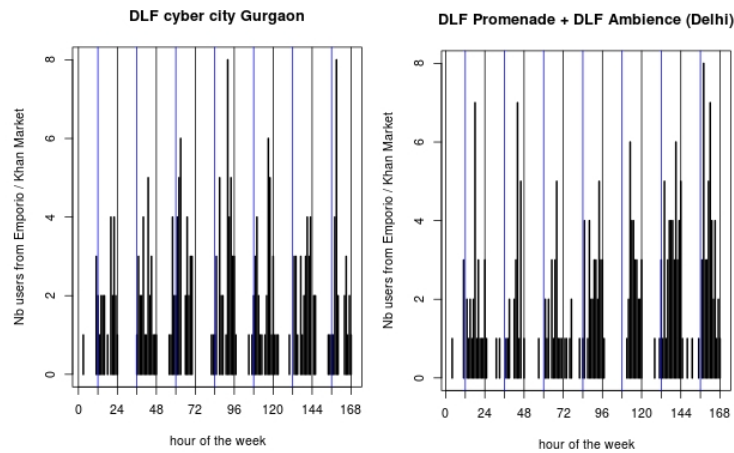


Figure 5: Visiting frequency of luxury shoppers in the two top shopping malls in terms of total frequentation.

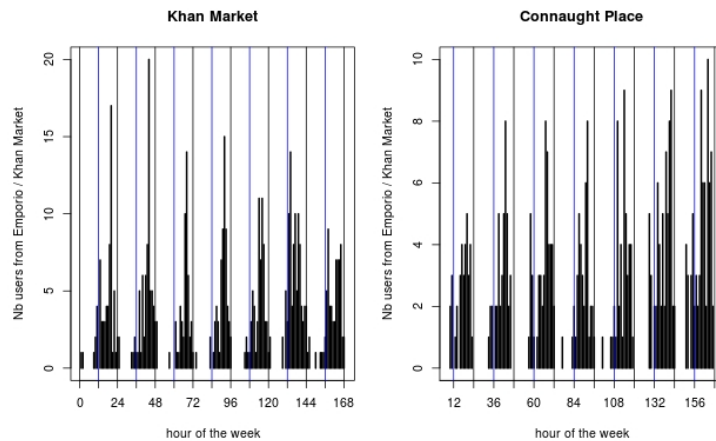


Figure 6: Visiting frequency of luxury shoppers in the two top high streets in terms of total frequentation.

The two high streets (Connaught Place and Hauz Khas village) also include walking spaces, historical monuments, and parks. Consequently, the tweets might not always be sent by consumers but also by visitors and walkers. This bias could be avoided through questionnaires. However, using geotagged Twitter data has the benefit to avoid imposing answers, for instance through the selection of a few places on the assumption that they are the most visited. However, it remains important to look at the configuration of each place to understand discrepancies in results. For instance, the Khan Market is a “high street”, a commercial concept consisting of a pedestrian street composed of various shops. On the other hand, the DLF Emporio is a “shopping mall”, i.e. a commercial form consisting of one or several buildings forming a complex of shops connected by air-conditioned walkways enabling customers to walk from stores to stores without going outside. While the mall mostly includes apparel and accessories stores with a large number of flagship international brands like Louis Vuitton or Chanel, located on the first floor and advertised on the building's facade, the Khan Market is a multifunctional high-street where we find garment stores, jewelers, book and electronic stores as well as grocery shops and restaurants, with a large price range, even though the products and services are relatively very costly. A study of 6 markets in Delhi undertaken by Sarma in 2005 (Sarma, 2006) also showed that locations of the shopping places in the urban grid explain large differences in consumer demographics. The Khan Market, centrally located on important routes of the city, has a high potential for attracting long range consumers from vehicular, bus and metro traffic, as well as several passing local pedestrian traffic due to its good connections to walkways, and the capacity to support vehicular and pedestrian movement on site and provide with a large quantity of parking space. This is not the case for the DLF Emporio which has been configured largely to host a public of car-owners.

