



HAL
open science

Understanding Human Movement Semantics: A Point of Interest Based Approach

Ionut Trestian, Kévin Huguenin, Ling Su, Aleksandar Kuzmanovic

► To cite this version:

Ionut Trestian, Kévin Huguenin, Ling Su, Aleksandar Kuzmanovic. Understanding Human Movement Semantics: A Point of Interest Based Approach. [Research Report] RR-7716, 2011, pp.21. inria-00616876v1

HAL Id: inria-00616876

<https://inria.hal.science/inria-00616876v1>

Submitted on 1 Feb 2012 (v1), last revised 1 Feb 2012 (v2)

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.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

***Understanding Human Movement Semantics:
A Point of Interest Based Approach***

Ionut Trestian — Kévin Huguenin — Ling Su — Aleksandar Kuzmanovic

N° 7716

August 2011

—— Distributed Systems and Services ——

R *apport
de recherche*

Understanding Human Movement Semantics: A Point of Interest Based Approach

Ionut Trestian* , Kévin Huguenin[†] , Ling Su* , Aleksandar Kuzmanovic*

Theme : Distributed Systems and Services
Networks, Systems and Services, Distributed Computing
Équipe-Projet ASAP

Rapport de recherche n° 7716 — August 2011 — 23 pages

Abstract: The recent availability of human mobility traces has driven a new wave of research on human movement with straightforward applications in wireless/cellular network algorithmic problems. However, all of the performed studies isolate movement from the environment that surrounds people, *i.e.* the points of interest that humans visit.

In this paper we revisit the human mobility problem with new assumptions. We find it reasonable to believe that human movement is not independent of the surrounding locations; most of the time people travel with specific goals in mind, visit specific points of interest, and constantly revisit favorite places. Points of interest are also differently spread. Motivated by this, we aim at studying the movement interaction that we humans have with the differently spread points of interest around us.

More specifically, by analyzing GPS mobility traces of a large number of users located across two distinct geographical locations we find that: (*i*) users do not particularly visit only locations that are close to them but also the functional aspect of the location matters with this respect too, (*ii*) while users in different part of the globe exhibit different time of day behavior, we also find that there is a striking correlation in the frequency of visits to the basic points of interest categories that we have analyzed.

Key-words: Human movement; Mobility traces; GPS; Points of interest

* Department of Electrical Engineering and Computer Science, Northwestern University

[†] School of Computer Science, McGill University

Mieux comprendre la sémantique des mouvements humains: Une approche par points d'intérêt

Résumé : La disponibilité récente de traces de mobilité humaine a engendré une vague de recherche sur les modèles de mobilité ayant des applications directes dans les réseaux cellulaires/sans fil. Cependant, les études menées jusqu'à présent ignorent l'environnement dans lequel se trouvent les personnes en mouvement : les points d'intérêt que ces derniers visitent et qui, par conséquent, guident leurs déplacements.

Cet article revisite le problème de la modélisation des déplacements humains en se basant sur de nouvelles hypothèses. Considérer que les déplacements humains ne sont pas indépendants des lieux environnants semble être une hypothèse raisonnable : la plupart du temps, les gens se déplacent avec un but précis en tête, visitent des lieux précis et se rendent régulièrement dans leurs endroits favoris. Ces points d'intérêt sont répartis de manière non uniforme. A partir de ces observations, on étudie les interactions entre les déplacements humains et la répartition des points d'intérêt.

Plus précisément, en analysant des traces de mobilité GPS de personnes évoluant dans deux zones géographiques distinctes du globe, il apparaît que : les gens ne basent pas le choix des lieux qu'ils visitent uniquement sur la proximité mais également sur des aspects fonctionnels de ces derniers ; si l'activité et la mobilité de gens originaires de régions distinctes présentent des motifs temporels différents, les fréquences auxquelles ils visitent certaines catégories de points d'intérêt ont des similarités frappantes.

Mots-clés : Mouvement humain ; Traces de mobilité ; GPS ; Points d'intérêt

1 Introduction

Human movement has been studied extensively in the past several years. This is mainly due to the rather obvious implications that results might have in fields such as urban planning, disease prevention, mobile advertising, mobile infrastructure placement, *etc.* As such, researchers have tried to understand statistical properties, and also proposed mathematical models that try to capture particular aspects of human movement, for example by relating to the movements of banknotes or subatomic particles¹. But the matter of fact is that humans are not banknotes and they are not subatomic particles. When they move they are driven by concrete goals such as buying groceries, dining at a restaurant, watching a movie, *etc.* And although it is indeed true that models (*e.g.* Lévy flights) manage to capture important aspects of movement (such as the distribution of trip sizes), they fail to capture the particular *semantics* that gets incorporated into human movement.

Furthermore, as we noted above, it is sensible to believe that most human movement is indeed driven by certain purposes. We can therefore only note that specific and important characteristics of human movement that should get captured in the aforementioned models are linked to two important factors. The first factor consists of *people's daily routines* (such as visiting a coffee shop in the morning, going to work, and/or shopping or going to a restaurant in the evening). A second factor is the actual *location* of the different *points of interest* that humans visit. For example, in a region there could be restaurants located almost everywhere but there might be only several stadiums that still manage to attract a large number of people at certain times. The combination of these two factors (human daily routines, and point of interest placement) could therefore explain the different characteristics that older studies on human movement have observed.

There are several applications that can benefit from understanding the above characteristics. With the current advent of mobile devices such as smartphones and tablets, understanding human interactions with points of interest can have a big impact on mobile infrastructure placement, transfer scheduling algorithms, mobile advertising, mobile social networking (to point out just a few). Furthermore, understanding how people interact with the points of interest around them could help urban planners into placing new ones. It is also important to mention that understanding how far humans move to visit places can help us better understand the propagation of viruses (including mobile phone ones), and contagious diseases with the hope of finding ways to prevent the spread.

In this paper we aim to answer the following research questions. What are the actual semantics of human movement? What are the locations that people visit, with what frequency, and at what time scales? What is the influence of the spread of these locations on human movement? We aim to address the above questions by extensively studying two movement traces from two different geographic regions. The traces contain users carrying GPS devices inside mobile phones (155 users are from China and 4,429 users are from the United States). We enhance the above traces by using categorized points of interest from GPS devices. Furthermore, because of the geo-diverse nature of our traces we manage to find and point out cultural differences with respect to humans and the points of interest around them. For example we found that the Chinese users have a lower movement span when compared to their United States counterparts (most probably related to the fact that people in the United States usually live in the suburbs and commute to work). Further details about the datasets we use will be given in Section 2.

Regarding our contributions, to the best of our knowledge, we are the first to systematically study the relationship between human mobility and points of interest at such a scale. This is our main contribution. Among a number of insights that we provide, the key one is that we present strong evidence relating user affinity at different time scales towards certain location types such as restaurants, shops, and stadiums (some locations, *e.g.* stadiums, are not located particularly close to users).

To understand the relationship that exists between human movement and the points of interest that surround them, we apply an energy model inspired from physics that we build over a modified

¹We are referring to the Brownian Motion Mobility Model [10], and the well known human movement study based on the one dollar bill [9].

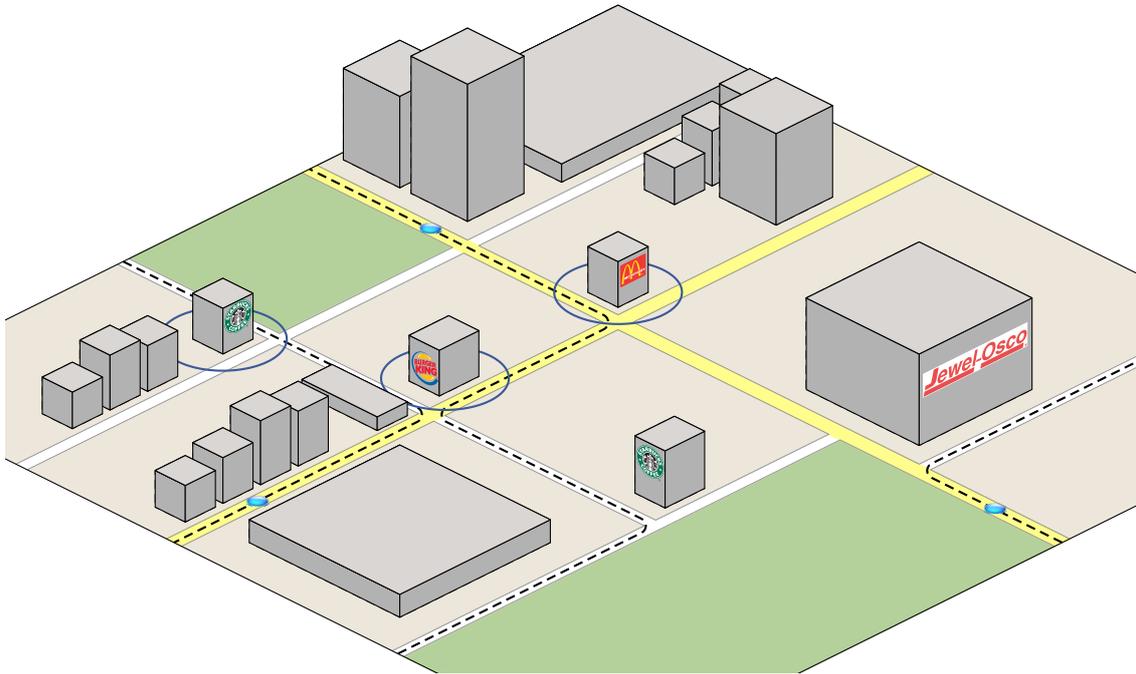


Figure 1: Overview of mobility and point of interest datasets.

kd-tree data structure. We further use this methodology to analyze three aspects. First, we perform user profiling in order to identify broad user interests in location types. We find that most users are interested in dining at restaurants, shopping, attending sports events while they are not so much interested in visiting libraries or hospitals. Second, we analyze the frequency with which users visit points of interest. We find that certain points of interest types preserve their popularity when it comes to how often users visit them. We also find that although some points of interest attract a large number of users, they do not attract them with a high visiting frequency (*e.g.* stadiums). Third, we analyze time of day effects related to users visiting points of interest and we find for example that the Chinese users are trying to optimize their activities. One example of such behavior is in users running errands during their lunch break.

We carried our analysis further by analyzing user interest in specific popular businesses such as Starbucks, McDonalds, and Burger King. We find that such businesses manage to attract a large number of users. Because some of the analyzed businesses carry WiFi access points, such a result could help mobile providers that are willing to form partnerships with businesses for content offloading purposes.

This paper is organized as follows. In Section 2, we describe the trajectory and point of interest datasets that we further use in our paper. At first, we analyze both user trajectories and points of interest spread in an independent manner to gather insights that would be useful in our later analysis. As such, in Section 3.1 we perform a preliminary analysis of user trajectories, and in Section 3.2 an analysis on the placement and spread of different types of points of interest. We further proceed by jointly analyzing the interactions that humans have with the points of interest around them. In Section 4 we present the interaction methodology that we use in this paper. In Section 5 we carry out our evaluation and present the results of our study. We discuss implications of our research in Section 6. We acknowledge related work in Section 7, and conclude our paper in Section 8.

2 Datasets

In order to correlate user movement with daily activities we require two types of data: (i) *user mobility data* (obtained from users carrying GPS tracking devices in mobile phones, dashed lines in Figure 1), and (ii) *point of interest data*—*i.e.* locations of *e.g.* restaurants (McDonalds in Figure 1), churches, shops (Jewel-Osco in Figure 1), parks, stadiums, hospitals, *etc.* . We detail how we obtain the two types of data below.

2.1 Mobility datasets

There are two mobility datasets containing user trajectories that we are using in this paper. In an effort to make our study more generic we use datasets that were collected in different regions of the globe. We further describe them.

- The first dataset we use is a GPS trajectory dataset collected by Microsoft Research Asia in their GeoLife project [15]. It contains the data of 155 users that were monitored over a period of over two years. This dataset records a broad range of outdoor movements, including not only daily routines like going home and going to work but also entertainment and sports activities such as shopping, sightseeing, dining, hiking, and cycling [41, 42]. We will further refer to this dataset in our paper as the Chinese dataset.
- The second dataset we are using is a dataset collected in the United States from the backbone of a mobile provider network. It contains the trajectories of 4,429 users that fielded applications that report their GPS locations back to the mobile provider for different purposes. This is actually a service offered by the mobile provider in order to provide location based services to subscribing users. This dataset was collected for the period of one week. We will refer to this dataset as the United States dataset. Although a lot of the movement is in the Chicago area, there is a large number of users that are located or travel to other parts of the United States.

Table 1: Characteristics of mobility datasets

Dataset	United States	Chinese
Timespan	1 week	2 years
Subjects	4,429	155
Samples	708,079	23,209,033
Mass center	36.880,-85.932	39.309,113.016

Table 1 summarizes the various datasets characteristics. Both the datasets contain the same type of information. This basically means that for each user we have an anonymized identifier, and timestamped latitude and longitude tuples. Such information suffices for the purposes of our analysis.

2.2 Point of Interest datasets

The above datasets are enhanced by extracting datasets containing the GPS coordinates of points of interest from China and the United States.

There are several ways in which one can proceed with regards to this. One way would be that based on the trajectories extracted in the previous section to crawl Google Maps (or similar services) and extract close by-interesting locations. This process would be very cumbersome and slow mainly because of the large number of location samples that can be noted in Table 1.

A second way would be to use the point of interest data that users share publicly on forums and specialized websites [28, 29]. For example, users might share favorite restaurants or the locations of

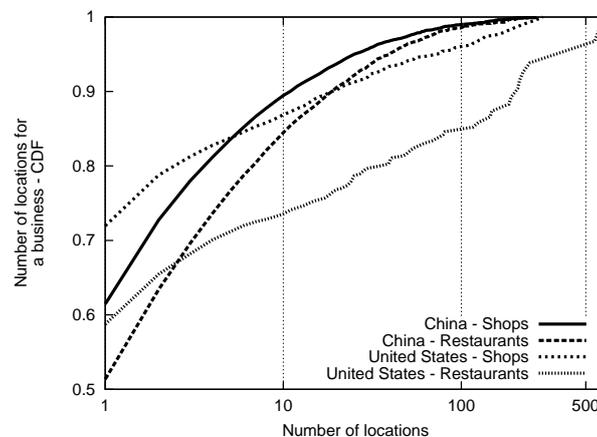


Figure 2: Distribution of the number of distinct locations that businesses have across China and across Illinois.

speed cameras. We have found this data to be non-comprehensive as the websites contain mostly locations that have sparked the interest of a few users and not of the general public.

The third way (and the method we have adopted in this paper) is to use the points of interest from GPS devices. In fact, the very notion of *point of interest* that we employ in this paper is part of the terminology used in GPS devices. Most useful, GPS devices categorize points of interest into about 40-60 classes such as restaurants, shops, airports, theaters, *etc.* We therefore have the latitude, and longitudes of the locations. Statistics for some of the categories we extracted can be seen in Table 2. We show the points of interest for China and the states of Illinois, Indiana, and Michigan that are located in the United States Midwest (that is the area where most of our users are). Note that for each point of interest there might be several entries corresponding to for example multiple entrances of the same building. In the table below we summarize only unique occurrences of the same point of interest.

Table 2: Point of interest statistics

Dataset	Illinois	Indiana	Michigan	China
Restaurants	19,400	9,231	14,671	116,095
Shops	56,768	25,454	46,199	267,541
Govt. Offices	4,645	2,890	5,185	137,837
Hospitals, Clinics	748	466	963	54,545
Libraries	692	372	585	40,361
Stadiums	71	24	61	17,868
Population [mi]	12.91	6.42	9.96	1,331
Size [1000 sq. mi]	57.9	36.4	96.7	3,705

Note that there is a bigger number of private businesses relative to the population (*e.g.* restaurants and shops) in the United States than in China. Oppositely when contrasting the number of government offices with the numbers shown for private businesses there is a much higher number of government offices (relative to private businesses) in China than in the United States. A few other insights can be extracted from Figure 2. It depicts the distribution of the number of locations that businesses have. If one compares shops and restaurants across Illinois (other states exhibit similar characteristics), and China we can see that although Illinois has a lot more businesses with a single location, it also leads in the number of businesses with a lot of locations (basically large corporations). Although China leads the medium number of locations range (medium sized businesses), it lacks some of the larger corporations popular across the United States.

In this paper we restrict ourselves to just a few types of points of interest. We will use in the remainder of this paper the following categories: restaurants, shops, companies, government offices, hospitals or clinics, libraries, and stadiums. We select restaurants, shops, and stadiums for obvious reasons; they attract a large number of users and they hold commercial interest. Another reason is that such popular locations often carry WiFi access points or additional mobile infrastructure to accommodate larger crowds. We select the rest based on the fact that the activities they represent are (similar to the previous ones) usually linked to people’s daily routines. It is important to understand how these locations are spread and we carry out this analysis in Section 3.2. Note also that when showing graphs, some of the categories might be missing in certain graphs. We do this in an effort to not overwhelm the presentation since some categories often exhibit similar behaviors.

3 Preliminary analysis

In this section we perform independent analysis on user trajectories and on point of interest placement. Not all humans exhibit the same behavior and although similar patterns can be observed across individuals there are definitely particularities in movement span that we point out in this section. Points of interest also have a specific placement and a specific spread (depending on the type) and we carry out analysis that potentially explains our evaluation results.

3.1 Trajectory analysis

Carrying out a preliminary trajectory analysis is important because we require a measure that will help us categorize users. Some humans travel over larger distances because of daily commutes related to a job location for example. Other humans optimize everything so that they do not move that much; they live close to work, they shop and dine at local stores and restaurants, *etc.* In order to carry out such an analysis we require an analysis measure that captures the physical space where users carry out their daily activities.

When dealing with trajectories in space, a useful analysis measure is the *radius of gyration*. When dealing with certain physical objects, it gives an insight about the actual size that the object takes in three-dimensional space. For user trajectories, we are considering the distance that all GPS points have with respect to the center of mass of the trajectory and as such, it captures the actual movement span that humans have. Basically in two dimensions the radius of gyration characterizes the linear size taken by a certain user’s trajectory up to time t and is defined as:

$$r(t) = \sqrt{\frac{1}{n(t)} \sum_{0 \leq i \leq n(t)} d(\mathbf{x}_i, \mathbf{x}_{\text{cm}}(t))^2} \quad (1)$$

The function $n(t)$ returns the number of trajectory samples until time t . The point \mathbf{x}_i is the i -th location sample for the user, and the point \mathbf{x}_{cm} is the center of mass of the trajectory until time t . Naturally, we use the Haversine distance² as the distance function $d()$ in the above evaluation.

Figure 3 presents the results of the radius of gyration evaluation for the two datasets. We make the evaluation over the total duration of the traces. Given that the United States trace is only one week long, we present for comparison reasons the radius of gyration of the Chinese users evaluated over the duration of one week. The results show that the United States users travel over bigger distances than their Chinese counterparts. This is likely because a large number of people in the United States live in the suburbs and commute daily to work over somewhat larger distances. Also note that the distribution of the radius of gyration values for the United States users carries a lower variance than for the Chinese users. To summarize the results, the Chinese users travel less but they exhibit more diverse behavior.

Furthermore, Figure 4 displays the radius of gyration for six representative users (three from each dataset). The steps in the figures correspond to movement that takes the user significantly out

²This is the geodesic or great-circle distance between two points that takes the earth’s sphericity into account.

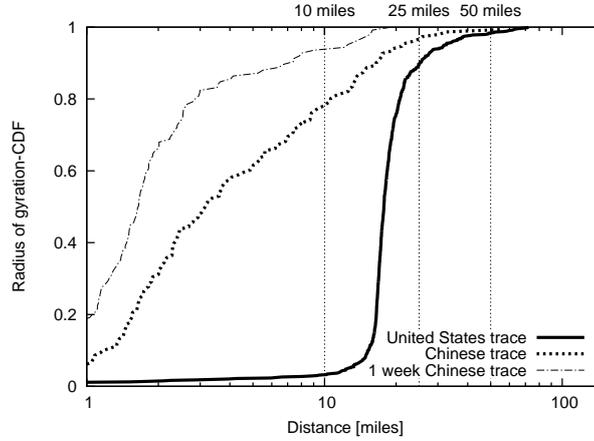


Figure 3: Radius of gyration at the end of the two datasets and across a week of the Chinese dataset. The United States dataset is a week long.

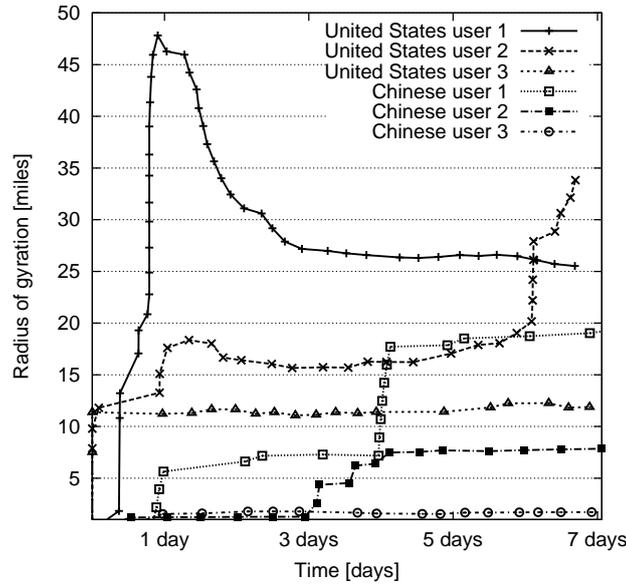


Figure 4: Radius of gyration evolution for six representative users across the two datasets.

of range from the area occupied so far (captured using the center of mass). This is in accordance to what previous studies have observed; humans usually move in a certain area, occasionally taking larger trips that take them out of the space where they usually carry out their activities [22].

To characterize user movement, other studies have used the radius of gyration in conjunction with the *moment of inertia*. This brings the trajectory of each user to a common frame of reference where common movement patterns can be observed. For the purposes of our analysis, considering only the radius of gyration for individual users suffices. We will further use the user radius of gyration in future sections.

Below we carry out an analysis related to the placement and spread of different types of points of interest.

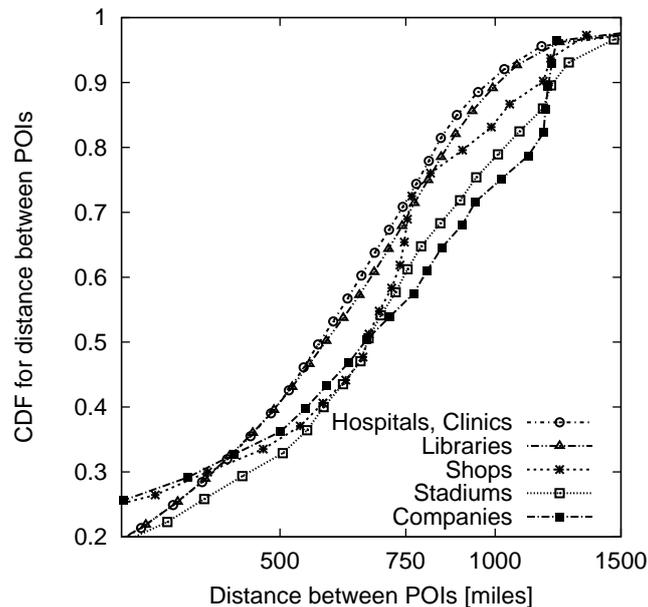


Figure 5: China – Distribution of distances between POIs of a certain type. The curves are ordered based on the legend.

3.2 Point of Interest analysis

In this section we are examining the spread of different point of interest types across the analyzed regions. We do this because the differently spread points of interest have an obvious effect on human trip sizes. For this, we take as metric the distribution of all distances between all points of interest of a certain type. Given that for some point of interest types the number of distances that make up the distribution can be particularly large (in the trillions) we sample the distributions by randomly computing a significant amount of values. Figures 5 and 6 contain the results for China and the state of Illinois respectively.

In Figure 5, one can see the distribution of distances between the points of interest in a certain category for several different categories. Some insights from the figure are as follows. One can note that companies and shops are clustered together in several centers; this is because of the large number of companies and shops in the distribution that have a small distance to one another and also the big jumps close to 750 and 1250 miles that correspond to the distances between large centers (the distance between Shanghai and Beijing for example is somewhere close to 750 miles). Stadiums are also clustered in certain regions; only keeping in mind that stadiums are not a necessity but more of a commodity. Necessarily they might be lacking in certain remote regions which would explain the appearance of the curve. Also note that hospitals, clinics and libraries all have smoother curves corresponding to the fact that they need to be close to evenly spread across the whole country. They are also public services and therefore they do not benefit from being located only in populous areas.

The same kind of insights can be drawn from Figure 6 corresponding to the state of Illinois where a lot of the users from the United States trace are located. Certain points of interest are spread over all the state of Illinois (*e.g.* libraries, clinics, government offices). Other points of interest (*e.g.* restaurants, shops, stadiums) have a different distribution, with more of them being clustered in the biggest centers, such as Chicago, Springfield, and Rockford. In fact, the big steps in some of the curves account for the distances from Chicago to Rockford, and Springfield. This is expected as private businesses are mostly clustered in the most populous areas.

Note that by computing the variance of the distributions we have shown above, we obtained similar insights (although not as expressive). The curves that are smoother have a bigger variance

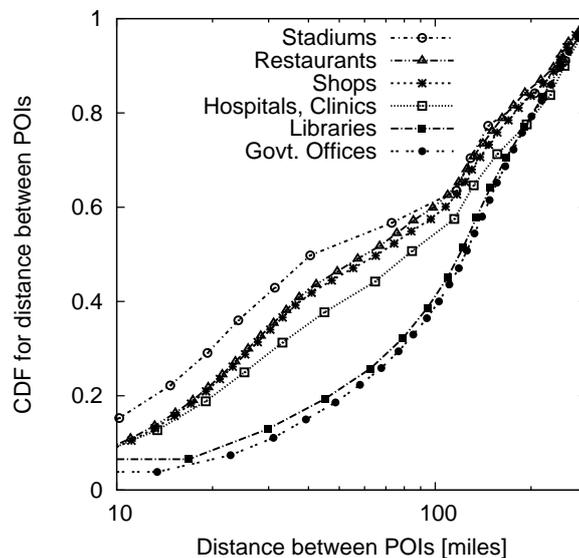


Figure 6: Illinois – Distribution of distances between POIs of a certain type. The curves are ordered based on the legend.

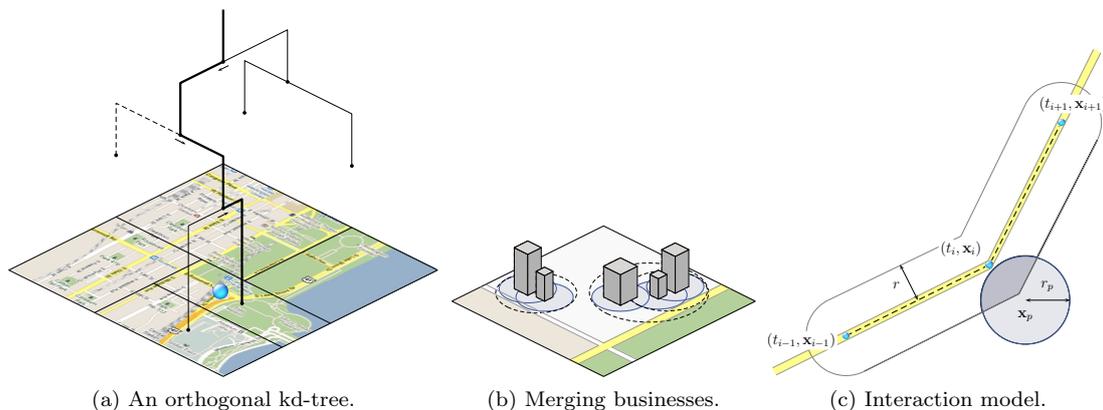


Figure 7: Methodology of our analysis.

than the ones corresponding to points of interest that are clustered together. One can note from this section that most points of interest corresponding to private businesses are clustered in the larger population places (where our users are located) while the public points of interest are more uniformly spread across the whole country/state.

Below, we build on the insights acquired in this section and develop a methodology for analyzing the connection that humans have with the points of interest around us.

4 Methodology

Both the analysis of mobility traces and the actual applications that run on websites or mobile phones for suggesting interesting points of interest around a user's position require an efficient way to find the closest points of interest to a given location. Basically, a request is a point in two-dimensional space and the k closest points of interest among all or a subset of them (when a certain type of point of interest is desired) must be returned. Obviously, this is a classic k nearest-

neighbors (kNN) search in a two-dimensional space with respect to a given distance measure. This distance measure is the Harvesine distance in our case.

To achieve fast k nearest-neighbors search, the set of points of interest needs to be structured based on the latitude and longitude. It is known that r-trees and kd-trees allow fast lookups, *i.e.* logarithmic on average for multiple-dimension data such as our two-dimensional coordinates of points of interest. Both are trees where nodes correspond to nested regions and leaf nodes contain pointers to the points of interest in the corresponding region. Each level of a r-tree splits a region in several possibly overlapping rectangular regions forming a minimum bounding box of the points it contains. Kd-trees partition a region in exactly two sub-regions along a hyperplane, *i.e.* a plane in three-dimensional spaces and a line in two-dimensional spaces. Orthogonal kd-trees split regions only along the axis. Both r-trees and kd-trees exist in an exact and approximate version and have similar performance.

In our analysis, we used orthogonal kd-trees (see Figure 7a) for their simplicity and their performance and opt for exact k nearest-neighbors search (the approximate version only reduces the worst case complexity). Other efficient solutions exist, *e.g.*, ANN [2] or FLANN [12] (set of kd-trees), but they are mostly optimized for high-dimension datasets—such as image descriptors that face the curse of dimensionality—and bring only small improvements for two-dimensional points compared to the overhead incurred in programming. Interestingly enough, the aforementioned techniques work for any symmetric distance function respecting the triangular inequality, *e.g.*, traffic-aware travel time.

The determination of the k nearest-neighbors of a location (shown with a circle in the figure) using a kd-tree operates in two stages (see Figure 7a). First, a depth-first search is performed to reach the leaf corresponding to the region containing the desired location (bold line). At each internal node, the searched location is compared to the corresponding splitting hyperplane thus determining which of the children must be therefore recursively explored (\leftarrow or \rightarrow). The k nearest-neighbors are selected among the points of interest in the same region as the location. Secondly, the recursion is unwound to refine the k nearest-neighbors by looking for closer points of interest in other regions. At each internal node, if the distance between the location and the splitting hyperplane is smaller than the distance to the k -th closest neighbor (so far) the node is recursively explored (plain lines), otherwise, it is ignored (dashed lines).

Our analysis requires a proper formalization of the notion of interactions between users and possibly grouped points of interest. The mobility trace of a user is given by a list of couples (t_i, \mathbf{x}_i) , called *snapshots*, ordered by increasing t_i . Such a couple indicates that the corresponding user was at position \mathbf{x}_i at time t_i . The velocity v_i of a user is assumed to be constant between two successive snapshots. The trajectory of a user is therefore a series of connected segments, namely a polygonal chain:

$$\mathbf{x}(t) = \mathbf{x}_i + \frac{(t - t_i)}{(t_{i+1} - t_i)}(\mathbf{x}_{i+1} - \mathbf{x}_i) \text{ for } t_i \leq t \leq t_{i+1}, \quad (2)$$

and the velocity is $v_i = d(\mathbf{x}_{i+1}, \mathbf{x}_i)/(t_{i+1} - t_i)$. A point of interest p is given by its coordinates \mathbf{x}_p on the map together with meta-information such as its name and type (*e.g.* McDonalds/Restaurant).

The first stage of the formalization consists in turning one-dimensional trajectories and positions into two-dimensional areas in such a way that an interaction can be thought of as intersecting areas over a time interval. User positions and points of interest are therefore replaced by discs of radius r and r' respectively. A linear trajectory thus becomes a rectangle (or a tube in three dimensions).

To capture the fact that businesses are often grouped, *e.g.* in malls and the fact that larger points of interest (*e.g.* stadiums) have multiple distinct close together coordinate entries in our datasets, we use a region-based segmentation approach, inspired from image processing, defining meta-points of interest for points of interest of the same type. The segmentation step is performed directly on the kd-tree. A point of interest or meta-point of interest p is represented as a disc. The radius r_p is fixed and equals to r' for points of interest while it can be larger than r' for meta-points of interest. When two (meta) points of interest are close enough, *i.e.* the intersection

between their respective discs is greater than a given threshold, they are merged into a single meta point of interest represented by the *minimum enclosing disc*, *i.e.* the disc enclosing both discs with the minimum radius (see Figure 7b). Given two (meta) points of interest at positions \mathbf{x}_{p_1} and \mathbf{x}_{p_2} with radii r_{p_1} and r_{p_2} respectively, the position and radius of the amalgamated meta-point of interest are:

$$\mathbf{x}_p = \left(\frac{d + r_{p_1} - r_{p_2}}{2d} \right) \mathbf{x}_{p_1} + \left(\frac{d + r_{p_2} - r_{p_1}}{2d} \right) \mathbf{x}_{p_2} , \quad (3)$$

and $r_p = (d + r_{p_1} + r_{p_2})/2$ where $d = d(\mathbf{x}_{p_1}, \mathbf{x}_{p_2})$ is the distance between the two points of interest. Consider the three right-most points of interest from the example depicted in Figure 7b. The radius of the two points of interest on the right have a large intersection. The corresponding points of interest are therefore merged and the discs are replaced by a larger one (dashed) tightly enclosing them. The second iteration of the segmentation algorithm merges the newly created meta-points of interest with the left-most point of interest yielding a new meta-point of interest with an even larger disc (dashed).

We define a model of interaction inspired from physics: given two successive snapshots (t_i, \mathbf{x}_i) and $(t_{i+1}, \mathbf{x}_{i+1})$ of a user and a point of interest p at position \mathbf{x}_p , the *power* of the interaction at time t is defined as a function $p()$ of the distance between the user's position $\mathbf{x}(t)$ (as defined in Eq. (2)) and the position \mathbf{x}_p of the point of interest. The *energy* E of the interaction during the time interval $[t_i, t_{i+1}]$ is obtained through integration:

$$E((t_i, \mathbf{x}_i), (t_{i+1}, \mathbf{x}_{i+1}), \mathbf{x}_p) = \int_{t_i}^{t_{i+1}} p(d(\mathbf{x}(t), \mathbf{x}_p)) dt . \quad (4)$$

While we impose the power function to depend only on the distance between the respective positions of the user and the point of interest, the choice of its expression is left open. We further constraint its support to positions where the disc around the user and the disc around the point of interest intersect: $p(d(\mathbf{x}(t), \mathbf{x}_p)) = 0$ if $d(\mathbf{x}(t), \mathbf{x}_p) > r + r_p$. This constraint allows efficient computation of the energy generated by a trajectory by pruning, using the kd-tree, all the points of interest too far from a position. In our analysis we used the area of intersection as power function:

$$\begin{aligned} p(d) = & -\frac{1}{2} \sqrt{(-d + r + r_p)(d + r + r_p)(d^2 - \delta^2)} \\ & + r^2 \arccos \left(\frac{d^2 + r^2 - r_p^2}{2dr} \right) \\ & + r_p^2 \arccos \left(\frac{d^2 + r_p^2 - r^2}{2dr_p} \right) , \end{aligned} \quad (5)$$

where $\delta = r - r_p$. If one of the two discs is included in the other one, *i.e.* $d < |r - r_p|$, the expression becomes $p(d) = \pi \min(r, r_p)^2$.

The energy received by a point of interest can be obtained by summing the energy of its interactions with all the segments of all users' trajectories. In case of an interaction with a meta point of interest, the energy of the interaction is evenly split among the points of interest composing the meta-point of interest.

5 Evaluation

In this section, by using the methodology we defined above, we evaluate user interaction with points of interest. We proceed as follows. First we take a look at user interests by profiling them based on the types of places they visit. Second, we take a look at the frequency of visits, a measure captured by the power model defined above. Third, we take a look at time of day effects aiming to answer when do users visit certain types of points of interest.

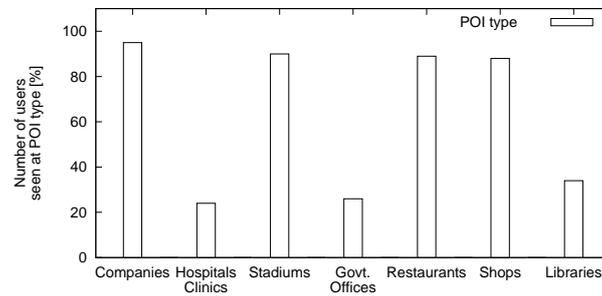


Figure 8: China – User profiling based on users visiting specific points of interest.

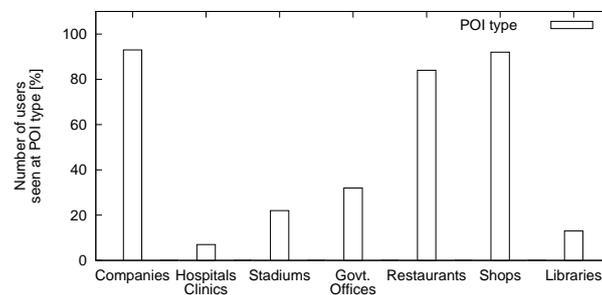


Figure 9: United States – User profiling based on users visiting specific points of interest.

5.1 User Interest in Locations

In this section we perform *user profiling* to identify broad user interests in the types of locations that they visit. To achieve this, we proceed as follows. We try to match user proximity to points of interest in a particularly conservative manner. If the user trajectory has been very close to a certain point or meta point of interest (r and r' are both chosen to be *five meters*) we say that the user has shown an interest in that specific type of point of interest. We perform this analysis by using our above methodology on the entirety of a user's trajectory.

The results of this study are presented in Figure 8 for the Chinese users and in Figure 9 for the United States users. One can see that a majority of the people in both datasets visit companies. This result is not unexpected, companies employ people and it is expected that users would visit their workplace. One can see that the users in both datasets visit restaurants and shops. An interesting fact is that libraries are not visited so much but this can be explained by the fact that people tend to read less and less in recent years [31]. The same result holds for government offices. People prefer to deal less with authorities and these days, when dealing with authorities, most of the actions can also be performed online. The results also show that most of the users in our datasets do not visit hospitals.

Further, we correlate the results obtained in this section with what we previously presented in Table 2 and Figures 5 and 6. One can notice that both hospitals, clinics and libraries are more spread than stadiums for example but we have a lot more users that visit stadiums than the other types we mentioned. Considering our previous analysis, stadiums have less points of interest and are not as uniformly located as libraries for example.

As one can see, most of the users tend to engage in regular activities such as dining in a restaurant or shopping yet they also visit stadiums even though it means travelling more.

Some other facts related to users visiting other types of points of interest (not shown in the figures) is that Chinese users tend to engage more in cultural activities than the United States users. A bigger percent of the Chinese users are seen visiting theaters and cultural centers than the United States users. Also, a small percent of the United States users (4%) are observed while engaging in religious activities (*e.g.* going to church). The United States users also seem to be

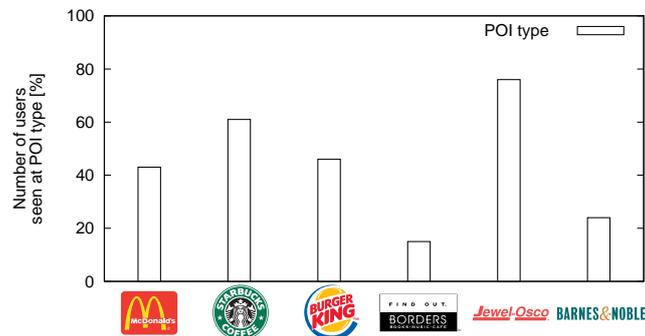


Figure 10: United States – User affiliation with specific businesses.

more engaged in car related activities. Approximately twice the number of United States users can be seen at gas stations, car dealerships, and car repair or maintenance places.

Finally, it is important to mention that when considering the radius of gyration measure that we analyzed in Section 3.1 a bigger radius of gyration corresponds to a more diverse user profile. Basically, users that have a larger radius of gyration manifest interests in more point of interest types than the users with a lower radius of gyration (result not shown).

5.2 Going Deeper: Business Case Studies

In this section we go deeper from the analysis we performed above and we analyze specific businesses and the users that visit these locations. We use the United States dataset and we select several businesses for our analysis. They are shown in Figure 10. From the restaurant category, we select the fast foods McDonalds and Burger King. We select Starbucks, Borders, Jewel-Osco, and Barnes and Noble from the shop category. Starbucks is a coffee shop. Borders and Barnes and Noble are book stores. Jewel-Osco is a popular Midwest grocery store. Note that all of the analyzed locations (except the Jewel-Osco grocery shops) carry publicly available WiFi access points that can be potential content offloading targets for mobile users.

Figure 10 contains the number of users in percents seen at the specific locations. For example, a majority of the users are seen shopping at Jewel-Osco grocery shops. Starbucks is more frequently encountered within the area than Jewel-Osco grocery shops and a large number of users are seen there as well. One can note that none of the book stores are popular in terms of the number of users that visit them (Borders recently filed for bankruptcy [8], and several analysts expect Barnes and Noble to follow suit [6]).

Furthermore, because we are dealing with actual business entities and not broad categories, one could imagine scenarios where mobile providers form partnerships with specific businesses and have their clients offload content through publicly available WiFi access points. A way to formalize this problem is through a set covering problem (a well known NP hard problem). For example each of the above businesses can jointly cover a certain amount of users from all the users. Therefore, we might desire to find the minimum number of businesses that a mobile provider can form partnerships with such that we cover all or a majority of the users. By carrying out this analysis on our data and restricting ourselves to businesses that carry WiFi access points, we find that McDonalds, Burger King, and Starbucks can jointly cover up to 94% of the mobile users.

Further analysis would be required to point out many other interesting aspects with regards to the temporal interaction between users and specific businesses. However, as this is not the main focus of our paper we leave such analysis for future work.

5.3 Frequency of Visits

In this section we answer the question related to which point of interest types are more popular. In general, people do not interact with different types of points of interest in the same way. In

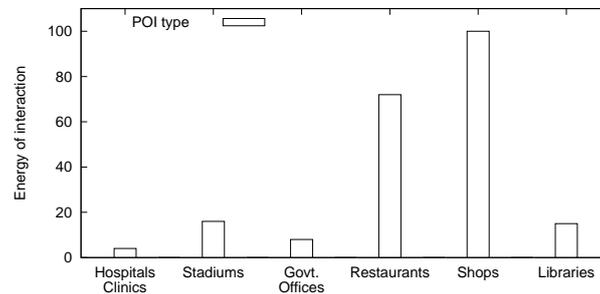


Figure 11: Energy of interaction for Chinese users interacting with the specific point of interest types.

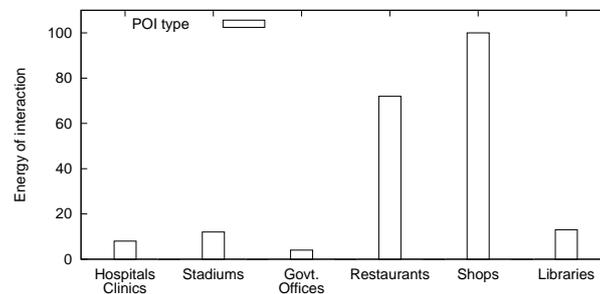


Figure 12: Energy of interaction for United States users interacting with the specific point of interest types.

particular, people visit certain places more often mainly because of functionality reasons; doing groceries, dining at a restaurant, watching a game in a stadium, *etc.* We perform this analysis by relating to the *energy* measure that we introduced in Section 4. We therefore try to capture the total amount of energy that specific points of interest manage to gather during the trace interval from users visiting them.

Figures 11 and 12 show the results. The figures are scaled to the biggest energy value that we observe for each dataset, in both cases shops. On purpose we do not show the values for companies mainly because the interactions that users experience with the companies they work at dwarf all other values.

A few insights from the figures are as follows. The most important insight is that the results are surprisingly consistent across the two datasets even though they were collected in different parts of the globe. Further, one can see that restaurants and shops lead the energy of interaction race in both figures. They hold the most energy out of all the analyzed point of interest types. This is not surprising as they are part of the activities that people carry out more frequently. Another insight is that although a lot more users have been seen at stadiums for the duration of the trace, the interactions that users carry out with stadiums are not so frequent. In fact libraries and stadiums compare in terms of energy of interaction most probably because the users that *do* visit libraries carry out such activities more frequently than the users that visit stadiums.

Next, we take a user perspective in order to capture the user activity related to points of interest. First we contrast the results shown in this section with the results shown in Figure 13. The figure shows the average distance over all United States users from the center of mass for each user until the closest point of interest of that specific type. Several categories of points of interest exhibit a direct correlation between the energy they have and the proximity to users. However this is not always the case as libraries and stadiums that have comparable or greater distances than hospitals, clinics exhibit a higher energy than the latter.

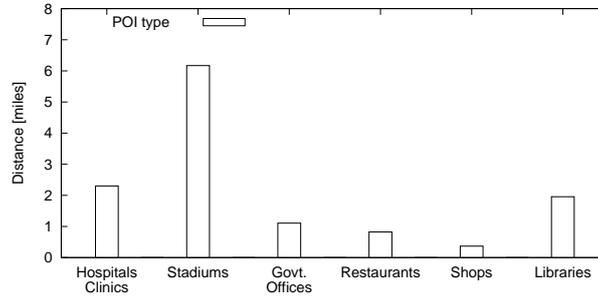


Figure 13: Average distance from the center of mass of each user until the closest point of interest of specific type for United States users.

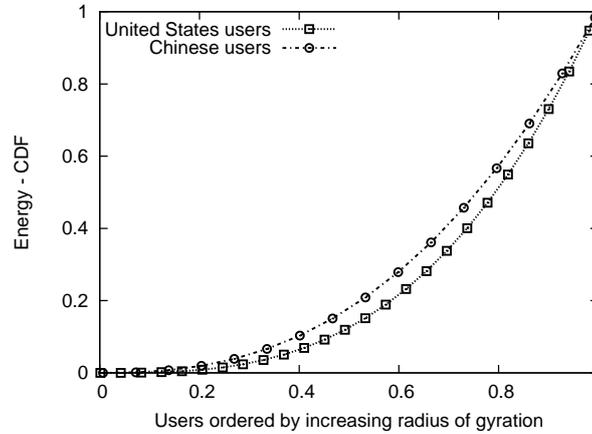


Figure 14: Energy of interaction for users ranked by increasing radius of gyration. The users that have bigger radius of gyration values have more interactions.

Here, we once again refer to the radius of gyration measure that we defined in Section 3.1. We compute the total energy that users gather by interacting with points of interest. Figure 14 shows the results. The x axis represents all users ranked by increasing radius of gyration values (computed until the end of the trace). A user farther up the x axis will always have a radius of gyration bigger or equal than a user who is lower on the x axis. The figure depicts a cdf over the total amount of energy gathered by users. Note that users who have bigger radius of gyration values interact more with the points of interest around them and hence have bigger energy values (the curve is steeper at the end). As shown, this result holds for both United States and Chinese users. This confirms our key hypothesis that people move on purpose even if it means traveling over larger distances.

5.4 Time of Day Effects

In this section we are analyzing the time of day effects that user interaction with points of interest might display. As such, we compute the interaction energy across different point of interest types for each of the hours of a day. We only show the times of the day when the energy is the highest or the second highest. In Figures 15, and 16 we present the results. For each of the point of interest types we have a bar 100% full for the time bin where the energy is the highest for the specific point of interest type shown. In cases where the second to highest energy value is bigger than 70% we show it as well. For example restaurant visits peak at 2 PM and 7 PM for the Chinese users (Figure 15) and at 7 PM for the United States users (Figure 16). A second bar for

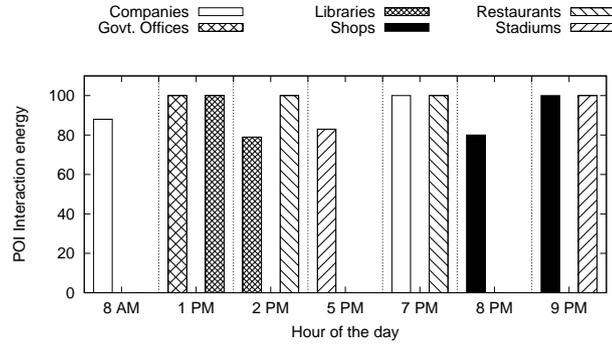


Figure 15: Hour of day effects for Chinese users interacting with points of interest.

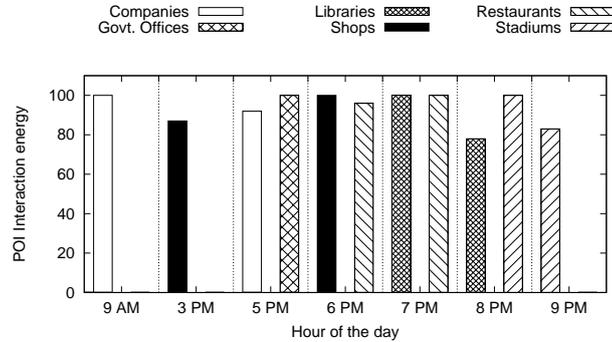


Figure 16: Hour of day effects for United States users interacting with points of interest.

restaurants stands at 6 PM and has a value of 96% of the highest energy for the United States users (Figure 16).

The results in the figures can be explained as follows. Restaurant visits peak during the early evening (after work hours) for both the Chinese and the United States users. For the Chinese users we have a second peak at 2 PM corresponding to lunch hours. The United States users also experience a peak at around noon (not shown) which is lower than the peaks we show for 6 PM and 7 PM. The difference in lunch hours might be explained by the different cultures involved.

User presence at companies peaks in the hours of the morning when people arrive at work (8 AM for the Chinese users and 9 AM for the United States users), and during the late hours of the afternoon when people leave (7 PM for the Chinese users and 5 PM for the United States users). An interesting phenomenon can be seen in Figure 15 for the Chinese users. During lunch hours (1 PM and 2 PM) some users engage in other activities such as visiting libraries or government offices. The peaks displayed for both libraries and government offices are significant. They are considerably larger than similar activity for the same points of interest in other time bins (3 times as much). This could be explained by the fact that users try to optimize their schedule and overlap some of the activities they carry out during their lunch break.

Both Chinese and United States users engage in shopping during the hours of the evening (8 PM and 9 PM for the Chinese users, and 6 PM for the United States users). Similarly, we note that sports related activities such as visiting stadiums occur during the hours of the evening again (9 PM for the United States users and 8 and 9 PM for the Chinese users).

When analyzing both figures we can note a tendency of United States users to engage in certain activities during late afternoon most probably after leaving work. This can involve dealing with the authorities, shopping, or dining at a restaurant. This is somewhat in contrast to what we see for some of the Chinese users that try to overlap some activities with their lunch break.

6 Discussion

In this section we discuss both issues related to our measurement data and implications that our research potentially has on intertwining areas.

6.1 GPS Accuracy

Studies of several GPS commercial device receivers puts the accuracy of GPS within a few meters 95% of the time [18]. Mainly because in this paper we took an extremely *conservative* approach when matching user trajectories with point of interest locations, such accuracy proves to be indeed sufficient. Also, to the best of our knowledge, this is the best level of accuracy that can be obtained by using commodity mobile phones. A better accuracy can only be obtained by using expensive military equipment or correlating with other positioning sources, both economically infeasible for large-scale studies such as ours.

6.2 Implications

In this section we point related areas that might benefit from the the research presented in this paper. As such we detail below about: (i) delay-tolerant networking and content offloading, and (ii) mobile social networking and mobile advertising.

Delay-tolerant networking and content offloading. Most research on delay tolerant networking considers random interactions between human carried mobile devices. Only recently there have been studies on delaying content until better connectivity infrastructure is encountered.

Not surprisingly, the content delay-energy savings trade off has been recognized in several other recent projects [4,20,40] that propose offloading 3G data to WiFi networks or other mobile infrastructure. Such offloading or scheduling approaches would definitely benefit from understanding human movement and particularly the relationship that users have with the businesses around them. Businesses often carry their own WiFi access points that could be useful offloading targets (*e.g.* Starbucks). Businesses could also potentially host new infrastructure for a mobile provider.

A recent body of work [1,27] proposed using both the fact that people always take the same routes (the predictability of human movement), linked together with external signs (*e.g.* the presence of certain Bluetooth devices) that could often signal better connectivity options. Our research has the potential to take this research one step further by noticing that modern phones are increasingly equipped with GPS receivers. Therefore, mobile phones could benefit when enhanced with location knowledge about points of interest, businesses that carry publicly available WiFi access points.

Mobile social networking and mobile advertising. Social networking websites have recently become very popular. Users are trying to keep in contact with each other all across the world. They try to follow the lives of their friends, the locations they visit, the things they buy, *etc.* There are different reasons for why users post information about their lives or follow their friends. Some people want to get help from their friends in their daily lives while some other people just want to ‘show off’ [25]. Although it is already an integral part of the social networking experience, human movement has started recently to fascinate social networking users. There are mobile social networking applications that are entirely built around users visiting places of interest. One can only note the popularity of social gaming systems such as Foursquare [13], and SCVNGR [32]. Some of these systems (*e.g.* SCVNGR) partner with existing retailers and offer badges or rewards such as gift cards or gaming items for doing everyday things such as buying groceries from Walmart, buying coffee from Starbucks, or having dinner at Burger King.

Such applications (and others that might emerge) involve users actively moving to get the desired result and integrated with newer technologies that allow users to use their mobile phones as credit cards for example [24,26] carry the potential of becoming the next big trend in mobile social networking [33,34].

Our study could definitely benefit all of the above. Understanding human interaction with points of interest could help mobile social networking applications better target their campaigns

when dealing with specific retailers. Mobile advertising could also be enhanced by targeting the users that manifest an interest in a specific businesses or point of interest types.

7 Related Work

Motivated by the emergence of Internet connected hand-held devices (*e.g.* smartphones), mobility models for human movement, such as Brownian Motion (*i.e.* random walks with Gaussian steps), Lévy Flights (*i.e.* random walks with possibly truncated heavy-tailed steps), Random Way Point (*i.e.* linear motion between randomly chosen points) and Fractal Way Point (*i.e.* random points forming a fractal) to name just a few, have recently received increasing attention in the networking research community [10]. This is mainly because of their applications in delay-tolerant networking (*e.g.* routing [11] and data-offloading [4,5,20,30]), mobile applications (*e.g.* location-aware data storage [35], and social gaming [13]), *etc.* These models have been inspired and/or usually validated using the so-called digital footprints obtained from popular GPS-enabled mobile devices. We use similar data in our research.

Even though some of the before mentioned models are able to capture spatial and temporal patterns of social behaviors, they do not take into account the particular semantics incorporated in human movement. While the mobility of a user may effectively be a polygonal line as in the random way point model, the destinations are not chosen (totally) at random but are determined by the semantics of each point of interest and the user interest at that specific time instant. This is precisely what we investigate in this work based on the notion of points of interest.

In [22], it is shown that human movement shows strong temporal and spatial patterns with a high probability of returning to a few locations. These locations can be of interest for a single person or a small social group (*e.g.* home or office), or common for a larger number of users (*e.g.*, restaurants). The first type of locations was studied in [21] through an analysis on the way that people name locations while the latter, effectively captured by the notion of points of interest, is at the core of our work. Our work confirms the findings from [22] and further details the semantics of these locations and the frequency at which they are visited. In [39] (our own work), we have shown that the application accessed by mobile users are correlated with their current location and their mobility patterns. In [9], a large-scale (*i.e.* the whole United States) mobility trace of banknotes is analyzed highlighting the limits of Lévy Flights to model human travel and therefore a two-parameter continuous-time random walk model is proposed instead.

Another noticeable body of work in the analysis of human mobility is the SLAW model [19] that unifies the most significant previously proposed models in order to generate realistic mobility traces for human walks. While we also acknowledge the usefulness of human mobility models we are more concerned with the lack of expressiveness of the characteristics that get incorporated into such models. As such, our work can lead to better, semantically enhanced mobility models.

Closest to our work, [42] studies the interactions between users and interesting locations (similar to the notion of point of interest). The user travel experience and location interest are jointly inferred using a new model, outperforming traditional rank-by-count and rank-by-frequency methods. The study is however limited as they only use a limited number of scenic locations mainly of interest to tourists and therefore they do not target points of interest at large.

Privacy is an increasing concern in mobile social applications that often leak personal information such as names, photos, *etc.*, as pointed out in [3]. This is particularly acute when users share their locations, be it social-driven, *i.e.* advertising the current location to friends for fun (*e.g.* micro-blogging), or purpose-driven, *i.e.* to perform an action that needs to know the current location (*e.g.* meeting someone or finding near restaurants) [36,37]. For instance, it was shown in [16] that anonymity could easily be broken in location traces by first inferring the work and home places and then using them to identify the corresponding person or at least narrow down the set of possible identities. Several efficient solutions have been proposed to enhance users privacy in location-sharing online services. Among them we can count: Path Confusion [17], Camouflage [23] that exploits intersecting paths, Mix Zones that consist of making close mobile nodes shuffle their identifiers [7,14]. This can also be combined with the usage of dedicated policies that control

when, where and to whom location information can be disclosed [38]. While we also agree that privacy related to locations and human movement is an important open problem, our work does not deal with it but is mostly aimed at understanding the actual semantics that gets incorporated into human movement.

8 Conclusions

In this paper we have proposed a new approach to analyzing the mobility of humans. Noting the lack of expressiveness incorporated in to existing studies, we proposed and performed the first-of-its-kind joint analysis of human mobility correlated with the surrounding environment, *i.e.* the points of interest that they visit. Our key finding is in demonstrating user affinity (with different visiting frequency depending on point of interest type) towards specific points of interest and specific businesses. We believe our results show noteworthy promise for further research in this area, clearing the way for future advances in understanding basic human behavior and impacting problems related to mobile transfer scheduling algorithms, mobile infrastructure placement, mobile social networking, mobile advertising, *etc.*

Summary. From the user perspective, our insights are the following: *(i)* Users with a more diverse user profile and a larger number of interactions with points of interest exhibit a larger movement span characterized by a larger radius of gyration. *(ii)* There is strong evidence related to users optimizing their time by for example overlapping certain activities with their breaks. *(iii)* While users in different parts of the globe exhibit different time of day behavior, we find that there is a striking correlation in the frequency of visits to the point of interest types that we have analyzed.

From the point of interest perspective, our insights are the following: *(i)* User presence at points of interest displays a strong time of day effect, certain points of interest being more popular in the afternoon while others are more popular in the evening. *(ii)* Point of interest proximity to users, together with functional aspects (the point of interest type) are the two factors that influence popularity.

Acknowledgements

Kévin Huguenin was partially funded by a scholarship offered by the University of Rennes I and an “explorateur” grant offered by INRIA.

References

- [1] G. Ananthanarayanan and I. Stoica. Blue-Fi:Enhancing Wi-Fi Performance using Bluetooth Signals. In *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services (MobiSys'09)*, 2009.
- [2] ANN: A Library for Approximate Nearest Neighbor Searching. <http://www.cs.umd.edu/~mount/ANN/>.
- [3] K. Balachander and W. Craig. Privacy Leakage in Mobile Online Social Networks. In *Proceedings of the 3rd Workshop on Online Social Networks (WOSN '10)*, 2010.
- [4] A. Balasubramanian, R. Mahajan, and A. Venkataramani. Augmenting Mobile 3G Using WiFi. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys'10)*, 2010.
- [5] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani. Energy Consumption in Mobile Phones: A Measurement Study and Implications for Network Applications. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (IMC'09)*, 2009.

-
- [6] Borders' Bankruptcy Shakes Industry. http://www.nytimes.com/2011/02/17/business/media/17borders.html?_r=1&partner=yahoofinance.
- [7] A. Beresford and F. Stajano. Mix Zones: User Privacy in Location-Aware Services. In *Proceedings of the 2nd International Conference on Pervasive Computing (Pervasive'04)*, pages 127–131, 2004.
- [8] Borders Files for Bankruptcy, to Close 200 Stores. http://www.msnbc.msn.com/id/41536256/ns/business-consumer_news/t/borders-files-bankruptcy-close-stores/.
- [9] D. Brockmann, L. Hufnagel, and T. Geisel. The Scaling Laws of Human Travel. *Nature*, 439(7075):462–465, Jan. 2006.
- [10] T. Camp, J. Boleng, and V. Davies. A Survey of Mobility Models for Ad Hoc Network Research. *Wireless Communications and Mobile Computing*, 2(5):483–502, Aug. 2002.
- [11] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE Transactions on Mobile Computing*, 6:606–620, June 2007.
- [12] FLANN: Fast Library for Approximate Nearest Neighbors. <http://www.cs.ubc.ca/~mariusm/index.php/FLANN/FLANN>.
- [13] Foursquare. <http://foursquare.com/>.
- [14] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. C. