Where? When? And how often? What can we learn about daily urban mobilities from Twitter data and Google POIs in Bangkok (Thailand) and which perspectives for dengue studies?

Alexandre Cebeillac, Éric Daudé, Thomas Huraux

Netcom
Réseaux, communication et territoires

31-3/4 | 2017
Mobilités et (r)évolutions numériques

Where ? When ? And how often ? What can we learn about daily urban mobilities from Twitter data and Google POIs in Bangkok (Thailand) and which perspectives for dengue studies ?

Où, quand et à quelle fréquence ? Que nous apprennent les données Twitter et Google (POI) sur les mobilités quotidiennes à Bangkok (Thaïlande), et quelles perspectives pour les études sur la dengue ?

Alexandre Cebeillac, Éric Daudé and Thomas Huraux

Electronic version
URL: https://journals.openedition.org/netcom/2725
DOI: 10.4000/netcom.2725
ISSN: 2431-210X

Publisher
Netcom Association

Printed version
Date of publication: 16 December 2017
Number of pages: 283-308
ISSN: 0987-6014

Brought to you by Université de Caen Normandie

Electronic reference
Alexandre Cebeillac, Éric Daudé and Thomas Huraux, "Where ? When ? And how often ? What can we learn about daily urban mobilities from Twitter data and Google POIs in Bangkok (Thailand) and which perspectives for dengue studies ?", *Netcom* [Online], 31-3/4 | 2017, Online since 26 March 2018, connection on 01 June 2022. URL: http://journals.openedition.org/netcom/2725 ; DOI: https://doi.org/10.4000/netcom.2725

Networks and Communication Studies,
NETCOM, vol. 31 (2017), n° 3-4
pp. 283-308

WHERE? WHEN? AND HOW OFTEN? WHAT CAN WE LEARN ABOUT DAILY URBAN MOBILITIES FROM TWITTER DATA AND GOOGLE POIs IN BANGKOK (THAILAND) AND WHICH PERSPECTIVES FOR DENGUE STUDIES?

OU, QUAND ET A QUELLE FREQUENCE ? QUE NOUS APPRENNENT LES DONNEES TWITTER ET GOOGLE (POI) SUR LES MOBILITES QUOTIDIENNES A BANGKOK (THAÏLANDE), ET QUELLES PERSPECTIVES POUR LES ETUDES SUR LA DENGUE ?

CEBEILLAC ALEXANDRE[1], DAUDE ÉRIC[2], HURAUX THOMAS[3]

Abstract *- Human mobilities in urban areas have an impact on the spread of infectious diseases, including those caused by mosquito-borne diseases like dengue and Zika virus. Therefore, finding appropriate data and methods to perform spatial analyses of mobilities is a critical issue. The emergence of substantial amount of geolocated data, which is easily available on the Internet, has a great potential in mobility researches, especially when paired with land use, following the activity-space concept. This paper, focused on the dengue endemic mega-city of Bangkok (Thailand), explores the potentialities of (A) using a land use classification from Google's points of interest (POIs) to assess the likelihood of performing an activity from (B) a large dataset of individual geolocated tweets to characterize and quantify daily mobility. These emerging data sources allow the characterization of (C) the rhythms of daily mobility in Bangkok, from the perspective of (1) the macroscopic urban pulse and (2) the rhythms and aims for individual movement. The advantages and limitations of this kind of data will be finally discussed regarding dengue epidemics.*

Keywords*: Twitter ; Google POI ; Urban mobility ; Bangkok.*

[1] UMR 6266-IDEES, CNRS, Université de Rouen, alexandre.cebeillac@etu.univ-rouen.fr
[2] UMR 6266-IDEES, CNRS, Université de Rouen, eric.daude@cnrs.fr
[3] UMR 6266-IDEES, CNRS, Université de Rouen, thomas.huraux@univ-rouen.fr

Résumé – *Les mobilités notamment quotidiennes sont un facteur qui contribue à la propagation de maladies infectieuses en milieu urbain. C'est le cas pour les arboviroses transmises par des moustiques du genre Aedes comme la dengue ou le Zika. L'utilisation de données sociales de plus en plus précises et de méthodes d'analyses spatiales contribue à enrichir notre compréhension de ces mécanismes sous-jacents. L'émergence d'une grande quantité de données géolocalisées aisément accessible sur Internet offre de nombreuses pistes de recherches sur les mobilités, notamment si ces dernières sont couplées à des données d'utilisation du sol, ce qui permet de construire des espaces d'activités. Cet article prend pour fil conducteur la dengue, endémique à Bangkok (Thaïlande), pour explorer les potentiels (A) d'une classification de l'utilisation du sol construite à partir des points d'intérêts (POI, Google Map) afin d'estimer la probabilité pour un individu d'effectuer une activité en mobilisant (B) une grande base de données de Tweets géolocalisés qui permet de quantifier les mobilités quotidiennes individuelles. Ces informations ainsi construites permettent (C) une caractérisation des mobilités quotidiennes à Bangkok, d'un point de macroscopique (pulsation urbaine) et des rythmes et objectifs des déplacements individuels. Les avantages et limites de ce type de donnée sont discutés dans le contexte des épidémies de dengue.*

Mots-clés – *Twitter ; Google POI ; Mobilité quotidienne ; Bangkok.*

INTRODUCTION

Intra-urban mobilities are analyzed and modeled in several domains: public transport planning, pollutant emission evaluation, epidemiological monitoring, etc. Human mobilities contribute to the spread of viruses at various scales: international (Tatem et al., 2006; Wesolowski et al., 2012) national (Sorichetta et al., 2016; Wesolowski et al., 2015) and local (Perkins et al., 2014; Stoddard et al., 2013). However, epidemiological studies throw little light on the significance of this contribution at infra-local levels, where physical interactions between vectors (*i.e. Aedes Aegypti m*osquitoes) and human beings take place. In fact, it is almost impossible to identify places where human are infected by mosquitoes. Yet, in the case of vectors with essentially daytime predatory activities, there is a significant probability of human beings getting the infection outside the place of residence. According to a large majority of epidemiological surveys, sick peoples are listed at their places of residence, so it is probable that several host-vector interactions escape the traditional systems of disease monitoring and control. But dengue is a complex pathogenic system as the onset of the disease is triggered by a complex association of environmental, biological and social factors. This led us to develop a multi agent-based model for dengue in order to explore two important issues: the role of local socio-environmental contexts in the vector population dynamics and the impact of human mobilities in the spread of viruses transmitted by *Ae. Aegypti*. The study of human mobilities in Bangkok should make it possible to explore viral propagation scenarii depending on diffusion mechanisms through proximity and for long distances (Daudé et al., 2015) and help to calibrate an agent-based model of urban mobilities. The mobility studied in this article relies on the concept of activity space, defined as "the portion of urban space that an individual visits during his daily activities" (Horton and Reynolds, 1971). It includes all the places visited

by an individual, places that are associated with activities defined in time as well as in space (Perchoux et al., 2013). The richness of the concept of activity space relies on two dimensions: (a) ego-centric mobilities, that is, different types of places visited according to a person's schedule and (b) place-centric or loco-centric aggregated visits, that is, the potential of a type of activity to attract individuals depending on the time. Hägerstrand (1970) identifies two major types of activities depending on their level of spatio-temporal flexibility and the freedom an individual has in choosing them:

- *Fixed activities*: these are carried out regularly and/or in a precise place. This can be, for example, work, school or a guitar class on Wednesday.
- *Flexible activities*: these are less restricted in time and space, less planned. We can consider, for example, the action "go out for a beer at an outdoor café" because the sun is shining, or "go shopping" at the last minute because there is an ingredient missing for dinner.

Typically, these activities are studied in household and activity-based travel surveys. Our objective is to reconstruct this type of data using alternative data sources that provide a direct measurement of mobility. For this, we use information obtained from Google and Twitter in order to construct a daily schedule for every agent giving time slot-wise probability for carrying out a given activity. Generating several thousands of individual schedules should make it possible to simulate daily rhythms of urban mobility in Bangkok.

Using data obtained from social networks has several advantages compared to the data obtained from mobile phones. The latter has already been used to observe, analyze and model people's mobility (*i.e.* Calabrese et al., 2013; Deville et al., 2014; González et al., 2008; Song et al., 2010 etc.). In addition, given a relatively low technical cost of acquisition, satisfying geographic precision and temporal continuity level in addition to a large sample size, depending on the operator, this data is in fact adequate for analyzing mobility at various scales. Yet, access to this type of data in partnership with telecom operators is not easy since it falls under a sector that is often highly competitive or linked to national security. Furthermore, implications on privacy protection and personal data raise ethical questions and are thus highly regulated in certain countries, notably France[4]. Finally, even if cooperation is possible in one place, it is not evident for it to be replicated in another place where the market could be under the control of other telecom operators.

One of our objectives is to develop a model of human mobility at the city-level, which would be as generic as possible and which could easily be applied to various geographic zones. This implies that we could use the same sources or information types in various places. In this context, we identified specific types of geographic data available on Internet. These rely on the combined growth of smartphones, social

[4] Data files processed in this project are subject to a declaration at the CNIL.

networks and online cartographic services which have led to a sharp increase in geographic information accessible online. Twitter's platform illustrates this phenomenon: users can send short messages publicly leaving digital traces that can be retrieved by others through APIs[5]. These messages are dated and are sometimes precisely geolocalized, revealing the GPS position of the user's mobile phone[6] - which is more precise than the telecom data, limited to the closest relay antenna. As Twitter geolocation feature was added in 2009, and rely on the growing availability of smartphones with a GPS chipset, we can consider this data as relatively new and make it a probable promising candidate for a study on daily mobility in a mega city. To create activity space profiles, we crossed these data sets with a land use typology of Bangkok generated using points of interest (POIs) listed in Google Places[7], notably, shops, educational institutions and business entities (for a detailed methodology, see Cebeillac et al., 2017). The general idea is that by attributing categories of functions (restaurants, service, etc.) to places visited by individuals, it is possible to reconstruct activity profiles over a typical week.

If the number of individuals in the sample is large enough, these profiles can be used to generate a synthetic group of individuals with mobility patterns similar to the target group. The latter corresponds to the group that is most exposed to dengue, that is, people under 30 years of age living in Bangkok (Limkittikul et al., 2014). Even if it's difficult to estimate the age of Twitter users, as it's only declarative, Sloan et al. (2015), assessed that nearly 80% of the users in UK were below 30 years old and Longley et al. (2015) found that men below 30 and women below 25 years of age were over-represented in London. Although Twitter's demographic data should vary from place to place, we assume that the order of magnitude is similar in Bangkok and that the use of Twitter is over-represented by the youth, which are also more exposed to dengue.

The technical side of this work is somehow linked with the study of Noulas et al. (2011), who collected Foursquare check-in data publicly shared on Twitter. As this data contains the name and the type of location (restaurant, train station, hotel etc.), they were able to analyze the temporal patterns of check-ins in different kinds of places and the overall movement from one place to another. Huang et al. (2014) inferred the potential activity linked with tweets by intersecting the location of the message by querying Google Place with a search radius of 20m. But their analysis focused only on one user with a large number of Tweets and cases with more than one POI in the search-radius are not evoked. Thus, our approach is more generic, as it doesn't focus on a sub-group of Twitter users (*i.e.,* the one that use foursquare or just one user), and supported by a thematic land use classification from an independent spatial database.

In the following sections we present (A) a typology of land use obtained from the Google Maps POIs in the city of Bangkok (Thailand) and (B) a database obtained

[5] Application Program Interface
[6] https://support.twitter.com/articles/118492#
[7] https://developers.google.com/places/

from the social networking platform, Twitter. These data sources make it possible to characterize (C) the rhythms of daily mobilities in Bangkok, from the perspective of (1) macroscopic urban pulse and (2) rhythms and potential targets of individual movement. This data will be useful in calibrating a simulation model for studying the role of daily mobility in the spread of dengue, and their limitations will be discussed.

A. BANGKOK, A MEGA BUT MONOCENTRIC CITY

Bangkok, Thailand's capital, is a mega city with a population of 9.2 million in 2015[8], which makes it the 36th among the most populated cities in the world, and the 3rd among the ASEAN countries, behind Manila and Jakarta. It is a sprawled out city with an area of 1 568 km², divided into 50 districts and 169 sub-districts (called Kwaengs).

Figure 1 highlights a heterogeneous distribution of the population with a centre/periphery type of population density structure and an over-densification along the communication lines. These densities suggest differentiated volumes and mobility profiles depending on the localization of individuals in the city and the types of activities found there. The latter were defined through an analysis of land use and Google's POIs (Point Of Interest). Information on the activities available in the territory of Bangkok will allow us to enrich the contexts of the various mobility profiles established: Which types of places are mostly visited? At what times? Land use in Bangkok was identified using ~75000 POIs obtained from Google Place and aggregated to 180m cells to which we applied an Ascending Hierarchical Classification (ACH) (Cebeillac et al., 2017). This method, inspired by Fleury et al. (2012) allows us to get a typology of land use and commercial units, divided into 6 major categories: accommodation and entertainment areas; major shopping and entertainment areas; shopping and entertainment areas; areas with basics shops; overestimation of the secondary sector activities (unidentified companies, car-related activities or electronic companies, etc.); undifferentiated areas with little economic activity.

[8] https://esa.un.org/unpd/wpp/

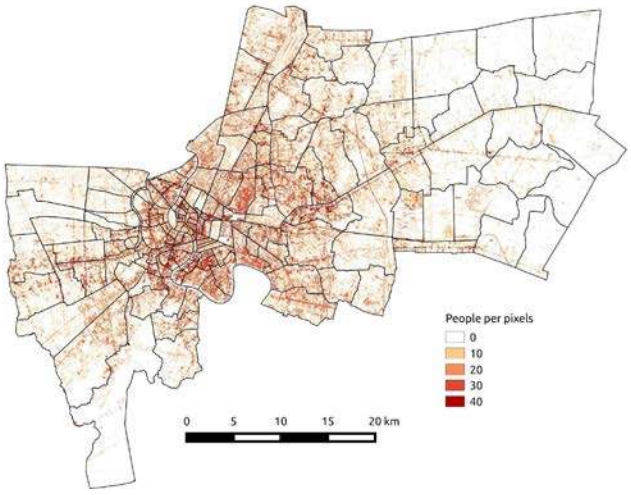

***Figure 1**: Estimation of spatial distribution of Bangkok's population using 30m pixels.*
Estimation based on dasymetric mapping using images from Landsat 8 and population census in 2010 (Misslin and Daudé, 2016).

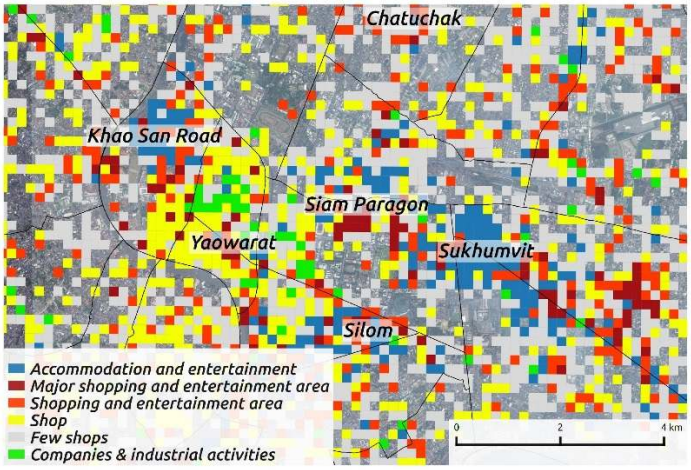

***Figure 2**: Typology of economic activities in central Bangkok, using Google POIs (Cebeillac et al., 2017).*

Figure 2 shows the presence of highly differentiated activity poles according to their main functions and spatial extent. Accommodation and entertainment areas (in blue) are thus found together mainly in three zones: backpacker tourism neighborhoods (Khao San Road), Sukhumvit neighborhood, which extends along the road towards Pataya and is very well-known for its nightlife and Silom neighborhood to the South of the Lumphini park. The major shopping areas (in red) are more scattered in the city with, in particular, a very high concentration in the Siam neighborhood, which has

gigantic shopping malls, interconnected through walkways (Siam Paragon, MBK Center, Siam Center, Central World). Apart from these highly specific neighborhoods, the city is mainly structured with small local shops and a few businesses or workshops as well as relatively occasional dense commercial areas (probably Malls) established at the crossroads of secondary communication lines.

This typology of commercial areas and units in Bangkok is further complemented by another layer containing educational and religious places (schools, libraries, universities, temples, churches, etc.) obtained by merging OpenStreetMap and Google Place data into a Boolean grid, and a transport layer (main roads, metros, buses, station and rails) obtained from OpenStreetMap.

These socio-economic functions of Bangkok, paired with global and individual mobility patterns from Twitter will be analyzed in the following sections.

B. FROM GEOLOCATED TWEETS TO INDIVIDUAL ACTIVITY SPACES

With more than 313 million active users in 2016, Twitter is one of the most popular platforms in the world for self-expression[9]. It is a "micro-blogging" service launched in 2006, which makes it possible to send short messages containing 140 characters (tweets). Most of these messages are public, according to the user's choice, and hence accessible to everybody, and some of them are even geolocated (a feature added in 2009). This geolocation is called "general" when the user adds a location to his message. To do so, he chooses a place from a list of places, which is generally but not necessarily close to his real location. This geolocation is called "precise" when the user authorizes the Twitter application to access geolocation data of his phone and subsequently decides to share it in his message. The precision of the localization then relies on the accuracy of the mobile phone's GPS, which may vary from a few meters to hundreds of meters, when the user is outdoors or indoors (Zandbergen and Barbeau, 2011).

Twitter gives free access to a part of its database and provides tools that allow collection of data, notably the API Stream[10] that provides, in real time, a maximum of 1% of the total public message flow. If a request returns fewer messages than the 1% threshold, which might be the case when we apply a geographical filter in Bangkok, all the public geolocated messages in the area should get recorded (Morstatter et al., 2013). To test these theoretical assumptions that rely on an opaque algorithm that could influence the data collection (Quesnot, 2016), we conducted an empirical test. This consisted in sending 2000 tweets with precise location in Bangkok through a bot and all

[9] 500 million tweets were sent daily in November 2013 while the number of active users was 215 million. Since then, the company has stopped releasing this type of information.
[10] https://dev.twitter.com/streaming/overview

of them were collected back by our program. This suggests that we might record most of the geolocated tweets sent in the area of Bangkok, as the live flow of messages is below the global 1% threshold.

Geolocalized tweets have great potential as we can be reasonably confident regarding their spatio-temporal accuracy, but there is no direct way, as far as we know, to check if the digitals footprint of a user is representative of his spatio-temporal trajectory. For this pilot study, we assume that aggregating all the tweets of an user on a typical week gives a satisfying sample of its spatio-temporal activities. We also assume an equal probability, in space and time, to have twitter activities for an individual.

A large number of studies conducted spatio-temporal analysis with Twitter data (see Steiger et al., 2015 for a review), and some studies have looked into geolocalized tweets to explore human mobility at the global scale (Hawelka et al., 2014) and study international migrations (Zagheni et al., 2014). Other studies at the national scale highlighted mobility trends in Australia (Jurdak et al., 2015) and United Kingdom (McNeill et al., 2016) or allowed identification of the types of places visited by analyzing message content in Japan (Fuchs et al., 2013) and United Kingdom (Steiger et al., 2015b). At the urban scale, some papers assessed the activity space of upper class people in Delhi, India (Cebeillac and Rault, 2017) and according to their "ethnicity" in Chicago (Luo et al., 2016) and London (Longley et al., 2015). Lenormand et al. (2014) showed that according to the aggregation level, tweets could be used as a proxy for mobility in Barcelona and Madrid. Khan et al., (2017) conducted a comparative study among the major cities in Australia, but didn't focus on the mobility pattern in one particular urban area.

1. Raw data from the Twitter social network and pre-processing

We recorded public and geo-localized data of Twitter users in Thailand in the extended area of Bangkok (13°5'-14°3'N, 99°8'-101°25'W) by limiting the extraction of data to the user's pseudonym (called unique identifier), time stamping and the precise coordinates generated at the time a Tweet is sent. Data was collected from 25th June 2014 to 4th December 2015. By the end of this period, we collected 20 126 426 tweets sent by 308 808 users within Bangkok and 25 752 072 tweets if we include messages sent by these users all over Thailand.

Some of the messages sent on Twitter do not come from real persons. The company announced in 2014 that close to 8.5% of its active users used other services. Therefore, a few of these users are potentially bots[11,12]. This percentage, which remained stable in December 2015, is explained by the fact that this social network is a good medium for promotion, publicity and geo-marketing. In addition, it is easy to hire a

11 http://qz.com/248063/twitter-admits-that-as-many-as-23-million-of-its-active-users-are-actually-bots/

12 https://www.sec.gov/Archives/edgar/data/1418091/000156459014003474/twtr-10q_20140630.htm

community manager or configure bots (twitbot), which can send promotional or informative messages in order to animate the social network of the account in question (Ferrara et al., 2015).

Regarding the data that we recorded, we cannot use message content or other metadata such as the medium of access to the platform (Android, iPhone, etc.) to remove these bots. Our filtering algorithm is thus based on a realistic average movement speed for each user between the dispatches of two tweets. Some authors define a threshold speed lower than that of a plane (Hawelka et al., 2014; Jurdak et al., 2015), but sometimes, two tweets may be recorded at the same time at very far places even though the author is clearly not a robot. This error could also be due to the Twitter servers or an undefined network connection problem. In order to avoid removing the users who are victims of this type of bug, we only removed those with more than 2% of the messages showing movement speeds higher than 600km/h. We also chose to remove users with an average speed of more than 70km/h, which seems unrealistic in the framework of intra-urban movement. Users with very high activity levels (more than 20 messages sent per day on an average) were also removed, as were occasional users (less than 15 days of activity recorded). Finally, we retained 21 006 900 Tweets for 99 307 users which meet the filtering criteria expressed herein above.

2. Individual activity spaces through the prism of their virtual social activities: data mining

As a first step, each user's tweets are put together according to the Dbscan algorithm (Ester et al., 1996) used largely in the context of creating clusters of geolocalized tweets (*e.g.*, Jurdak et al., 2015; Luo et al., 2016; Steiger et al., 2015a). The algorithm brings together the points (Tweets) into a single group based on the criteria of distance and minimum number of constituent elements (Figure 3). In our case, we chose a distance of 50m as the maximum distance between two points of a cluster (Figure 3.a) – even though a cluster may contain only one tweet. Subsequently we assigned the cluster's barycenter as a place in the individual's activity space (Figure 3.c). Each place created thus contains anything between a single dated and geolocalised trace (isolated tweet) and several hundreds of dated traces grouped together into a single point (the cluster's barycenter).

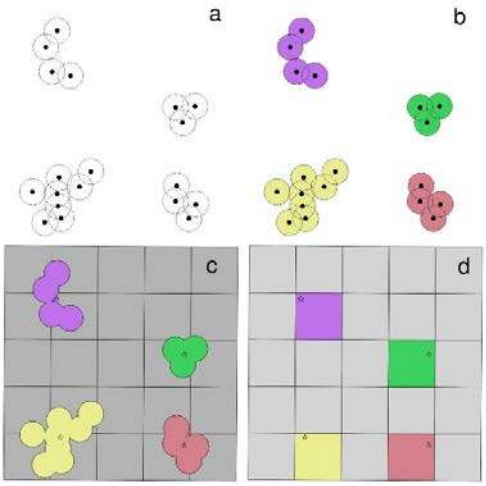

***Figure 3**: Principle of activity space creation*
A user tweeted 20 times (a). Use of the DBSCAN algorithm makes it possible to group together these tweets into 4 clusters (b) for which the barycenters are calculated (c). These barycenters are then assigned to a cell (d). The yellow cell becomes a place in the individual's activity space and contains 8 dated traces.

This approach allows us to reduce the lack of precision in the telephone GPS coordinates by retaining information on the number of tweets sent and the time when they were sent. We then assigned barycenters of each of the places created for all the users to a grid of cells measuring 180m a side (Figure 3.d), which corresponds to the spatial division of the typology established for activities in Bangkok (see figure 2). This last operation also makes it possible to add a spatial blur on the exact location of an individual in a cell and know immediately the number of people who visit a cell at a given instant.

In addition to this spatial aggregation, we constructed a temporal aggregation protocol. This became necessary because of the data's sporadic nature resulting from the differentiated behavior of Twitter users. Figure 4 illustrates this variability in the temporal frequency of geolocated message dispatch depending on the users, illustrated here for 9 users taken randomly.

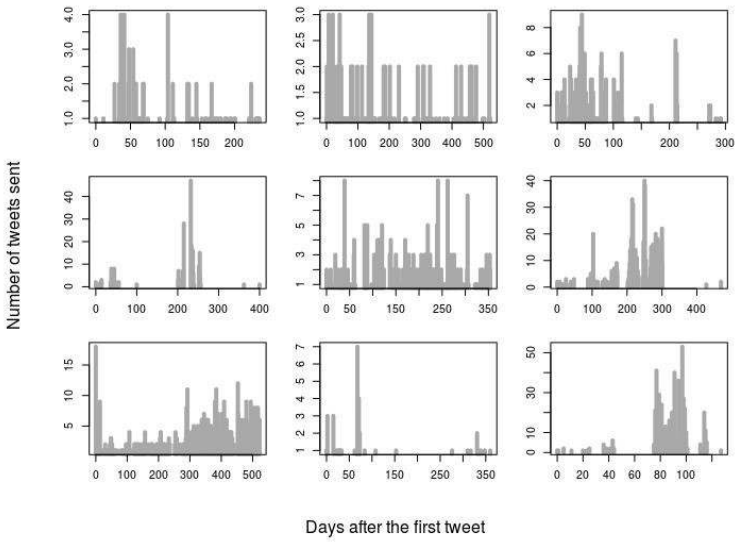

***Figure 4**: Temporal frequency of geolocalized message dispatch by 9 users taken randomly from the Twitter database.*

Some users use the social network regularly, tweeting a little every day. Others, on the contrary, have very high activity periods followed by a period of absence on the social network. In order to restrict the episodic nature inherent to this kind of dataset (Andrienko et al., 2012), we made the choice of aggregating the tweets per hour and per day of the week. This allowed us to generate a typical week but made us lose temporal sequences. Thus, we did not differentiate between a tweet sent on Thursday 12th January at 6.07 pm and another one sent on Thursday 19th January at 6.20pm. The two of them were considered as sent on a Thursday at 6pm. This loss of information is compensated by recording the frequency of visits to each place in the activity space. We thus have, for each individual, the number of a different day when this person tweeted from that place.

At the end of this first phase of raw data processing, we qualified, for each person, an activity space, defined geographically as a set of cells (places) in a grid, with each of these cells getting assigned timings and frequencies of visits. We will subsequently qualify the activity probably carried out in these cells according to the time and the main function (residential, trade, restaurant, work, school, transport) of the predefined cell.

3. Detection of the users' domicile

Estimating our sample individuals' place of residence is the starting point for the analysis of tweets in Bangkok since it allows us to verify the quality of the dataset by comparing it with the census data. Our first hypothesis is that, for a person, the most visited Tweet place between 8pm and 8am corresponds to his home (Luo et al., 2016). However, in the context of Bangkok, the effervescence of the nocturnal activity in

certain neighborhoods, which could skew this domicile attribution process, must be taken into account. We thus added a few conditions for the determination of the place of residence for Twitter users:

- The residence is one of the 3 places where the person records the most number of days with tweets sent notably between 8pm and 8am;
- The residence is located in a residential area; we thus remove the markets, railway stations, airports and parks;
- The residence is the place where we recorded the most number of days with sent messages.

After applying these filters, users that still have more than one potential place of residency are discarded. A total of 47 378 users fulfill these conditions including 35 607 users who are residing in the region of Bangkok, and 24 972 in the city of Bangkok. These users sent 16 000 384, 12 419 068 and 8 958 357 tweets respectively over the entire recording period.

In order to estimate the quality of our protocol and assess this sample's distribution with respect to the total population's distribution, we crossed these results with sub-district-wise census data for the entire population and age-group-wise data (Thailand Census 2010[13]) for the entire city of Bangkok (Figure 5). The Ripley's K was also computed using the border method (Ripley, 1988), and this second order statistic shows that the distribution of the estimated residence is clustered and not random (figure 5.c).

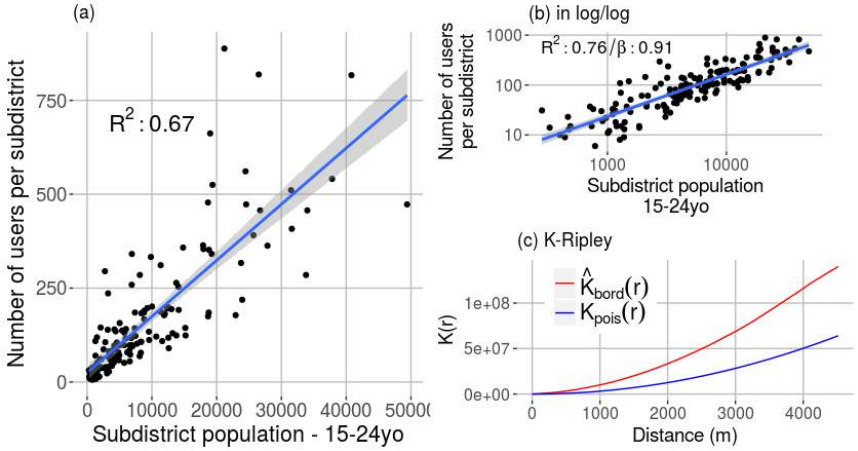

***Figure 5**: Correlation between the Twitter social network user sample in the place deemed as their residence and the 15-24 year old population at their place of residence (census 2010) following a linear (a) or a log/log regression (b). The computed K-Ripley (c) in red is far from a Poisson process, in blue, suggesting a clustered and non-random distribution of residences.*

[13] http://web.nso.go.th/

There is also a relatively high positive linear correlation between the total population at the place of residence and our sample ($R^2 = 0.655$). It increases slightly in the 15-24 age group ($R^2 = 0.67$). Interestingly, the correlation is higher if we consider both the population per sub district from the census and the estimation of the place of residence as a log/log function ($R^2 = 0.76$, figure 5.b). This suggests that the number of Twitter users in a sub-district depend on a power function of the local population, but it is difficult to explain why. These coefficients thus indicate that the sample provides a good spatial representativity during the data collection period, notably among the youth. The residue map based on the linear correlation shows an overestimation of the population in the city center (figure 6), notably in the tourism and entertainment neighborhood of Khao San Road. This can be explained with the algorithm for detecting residences based mainly on the messages sent in the evening to characterize the place of residence. This choice can be erroneous when several persons go regularly to the same night-time recreation neighborhood such as Khao San. In fact, this neighborhood accommodates a number of foreign tourists (Le Bigot, 2016) whose use of social networks in a holiday context is quite specific. University neighborhoods, notably that of Chulalongkorn, are also over-represented, probably because there are several students in university dormitories who are more active than the average on social networks and are not necessarily counted as residents of the neighborhood. Finally the population in the Eastern periphery of the city with few inhabitants is overestimated. On the other hand, the southwestern inhabitants of the megacity use less Twitter than the rest of the city. These sub-districts are highly populated, more rural and less privileged, which could explain this under-representation.

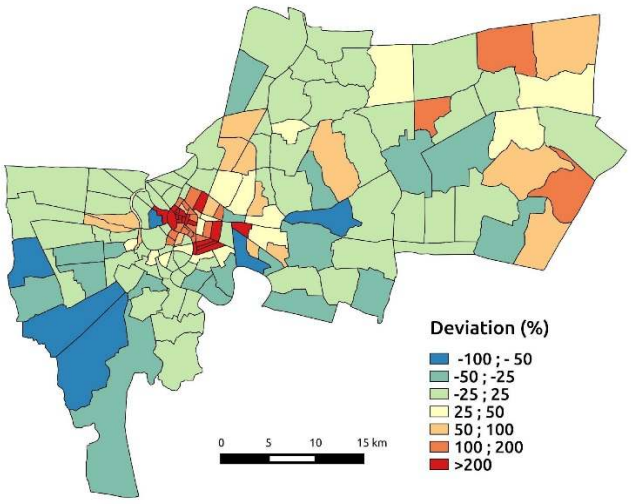

***Figure 6**: Cartography of the deviation between the estimated population according to the Twitter data and the population census in 2010.*

Nevertheless, the fact that geolocated tweets represent only a small part of the overall flow of messages - between 1.5 to 3% (Murdock, 2011) – does not seem to be a

limitation for our study, as we managed to collect a sample of ~25 000 persons residing in Bangkok, which seems to be, for the most part, spatially representative of the total population in the place of residence - even if it is difficult to assess its socio-economic representativity. Furthermore, our dataset is probably made up of essentially adolescents and young adults who are the main Twitter users (Luo et al., 2016; Sloan et al., 2015). This hypothesis is substantiated by the fact that several Tweets are sent from university neighborhoods. However, this population bias can be an asset to the extent that seroprevalence is lower in this population group, and then represent an at-risk group for dengue. In addition, as shown by McNeill et al. (2016) despite the demographic bias originating from the Twitter data, the latter is a good proxy to estimate short distance commuting – at least in the UK.

C. DAILY MOBILITY RHYTHMS IN BANGKOK

The objective of this section is now to characterize more precisely the rhythms, the directions and the distances covered by these movements at the macroscopic scale of the city (urban pulse) and the microscopic scale of the individuals (commuting schedules).

1. From the urban pulse …

Daily urban mobility is generally driven by at least two major almost symmetrical waves of massive population movement, one in the morning and the other in the evening. A third mid-day wave is also generally observed depending on the socio-cultural context. The city of Bangkok illustrates these overall dynamics relatively well. The maps in Figure 7 show sequentially these two main waves of movement on Tuesday and Sunday. A very scattered distribution of our sample is observed in the morning, then concentrated in the city center and along the transport hubs early in the afternoon and once again more dispersed in the evening. This general periphery-center-periphery type of mobility behavior varies depending on the day of the week with the population distribution generally more concentrated on Sunday than on Tuesday in Bangkok. This differentiation can essentially be explained by the more concentrated recreational activities in the weekend in comparison to the more diffuse professional activities on weekdays.

This phenomenon of spatial concentration is also observed when we consider the attractiveness of the places. Figure 8 shows the number of persons visiting the various cells[14] in Bangkok and a high number of areas receive only a few visitors while a small number of areas attract a large number of visitors. These two extreme positions correspond respectively to the peripheral areas and the city center of Bangkok.

[14] 180mx180m.

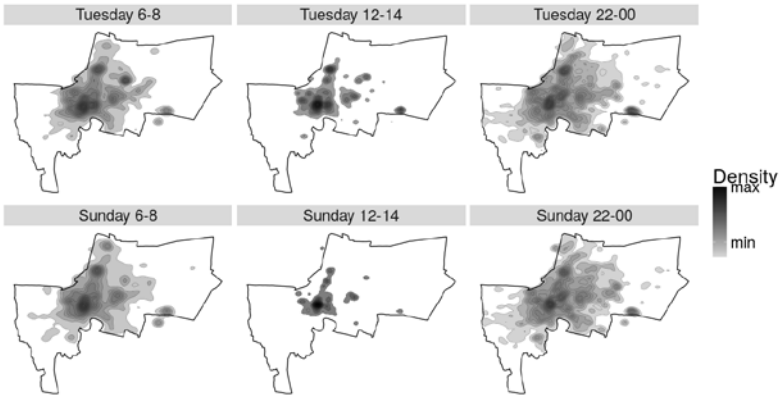

***Figure 7**: Densities of Twitter users on Tuesday and Sunday between 6 and 8am, 12 and 2pm, 10pm and midnight.*

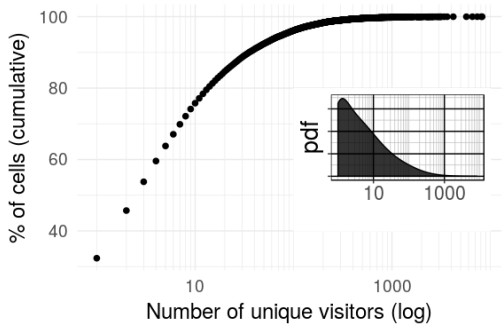

***Figure 8**: Distribution curve of the number of visitors in Bangkok agglomeration with (y) the number of persons present in an 180mx180m cell and (x) the proportion of cells in the entire agglomeration. Close to 75% of the areas receive less than 10 persons and less than 5% receive several hundreds of persons.*

Evaluating sub-district-wise spatial range of movements by individuals could allow us to identify an effect of distance from the centre in this centre-periphery type of pendular movement dynamics: the sub-districts far from the city centre probably lead to movements that are more distant than the ones originating from the centre. In order to explore this hypothesis, we calculate the gyration radius (equation 1). Used originally to describe the behavior of an object around an axis, the gyration radius was applied to the analysis of the scope and trajectories of mobilities (González et al., 2008). This is an indicator for measuring the dispersion of all the places a_i in an individual's activity space of n places, from a reference point, a_{cm}, which is the residence in this case[15].

[15] But maybe the place where the individual tweeted most often or the centre of gravity of his activity space.

$$r_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|a_i - a_{cm}|^2}$$

***Equation 1**: radius of gyration.*

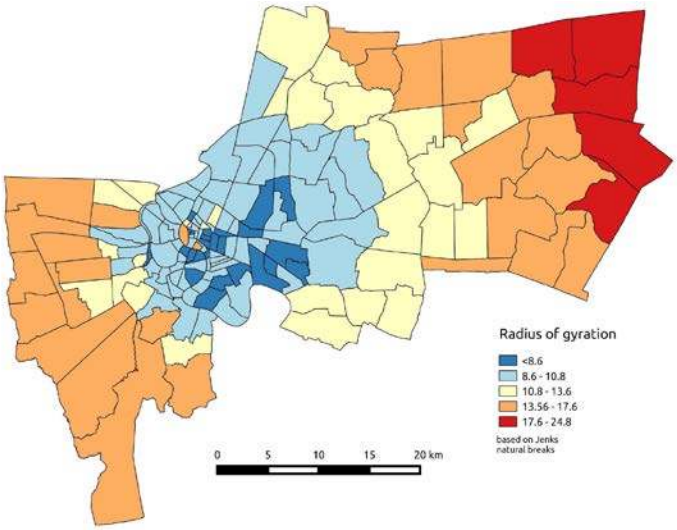

***Figure 9**: Cartography of the sample's average radius of gyration (in km) using their estimated place of residence, aggregated at the sub-district level.*

Figure 9 shows the average sub-district-wise radius of gyration for the movement occurring in Bangkok and its inner suburbs. The range of the movement confirms the centre-periphery effect induced by the mono-centric type of the Thai capital's urban structure. The residents of the city centre have, on an average, a less dispersed activity space (less than 10 km from the place of residence) than the inhabitants at the periphery (up to 25 km).

The main urban pulse of the city of Bangkok thus reveals (1) periphery-center-periphery symmetrical population flows, in the beginning and at the end of the day, (2) a high concentration of population in a limited number of central places, (3) a tendency of individuals residing in the margins of the territory to cover longer distances.

2. … to the rhythms and purposes for individual movement

Highlighting these macro-dynamics of movement in the city doesn't tell us anything about the rhythms of these individual movements (how many movements per day?) or about the purposes of this movement. This issue can be resolved by creating sub-groups of people according to their individual mobility behavior, and pairing the visited place with the land-use shape to assess the probable activity.

The work of González et al., (2008) and Song et al., (2010), based on mobile phone call record (CDR) details suggested that 3 main parameters are important to describe human mobility:

- The radius of gyration
- The number of places visited as an indicator of inter-individual movement variability
- And the trip length distribution, *i.e.,* the distance probability between two distinct locations, approximated here as the mean distance between two time-consecutives tweets locations.

We have thus established a classification of the sample individuals according to these three movement criteria and then observed their sub-district-wise spatial distribution. We apply a classification according to the K-means algorithm into 5 categories, as suggested by the silhouette methods (Kaufman and Rousseuw, 1990). For each of the categories, we get the average statistics of the individuals (Figure 10, black density plot) compared to the global trend (blue density plot). The proportion of the individuals belonging to each category is then mapped (Figure 10). This highlights that user groups can have similar mobility pattern depending on the sub-districts. Class 1 shows individuals with a relatively high radius of gyration and mean trip distances, but with a number of visited places that follow the global trend. This suggests an ability to travel over long distances. They live mainly in the periphery of Bangkok, along with users from the Class 5, but the latter exhibit a lower mean trip distance. Class 3 describes 42% of our sample and stands for people with a low dispersion level that live mostly downtown, where there are numerous and diverse services and thus do not require or require little long distance movement. Users from Class 4 lives mainly in the pericenter and have a higher mean trip distance than the users of Category 3. Finally, people from Class 2, defined by a large number of visited places are not clustered in the city.

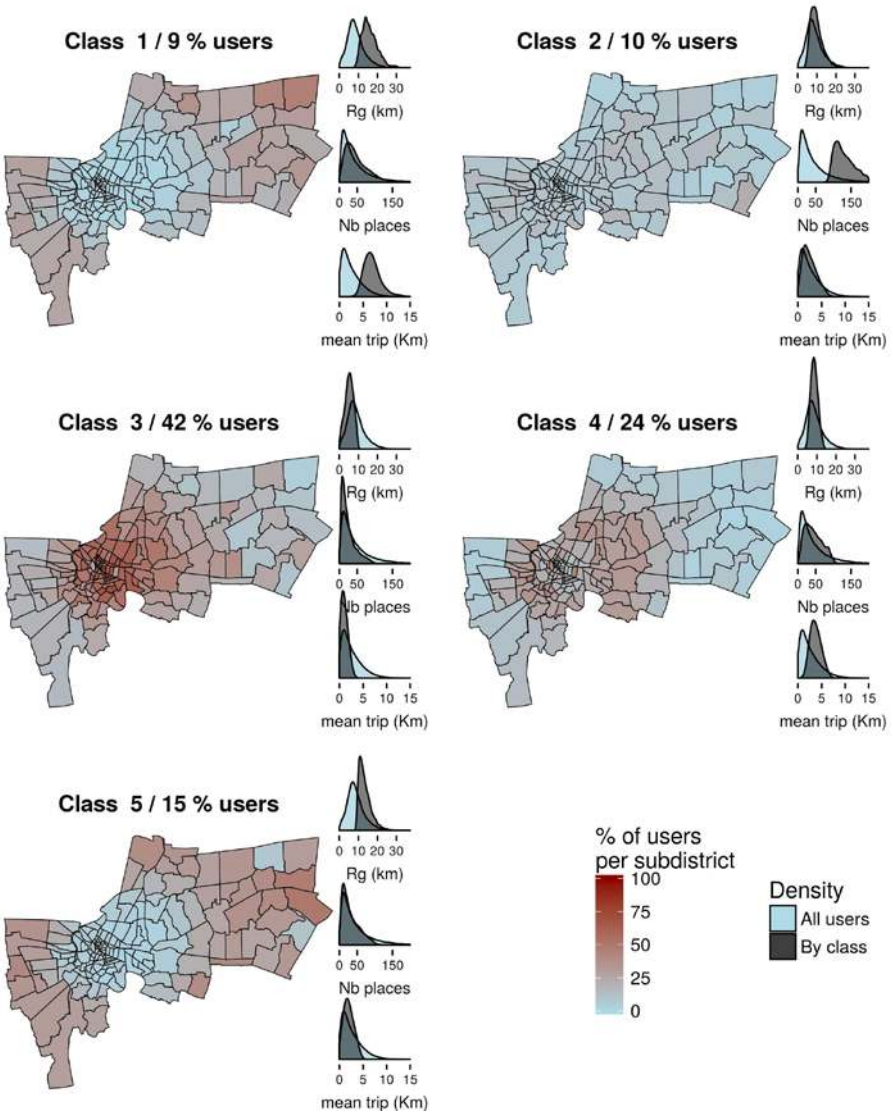

***Figure 10**: Classification of mobility profiles into 5 categories.*

This kind of information is useful and easy to implement in an agent based model, as a synthetic agent could have a differential probability to pick mobility characteristics according to the place he lives. These movement rhythms can be then enriched with the description of probable places visited by crossing these results with Bangkok's land-use typology.

A time slot-based analysis (Figure 11) demonstrates that the entertainment and trade places are very popular in the afternoons and early evenings, especially on

Saturdays. The commercial areas with fewer entertainment places follow the same trend but to a lesser extent. Schools also have interesting curves, with notably a very prominent peak in visits between 8am and 9am on weekdays and absent on weekends. Universities have a lot of visitors on weekdays between 9am and 5pm and much less on weekends. The places situated on transport lines display two peaks of visits during the week, one around 9am and the other around 6pm, which correspond to the peak hours of the two main waves described above. Places of worship, which can also be touristic places, are visited throughout the week and most particularly on Sundays.

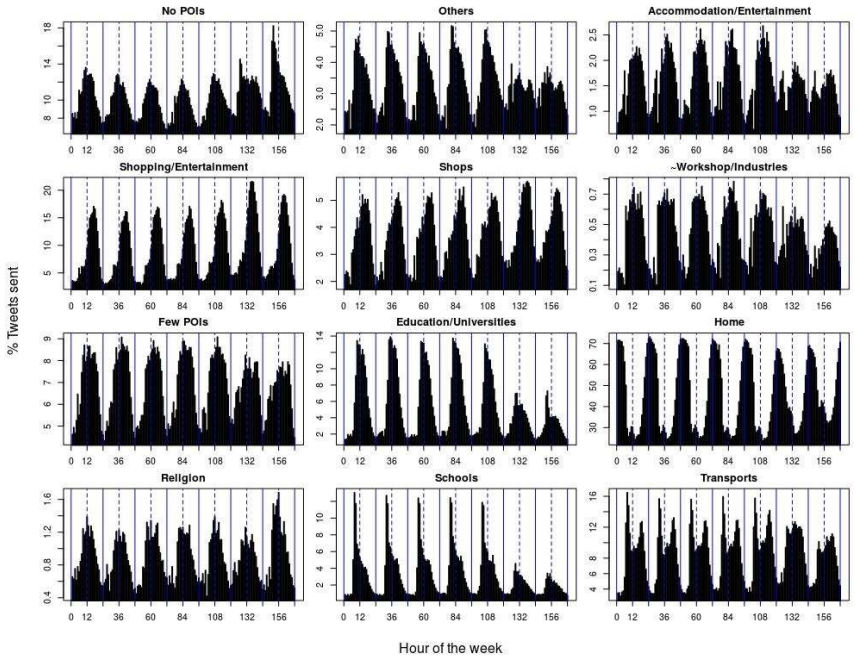

***Figure 11**: Percentage of time slot-wise tweets according to the type of place visited over a typical week.*

Places for which we do not have information on their land use (no POI) have more visitors in the weekend than during the week. These may be peripheral residential areas probably linked to physical social relations, typically to *go see friends or family*. The places where there is no commercial activity are visited mostly in the morning during weekdays and very less during the weekend. These could be transition places between the residence and the workplace.

We've also calculated the distances between home and various types of places where an activity is carried out, which makes it possible to know the scope of attraction for certain activities and efforts made by the users to go there (Figure 12).

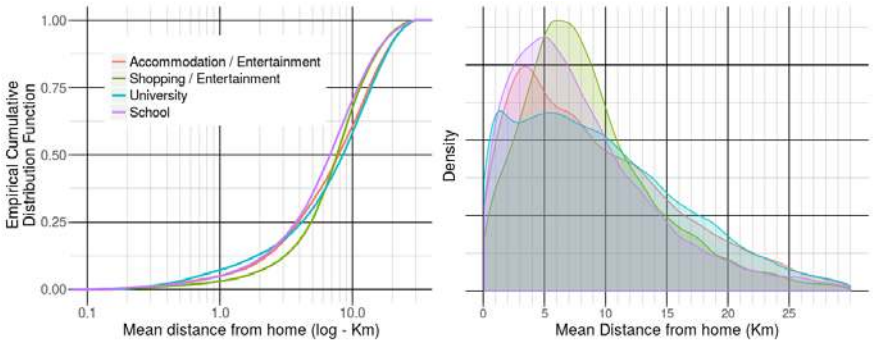

***Figure 12**: Proportion of movement made to go for an activity according to its distance from the place of residence.*

Figure 12 shows that among the people that were in a cell with a school, 25% of the latter were located at less than 4 km from their home. Among the people that frequent universities, 20 % cover less than 2km from their place of residence, which therefore might be close to the campus and 20 % cover more than 20km, which could be the case when a student stays with his parents and has to go across the city to reach his place of study. 25% the people that did some activities linked with shopping and entertainment traveled less than 5 km, but the highest density is found at 6 and 7 km, which suggest that people are more likely to travel over long distances to perform this kind of activities.

We then looked into the frequency with which the users visit places depending on their function, which can be associated with a time of return to the place. The functions of the places have thus been classified according to the frequency of the visits. For example, a school receives a limited number of visitors but the individuals who visit it go there very often. On the contrary, a shop may receive many visits but the visitors taken individually may go there only rarely. Figure 13 shows the preferences for the types of places visited through their rank in the activity space of our sample. Higher the ranking, more this type of place is visited occasionally, which makes it possible to get a hierarchy of the activities according to the persons. Thus, educational institutions (schools and universities) are very often the places visited the most by the people who go there directly from their residence (Rank 1). Entertainment and retail places are among the places most visited by the entire sample with a large number of persons visiting them occasionally (rank higher than 4-5).

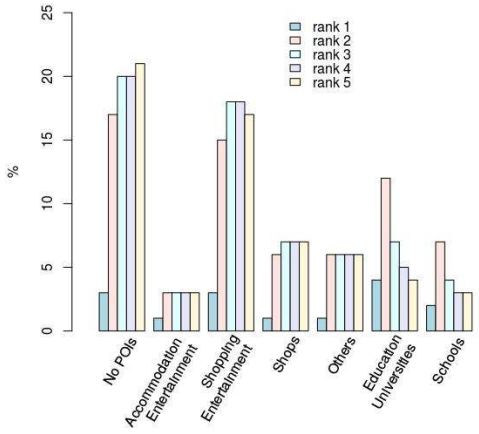

***Figure 13**: Rank-wise percentage of the functions of the places visited.*

All these results are thus in line with the intuitions that we might have about hourly visits to the types of places and they allow us to quantify them quite precisely. Figures 11 and 13 provide overall information on the temporal attractiveness and frequency of visits to different assemblages of places, and Figure 12 show the density distribution of distances that Twitter users are likely to cross starting from their home to perform some activities. These informations are useful to elaborate an agent-based model on mobility, as an agent could pick different time schedules and mobility behavior according to a particular category (children that go to school, student, or worker).

DISCUSSION & CONCLUSION

In this paper we have presented a study of daily mobilities in Bangkok using Twitter data. We chose to look into mobility using the concept of activity space, that is, all the places visited and activities defined at an individual level in time and in space. We relied on large emerging databases for analyzing urban mobilities using geo-localized and dated messages from thousands of Twitter users as well as for characterizing land use by using points of geographic interest from Google Places database.

Using geo-localized Tweets in Bangkok, we have presented different methods that allowed us to know the time when the users visit the places and establish mobility profiles according to the place of residence, number of places visited and distances covered. These results might have been different in another time period, even if the high duration of data collection may tend to mitigate these differences.

Using a layer of land use in Bangkok generated with Google POIs allowed us to link potential goals for these individual movements. Thus, the places that are the most visited and closest to home, the school for example, are less dependent on a

center-periphery structure than the places less visited but further away from the residence such as nightlife places.

Thus, this data should allow us to create individual schedules by relying on the estimation of the place of residence, distance from the activities as well as function-wise and time slot-wise attraction potential of the places and help us to calibrate our agent-based mobility model (Huraux et al., 2017). The latter can thus be expected to reproduce daily mobility dynamics in Bangkok, which could be used to geolocalize the interactions between human beings and dengue vectors in a precise manner. These will then provide an original cartography of the risk of infection at a scale, which has never been carried out till today.

But sensitive analysis has to be performed as the landuse classification might depend on the size and shape of the spatial entity. Nevertheless, this approach could also be used for instance in a CDR activity-based analysis by classifying landuse in Voronoi polygons, even if we might end up with slightly different results. Furthermore, Google is not the only provider of Points of Interest and other sources could be used, such as Foursquare or Facebook. Comparison of these spatial datasets will be done in a future work.

The question of representativity of Twitter users with respect to the population of the city is always an important issue. We've shown here that the estimated home location is correlated with the sub-district population, but we don't know if our sample is correlated with the socio-cultural and economic background of Bangkok's citizens. However, we assume this general criticism in the specific context of the study of dengue, given that the population age groups at risk in Bangkok might correspond to the age group of the main users of virtual social networks. Furthermore, Twitter data only gives a snapshot of where and when people have been, and as far as our knowledge, there is no proper way to quantify the distortion with the reality on such a large sample. People could have over tweeted in anecdotal places according to their real activity space and under-tweeted in more important places such as their office or places they visit more often.

The question of the intention beyond the creation of such digital footprint is also raised by social network studies. We should also consider that there are more and more gateways between social networks, as it's possible to post an Instagram picture or a check-in from Swarm on Twitter. Unfortunately, we have not recorded the source of the message, so we cannot quantify the share of each sources (Twitter from Android, Instagram, Foursquare, etc...) but this should be considered for future studies. For instance, a geolocated picture from Instagram shared publicly on Twitter does not exhibit the precise location of the user but the location of the picture, selected from a list of nearby places - but the user can choose anyplace in the database. Even if our bot detection algorithm might have deleted some of such users, some bias might still exist in our dataset, which could also explain the over-representation of some places (i.e., touristic and commercial places downtown).

Moreover, geolocated tweets shouldn't be used as a standalone dataset, and should be merged with other large and easily accessible Internet databases, such as Facebook or Foursquare check-ins, to enhance the attractivity pattern of the city. These relatively new sources of data are full of biases, (*i.e.* does the numerical footprint reflect the real mobility pattern of the targeted group?) and should therefore be mixed with more "classic" mobility data, such as institutional time use survey or qualitative interviews from the field, to have a more precise view of the studied phenomena.

Finally, using the Twitter data without the user being aware of what the data he produced might be used for raises the same kind of ethical problems as the use of CDR. But in the first case, the user might know that he shared publicly some information with a geographical context and is therefore widely accessible, contrary to CDR, where the users don't expect that the meta-data resulting from a phone call could be re-used for other purposes than billing and network maintenance. We provided here a 3 stage anonymization: spatially, by merging tweets into a 180m cell, temporally by aggregating the tweets into a typical week and thematically with our land-use typology that doesn't display the precise activity on an user. This approach make it possible to perform some spatial analysis with a relatively privacy protection of the users of the dataset, as it is not possible to ensure a 100% secure anonymization for all the users in the dataset (de Montjoye et al., 2013)

REFERENCES

ANDRIENKO N., ANDRIENKO G., STANGE H., LIEBIG T., HECKER D. (2012), "Visual Analytics for Understanding Spatial Situations from Episodic Movement Data", *KI - Künstliche Intelligenz,* 26, pp. 241–251. doi :10.1007/s13218-012-0177-4

CALABRESE F., DIAO M., DI LORENZO G., FERREIRA J., RATTI C. (2013), "Understanding individual mobility patterns from urban sensing data : A mobile phone trace example", *Transportation Research Part C : Emerging Technologies*, 26, pp. 301–313. doi :10.1016/j.trc.2012.09.009

CEBEILLAC A., DAUDE É., VAGUET A. (2017), *Discontinuités spatiales, santé et mobilités. Analyses et typologies de Google POI et de Tweets pour caractériser les structures spatiales et les dynamiques d'attractivités de Bangkok (Thaïlande)*, Actes du colloque SAGéo, Conférence internationale de Géomatique et Analyse Spatiale.

CEBEILLAC A., RAULT Y.-M. (2016), "Contribution of geotagged Twitter data in the study of social groups' activity space : The case of the upper middle classes in Delhi, India", *Netcom*, 30-3/4 http://journals.openedition.org/netcom/2529

DAUDE É., VAGUET A., PAUL R. (2015), « La dengue, maladie complexe », *Nat. Sci. Soc.*, 23, pp. 331–342. doi :10.1051/nss/2015058

DE MONTJOYE Y.-A., HIDALGO C. A., VERLEYSEN M., BLONDEL V. D. (2013), *Unique in the Crowd : The privacy bounds of human mobility*, Scientific Reports 3. doi :10.1038/srep01376

DEVILLE P., LINARD C., MARTIN S., GILBERT M., STEVENS F. R., GAUGHAN A. E., BLONDEL V. D., TATEM A. J. (2014), "Dynamic population mapping using mobile phone data", *Proceedings of the National Academy of Sciences*, 111, pp. 15888–15893. doi :10.1073/pnas.1408439111

ESTER M., KRIEGEL H.-P., SANDER J., XU X. (1996), *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, KDD-96 Proceedings.

FERRARA E., VARUL O., DAVIS C., MENCZER F., FLAMMINI A. (2015), *The rise of social bots*, Communications of the ACM, pp. 96–104. doi :10.1145/2818717

FLEURY A., MATHIAN H., SAINT-JULIEN T. (2012), « Définir les centralités commerciales au cœur d'une grande métropole : le cas de Paris intra-muros », *Cybergeo : European Journal of Geography [Online], Space, Society,Territory*, doi :10.4000/cybergeo.25107

FUCHS G., ANDRIENKO G., ANDRIENKO N., JANKOWSKI P. (2013), "Extracting Personal Behavioral Patterns from Geo-Referenced Tweets", *AGILE*.

GONZÁLEZ M. C., HIDALGO C. A., BARABÁSI A.-L. (2008), "Understanding individual human mobility patterns", *Nature*, 453, pp. 779–782. doi :10.1038/nature06958

HÄGERSTRAND T. (1970), *What about people in regional science ?* Papers of Regional Science Association, 24, pp. 7–21.

HAWELKA B., SITKO I., BEINAT E., SOBOLEVSKY S., KAZAKOPOULOS P., CARLO R. (2014), "Geo-located Twitter as the proxy for global mobility patterns", *Cartography and Geographic Information Science*.

HORTON F. E., REYNOLDS D. R. (1971), "Effects of Urban Spatial Structure on Individual Behavior", *Economic Geography*, 47, 36. doi :10.2307/143224

HUANG Q., CAO G., WANG C. (2014), "From Where Do Tweets Originate ? : A GIS Approach for User Location Inference", In : *Proceedings of the 7th ACM SIGSPATIAL International Workshop on Location-Based Social Networks, LBSN '14*. ACM, New York, NY, USA, pp. 1–8. doi :10.1145/2755492.2755494

HURAUX T., MISSLIN R., CEBEILLAC A., VAGUET A., DAUDE É. (2017), *Modélisation de l'impact des îlots de chaleur urbains sur les dynamiques de population d'Aedes aegypti, vecteur de la dengue et du virus Zika*, Actes du colloque SAGéo, Conférence internationale de Géomatique et Analyse Spatiale.

JURDAK R., ZHAO K., LIU J., ABOUJAOUDE M., CAMERON M., NEWTH D. (2015), "Understanding human mobility from Twitter", *PloS one*, 10, e0131469.

KAUFMAN L., ROUSSEUW P. J. (1990), "Finding Groups in Data : An Introduction to Cluster Analysis", *Biometrics*, 47, 788. doi :10.2307/2532178

KHAN S.F., BERGMANN N., JURDAK R., KUSY B., CAMERON M. (2017), *Mobility in Cities : Comparative Analysis of Mobility Models Using Geo-tagged Tweets in Australia.*

LE BIGOT B. (2016), "The backpackers' "round the world" trip, standardised journey ?" *Via@* [online].

LENORMAND M., PICORNELL M., CANTU-ROS O. G., TUGORES A., LOUAIL T., HERRANZ R., BARTHELEMY M., FRIAS-MARTINEZ E., RAMASCO J. J. (2014), "Cross-Checking Different Sources of Mobility Information", *PLoS ONE*, 9, e105184. doi :10.1371/journal.pone.0105184

LIMKITTIKUL K., BRETT J., L'AZOU M. (2014), "Epidemiological Trends of Dengue Disease in Thailand (2000–2011) : A Systematic Literature Review", *PLoS Neglected Tropical Diseases*, 8, e3241. doi :10.1371/journal.pntd.0003241

LONGLEY P. A., ADNAN M., LANSLEY G. (2015), "The geotemporal demographics of Twitter usage", *Environment and Planning A*, 47, pp. 465–484.

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*, 70, pp. 11–25. doi :10.1016/j.apgeog.2016.03.001

MCNEILL G., BRIGHT J., HALE S. A. (2016), "Estimating Local Commuting Patterns From Geolocated Twitter Data", *arXiv preprint* arXiv :1612.01785.

MISSLIN R., DAUDE E. (2016), *Génération d'environnements artificiels pour la simulation spatiale d'arboviroses*.

MORSTATTER F., PFEFFER J., LIU H., CARLEY K. M. (2013), "Is the sample good enough ? Comparing data from twitter's streaming api with twitter's firehose", *arXiv preprint* arXiv :1306.5204.

MURDOCK V. (2011), "Your mileage may vary : on the limits of social media", *SIGSPATIAL* Special 3, pp. 62–66.

NOULAS A., SCELLATO S., MASCOLO C., PONTIL M. (2011), "An empirical study of geographic user activity patterns in foursquare", *ICwSM,* 11, pp. 70–573.

PERCHOUX C., CHAIX B., CUMMINS S., KESTENS Y. (2013), "Conceptualization and measurement of environmental exposure in epidemiology : Accounting for activity space related to daily mobility", *Health & Place*, 21, pp. 86–93. doi :10.1016/j.healthplace.2013.01.005

PERKINS T. A., GARCIA A. J., PAZ-SOLDAN V. A., STODDARD S. T., REINER R. C., VAZQUEZ-PROKOPEC G., BISANZIO D., MORRISON A. C., HALSEY E. S., KOCHEL T. J., SMITH D. L., KITRON U., SCOTT T. W., TATEM A. J. (2014), "Theory and data for simulating fine-scale human movement in an urban environment", *Journal of The Royal Society Interface*, 11, 20140642–20140642. doi :10.1098/rsif.2014.0642

QUESNOT T. (2016), « L'involution géographique : des données géosociales aux algorithmes », *Netcom*, pp. 281–304. http://journals.openedition.org/netcom/2545

RIPLEY B. D. (1988), *Statistical inference for spatial processes*, Cambridge: Cambridge University Press.

SLOAN L., MORGAN J., BURNAP P., WILLIAMS Ma. (2015), "Who Tweets ? Deriving the demographic characteristics of age, occupation and socialClass from Twitter User Meta-Data", *PloS ONE*, doi :doi :10.1371/journal.pone.0115545

SONG C., KOREN T., WANG P., BARABÁSI A.-L. (2010), "Modelling the scaling properties of human mobility", *Nature Physics*, 6, pp. 818–823. doi :10.1038/nphys1760

SORICHETTA A., BIRD T. J., RUKTANONCHAI N. W., ZU ERBACH-SCHOENBERG E., PEZZULO C., TEJEDOR N., WALDOCK I. C., SADLER J. D., GARCIA A. J., SEDDA L., TATEM A. J. (2016), "Mapping internal connectivity through human migration in malaria endemic countries", *Scientific Data*, 3, 160066. doi :10.1038/sdata.2016.66

STEIGER E., ALBUQUERQUE J. P., ZIPF A. (2015a), "An advanced systematic literature review on spatiotemporal analyses of twitter data", *Transactions in GIS*, 19, pp. 809–834.

STEIGER E., WESTERHOLT R., RESCH B., ZIPF A. (2015b), "Twitter as an indicator for whereabouts of people ? Correlating Twitter with UK census data", *Computers, Environment and Urban Systems*, 54, pp. 255–265. doi :10.1016/j.compenvurbsys.2015.09.007

STODDARD S. T., FORSHEY B. M., MORRISON A. C., PAZ-SOLDAN V. A., VAZQUEZ-PROKOPEC G. M., ASTETE H., REINER R. C., VILCARROMERO S., ELDER J. P., HALSEY E. S., KOCHEL T. J., KITRON U., SCOTT T. W. (2013), "House-to-house human movement drives dengue virus transmission", *Proceedings of the National Academy of Sciences*, 110, pp. 994–999. doi :10.1073/pnas.1213349110

TATEM A. J., ROGERS D. J., HAY S. I. (2006), "Global Transport Networks and Infectious Disease Spread", In : *Advances in Parasitology*, Elsevier, pp. 293–343.

WESOLOWSKI A., EAGLE N., TATEM A. J., SMITH D. L., NOOR A. M., SNOW R. W., O B. C. (2012), "Quantifying the Impact of Human Mobility on Malaria", *Science*, 338, pp. 267–270. doi :10.1126/science.1223467

WESOLOWSKI A., QURESHI T., BONI M. F., SUNDSØY P. R., JOHANSSON M. A., RASHEED S. B., ENGØ-MONSEN K., BUCKEE C. O. (2015), "Impact of human mobility on the emergence of dengue epidemics in Pakistan", *Proceedings of the National Academy of Sciences*, 112, pp. 11887–11892. doi :10.1073/pnas.1504964112

ZAGHENI E., KIRAN V. R., STATE B. (2014), "Inferring international and internal migration patterns from Twitter Data", In : *World Wide Web Conference Committee (IW3C2).*

ZANDBERGEN P. A., BARBEAU S. J. (2011), "Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones", *Journal of Navigation*, 64, pp. 381–399. doi :10.1017/S0373463311000051