2.3. Study of places of residence

Amongst all the places in the activity space, those which provide the most valuable information on the social backgrounds are probably the places of residence. We explained earlier how we were able to define clustered locations for each user. In order to identify the residential places, we created an algorithm to find out which one of them were homes of the Twitter users. Firstly, candidate locations could not be in places considered by OSM as leisure places, malls, park, or road. Consequently, we removed all those places. Then, we adapted the method used by Ahas et al., 2010 and Calabrese et al., 2013, defining home locations as the cluster where the users were active during the most different days, and especially during the night hours – 9 pm to 7 am (Cebeillac et al, 2017). Since the Twitter database is not representative of all Delhi inhabitants, this algorithm just provide an estimation of home locations of Twitter users. However, we were able to detect the home of 6324 users in Delhi for 881333 tweets in 130308 locations.

If we look at the entire Twitter dataset, we can observe four main poles, namely Gurgaon, South Delhi, East Delhi and West Delhi, with higher densities in South and East Delhi (Figure 7), with reducing density in the shopping mall dataset (Figure 7). As for the luxury shoppers dataset (Figure 7) accounting for only 164 users who tweeted in the Khan Market or in the DLF Emporio, it is noteworthy that two poles appear

more clearly in South Delhi and Gurgaon, the most expensive areas of the city. On the contrary, the middle-income areas of West and East Delhi (Noida excepted) are not on the map anymore.

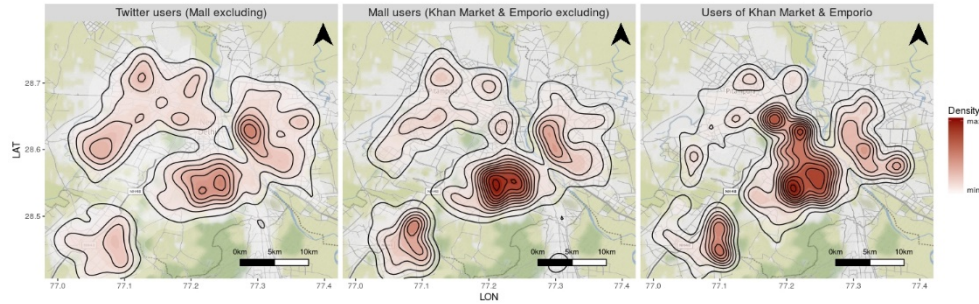


Figure 7: Home density of Twitter users.

From left to right: Twitter users who never visit malls; Twitter users who visit malls (Khan Market and Emporio excluded); Twitter users who visit Khan Market and/or Emporio. **Sources:** Twitter dataset. Made with R using ggplot2.

In order to estimate the social class of individuals of our two samples, we've linked the home locations (declared during the survey and estimated for Twitter users) with the property tax level, ranging from H (minimum) to A maximum) for 2106 colonies in Delhi in 2010 (Figure 8). We assume that the property tax is a good proxy and provide valuable information on the socioeconomic categories of the users, even if it is worth noting that urban villages like Hauz Khas village tend to have a very low tax level even though they host significant numbers of high-income households. Indeed, in wealthy neighbourhoods, active neighbour welfare associations tend to militate to downgrade the rate of their property tax (Telle, 2011).

Finally, 131 home locations from the survey were defined at the colony wise level. It is relatively small due to the inaccuracy of the locations given by the interviewees. On the other hand, 164 home location were estimated for the Twitter users that visit Khan Market and Emporio.

In order to compare the two dataset with Delhi global population, we've estimated the population per colony (as the population is only available ward wise). To do so, we've detect built-up areas (using Kmeans algorithm on Landsat-8 images) and merged them with the 2011 ward-wise census, assuming a homogeneous population per ward. We've created a grid of 250m cells with the population estimation that we've intersected with the colony shapefile.

Figure 9 allow to compare the percentage of population from the survey and from the census that lives in a colony with a specific property tax category. It shows that less of 10% of the Delhiites live in wealthy colonies (A or B), this percentage rise to more 80% for our survey sample, confirming that customers of Khan market and Emporio lives in the most expensive areas of the city. While 70% of Delhi inhabitants

stays in quite low property tax colonies (from D to H), it concern only few interviewees, and they lives in wealthy urban villages.

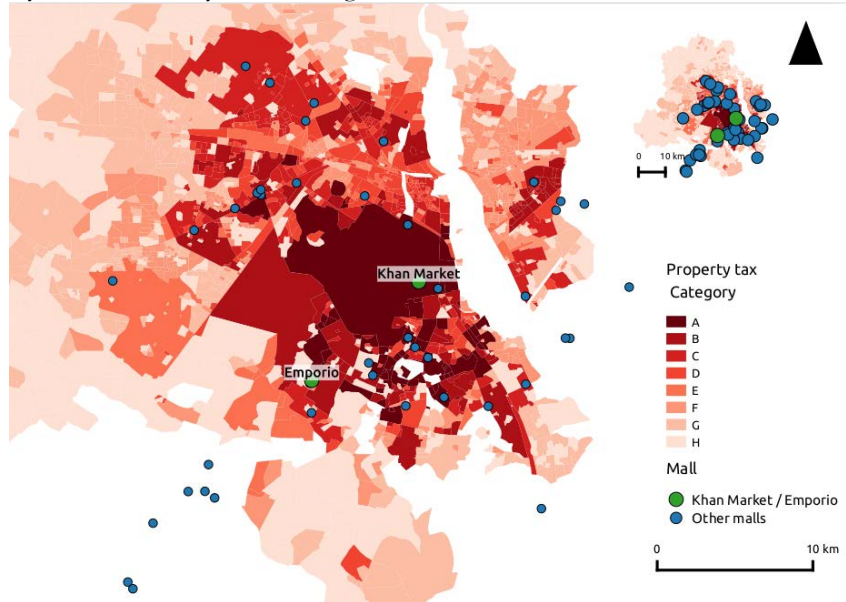


Figure 8: Map with property tax categories and location of shopping malls in Delhi.

Sources: www.mcdpropertytax.in, provided by O. Telle et B. Lefevbre. Made with qgis.

Distribution of survey samples according to property tax



Figure 9: Percentage of Delhi inhabitants and persons from the survey in each property tax category. From a dasymetric map of the census 2011.

If we compare the results of the survey to the Twitter database (Figure 10), we can notice that sample of Twitter users is more wide-spread among the different property tax categories. It thus clearly appears that there is an under-representation of

high-income users in the Twitter database compared to the interviewees (Figure 10). Nevertheless, in both cases, luxury shoppers are over-represented in high-income areas as compared to the actual population who live there.

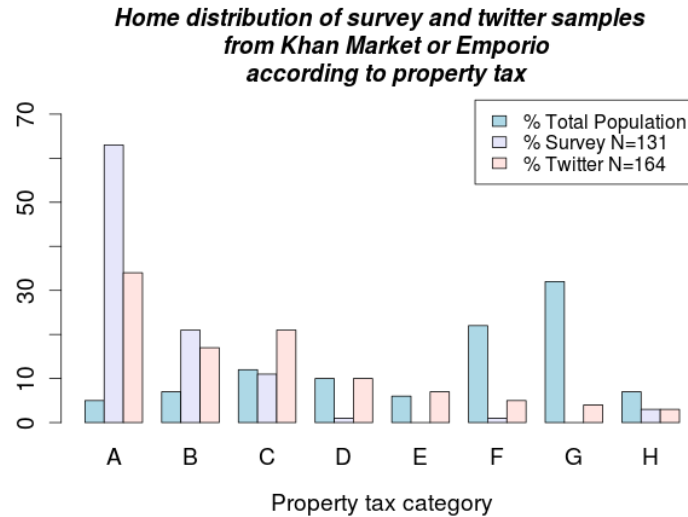


Figure 10: Percentage of Twitter and surveyed luxury shoppers in each property tax category.

The discrepancies between the survey and Twitter datasets indicate that people who tweet in luxury shopping places are not as wealthy as most of the typical luxury shoppers. In the last part of this article, we will try to understand why members of the lower middle class tend to tweet and geotag their tweets more than other groups in luxury shopping places, thus highlighting the limitations of using Twitter data to identify a social group based on its presence in a specific place.

3. TWITTER IN CONTEXT: WHO REALLY TWEETS IN LUXURY SHOPPING PLACES?

To our knowledge, there has been no sociological research on Twitter users in India, despite one study on the use of social media in South India (Miller et al., 2016) making it difficult to integrate potential biases in the dataset based on age, gender or social categories of the users. We do not have much information on why Indian Twitter users generally post messages. However, studies in other countries have shown that Twitter is mainly used by the younger generations, both as a platform to express opinions and as a social network. The first use likely reduces the number of geotagged tweets people send, since it mostly consists of sharing personal thoughts, general information, or media content. In the other case, when users geotag the tweets they send, we can assume that they are using it as a social network since such information could be only of interest to interpersonal relation circles (Figure 11).

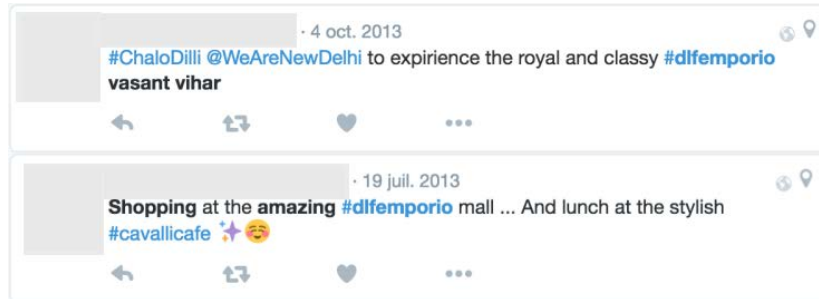


Figure 11: Two geotagged tweets sent from the DLF Emporio.

Twitter users often share a photo through Facebook or Instagram. In that case, tweeting becomes a way to put forward a meliorative personal information with their followers (Figure 12).

Drawing on the concept of conspicuous consumption⁸, we can further assume that people tweet in luxury places in order to gain prestige, esteem and sense of importance. Indeed, luxury consumption in India, like in many places globally, particularly in developing countries, is a marker of success. Moreover, shopping malls and international brands are major symbols of modernity, giving consumers a sense of “world-class” belonging and a taste of what is perceived as “first-class” or “five-star” lifestyles (Brosius, 2013). When surveyed, several consumers of the DLF Emporio complained about the noisiness and bad crowd of traditional Indian shopping places and reported that they preferred comfortable environments like shopping malls rather than open markets, which reminded them of the Indian “bazaars”, which convey archaic imaginaries. Among the luxury shoppers, we would thus find a considerable number of individuals with limited purchasing power, coming to experience new luxurious and so-called modern environments. In India, consumption has indeed become a marker of middle-class belongingness in the last two decades (Varma, 2014; Rault, 2017). This is notably illustrated by the slogan of the DLF Emporio, “experience luxury”.

⁸ The concept was coined by the sociologist and economist Thorstein Veblen in his seminal work « Theory of the Leisure Class », in 1899. He defines « conspicuous consumption » as the act of spending on unnecessary items in a way that makes people notice it in order to display or attempt to increase one’s social status.

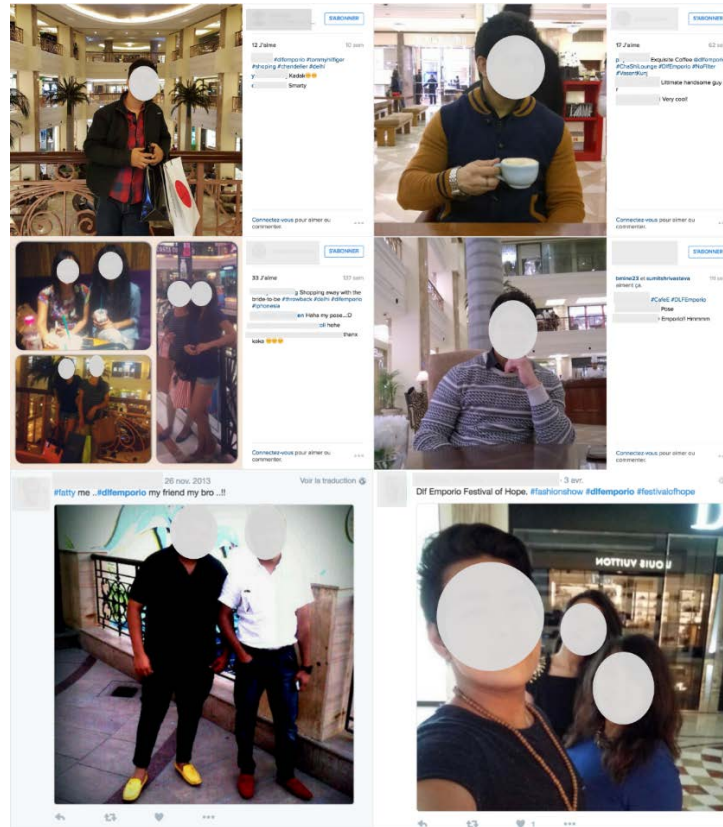


Figure 12: 4 photos posted on Twitter via Instagram (above) and 2 photos posted directly on Twitter (below) from the DLF Emporio.

Consequently, the users are not necessarily consumers. For instance, in a study on shopping malls in India, sociologist Nita Mathur, showed that among the people who visited the mall regularly, only 20%, the “serious buyers”, had disposable income at hand, while the remaining 80% consisting of “casual buyers”, mainly youngsters, came to familiarize themselves with the latest fashion trends or spend time with friends (Mathur, 2010). Accounts of shop sellers in the luxury shopping places, reporting the lack of customers in their stores, confirmed that many of the visitors did not make any purchases. We could also observe that many users did not feel comfortable enough to enter the shops and would remain in the alleys, observing products through the window, limiting themselves to buying a beverage in the cafés. The low consumption levels actually have consequences on the profitability of international luxury brands. Several of them had to leave India prematurely or are still struggling to become profitable and convince price conscious Indian customers⁹. For instance, Bottega Veneta, an Italian

⁹ “Luxury retailers shrink outlets to maximise profits”, 09.04.2012, *The Economic Times*, http://articles.economictimes.indiatimes.com/2012-04-09/news/31313147_1_luxury-brands-abhay-gupta-retail-space.

house specialized in leather goods, had to downsize its store area by half and move it in an older section of the DLF Emporio mall. With the resistance of small retailers, the difficulties to adjust to the diversity of the Indian consumer, the lack of commercial infrastructures and an adverse legislation, retailing multinationals are facing considerable hardships to penetrate the Indian market (Rault 2015). Actually, with only 2% share of the global luxury market, India trails well behind other emerging nations, China in particular (30%). From the Indian ultra-high-net-worth individuals (127 000)¹⁰ to those who own both a TV, a computer, a vehicle (two or four wheeler) and a phone (only 4,6% of total households), there seem to be relatively fewer typical luxury consumers in India¹¹. Actually, the old elite has mostly continued to buy luxury goods abroad due to limited availability of products and late arrival of new collections.

As a result, luxury shopping places are home to different social strata, also because, as Veblen wrote, middle class members imitate the consumption practices of the upper class in order to improve their social status (Veblen, 1994). On the contrary, regular customers and serious buyers do not feel the need to display activities of luxury shopping because it is a habit. They do not build their social status by showing their capacity to consume but by their ways of consuming, as it was highlighted in the monumental work by Pierre Bourdieu, *La Distinction* (Bourdieu, 1979). In doing so, they distinguish themselves from the lower groups. Several users who considered themselves as regular and serious customers thus look down upon people who wear loud colors and logos, show off brands, take selfies¹² and use social networks to display their habits. This is the case of Arjun, a 30 years old architect and a regular customer at DLF Emporio.

“If I open my Facebook right now there will plenty of friends posting selfies with friends in Emporio. Digital world has made people stupid, imbecile. It is about keeping displaying their habits and everything. It is about being loud in the open. If you see my activity on Facebook, I don't post anything on my page. It is a social service supposed to connect people, not to display your entire habits.”

According to their experience, social background, caste, class, origin or income, individuals do not have the same ways of tweeting, and in India, society is highly fragmented on those many lines (Deshpande, 2004). In the case of the users of luxury shopping places, people who are not familiar with luxury environments might be more prone to tweet there than regular customers. Consequently, the assumption that luxury shoppers would tweet equally seems to be wrong, partly explaining why middle-income groups are over-represented in the Twitter sample. We can thus conclude that the *Twitter* users in luxury shopping places are not representative of the Indian middle upper class.

¹⁰ Top of the Pyramid 2015. Decoding the Ultra HNI., Kotak, 2015.

¹¹ Government of India, “Census of India”, 2011.

¹² Self-portrait photos.

CONCLUSION

To conclude, we can affirm that the use of social media geotagged data provides us with new perspectives on social groups' activity space and mobility. However, if selected on the basis of the sending of a message from a place, the sample might not be representative of the social group the study initially intended to follow, due to the various social incentives to post on social media and the varying social backgrounds of the individuals.

Consequently, there is a need for a mix use of Twitter data and ethnographic methods. First of all, Twitter data can be a major help to the ethnographer, since it facilitates the inclusion of a large variety of locations in a study. This way, there are more possibilities for comparative approaches at various scales. When merged with other spatial information data (such as OpenStreetMap or Google place), such data can also help in a great way in analyzing social groups' activity space and mobility, as well as lifestyles. For instance, luxury shopping places could seem to be an accessible and easy-going place of study, but it is not. Customers seldom want to spend time answering questions during their leisure moments; their answers are often inaccurate and partial; it consequently takes a lot of time and resources to collect a significant sample of customers; and finally, official authorizations for studies in private places are often difficult to obtain. With Twitter data, the researcher does not need the agreement of the users to inquire about their practices. Even though the Twitter users supposedly know that the information they give out is public since they signed the user agreement, they do not know that this data could serve in tracking their movements, even for scientific purposes. This is an ethical question which remains unsolved in this paper, but would deserve a larger discussion.

Eventually, and we would like to stress it, there is a need to do ethnographic research in order to accurately use social media data. Integrating the social representations of the users can help in identifying potential limitations in geotagged data, and perhaps, being aware, integrate the potential biases in the analysis.

REFERENCES

- AHAS R., AASA A., SIILM S., TIRU M. (2010), "Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data", *Transportation Research Part C: Emerging Technologies*, vol. 18, n° 1, pp. 45-54.
- BOURDIEU P. (1979), *La Distinction. Critique sociale du jugement*, Paris : Les éditions de minuit.
- BROSIUS C. (2013), *India's middle class: new forms of urban leisure, consumption and prosperity*, New Delhi: Routledge India.

- CALABRESE F., BERLINGERIO M., LORENZO G.D., NAIR R., PINELLI F., SBODIO M.L. (2013), "AllAboard: A System for Exploring Urban Mobility and Optimizing Public Transport Using Cellphone Data", in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Berlin Heidelberg: Springer, pp. 663-666.
- CEBEILLAC A., DAUDÉ E., HURAUX T. (2016), "Where? When ? And how often ? What can we learn about daily urban mobilities from Twitter data and Google Map in Bangkok (Thailand) and what are the perspectives for Dengue studies ?", *Actes du 15e colloque du GT Mobilités Spatiales, Fluidité Sociale (MSFS)*.
- DESHPANDE S. (2004), *Contemporary India: a sociological view*, New Delhi: Viking.
- ESTER M., KRIEGEL H.-P., SANDER J., XU X. (1996), "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", *KDD*, vol. 96, n° 34, pp. 226-231.
- FERRARA E., VARUL O., DAVIS C., MENCZER F., FLAMMINI A., (2015), "The rise of social bots", *Communications of the ACM*, vol. 59, n° 7, pp. 96–104. Doi: 10.1145/2818717
- HAWELKA B., SITKO I., BEINAT E., SOBOLEVSKY S., KAZAKOPOULOS P., RATTI K. (2014), "Geo-located Twitter as the proxy for global mobility patterns", *Cartography and Geographic Information Science*, vol. 41, n° 3, pp. 260-271.
- JURDAK R., ZHAO K., LIU J., ABOUJAOUDE M., CAMERON M., NEWTH D. (2015), "Understanding human mobility from Twitter", *PloS one*, vol. 10, n° 7, e0131469
- KAHLE D., WICKHAM H. (2013), "ggmap: Spatial Visualization with ggplot2", *The R Journal*, vol. 5, n° 1, pp. 144-161. <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>
- LIU J., ZHAO K., KHAN S., CAMERON M., JURDAK R. (2015), "Multi-scale population and mobility estimation with geo-tagged tweets", *Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference*, pp. 83-86.
- LUO F., CAO G., MULLIGAN K., LI X. (2016), "Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago", *Applied Geography*, vol. 70, pp. 11–25.
- MATHUR N. (2010), "Shopping Malls, Credit Cards and Global Brands: Consumer Culture and Lifestyle of India's New Middle Class", *South Asia Research*, vol. 30, n° 3, pp. 211–231.
- MILLER D. (2016), *How the world changed social media*, London: UCL Press.
- MORSTATTER F., PFEFFER J., LIU H., CARLEY K. (2013), "Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose", *Association for the Advancement of Artificial Intelligence*.
- PERKINS T. A., GARCIA A. J., PAZ-SOLDAN V. A., STODDARD S. T., REINER Jr. R.C., VAZQUEZ-PROKOPEC G., BISANZIO D., MORRISON A. C., HALSEY E. S., KOCHER T. J., SMITH D. L., KITRON U., W. SCOTT T., TATEM A. J. (2014), "Theory and data for simulating fine-scale human movement in an urban environment", *Journal of The Royal Society Interface*, vol. 11, n° 99.

- PREOȚIUC-PIETRO D., LAMPOS V., ALETRAS N. (2015), “An analysis of the user occupational class through Twitter content”, *The Association for Computational Linguistics*.
- RAULT Y.-M. (2015), « Le lent développement de la grande distribution en Inde », in CADENE P. et DUMORTIER B. (Eds.), in *L'Inde : une géographie*, Paris : CNED / A. Colin, pp. 368- 375.
- RAULT Y.-M. (2017), « Classes moyennes et construction d’une modernité indienne. Le cas des « shopping malls » à Delhi », *Bulletin de l’Association de géographes français*, vol. 1, pp. 77-94.
- SARMA A. K. (2006), *The social logic of shopping: a syntactic approach to the analysis of spatial and positional trends of community centre markets in New Delhi*, University College London.
- SRIDHARAN E. (2011), “The Growth and Sectoral Composition of India’s Middle Classes: Their impact on the politics of economic liberalization”, BAVISKAR A., RAY R. (Eds.), in *Elite and Everyman: The Cultural Politics of the Indian Middle Classes*, London: Routledge, pp. 27–58.
- STEIGER E., DE ALBUQUERQUE J.P., ZIPF A. (2015), “An Advanced Systematic Literature Review on Spatiotemporal Analyses of Twitter Data”, *Transactions in GIS*, vol. 19, n° 06, pp. 809–834.
- TELLE O. (2011), *Aedes : Analyse de l’émergence de la dengue et simulation spatiale*, Université de Rouen.
- VARMA P. K. (2014), *The new Indian middle class: the challenge of 2014 and beyond*, Noida: Harper Collins Publishers India.
- VEBLEN T. (1899), *The theory of the leisure class*, Dover thrift editions (1994), New York Dover Publications.
- WICKHAM H. (2009), *ggplot2: Elegant Graphics for Data Analysis*, New York: Springer-Verlag.
- WANG Y., CALLAN J., ZHENG B. (2015), “Should We Use the Sample? Analyzing Datasets Sampled from Twitter’s Stream API”, *ACM Transactions on the Web*, vol. 9, n° 3, pp. 1–23.
- ZANDBERGEN P. A., BARBEAU S. J. (2011), “Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones”, *Journal of Navigation*, vol. 64, n° 3, pp. 381–399.

ACKNOWLEDGMENT

The authors would like to thank the organizers of the 18th AJEI (Association des Jeunes Etudes Indiennes) workshop for providing a stimulating forum to present and discuss their results.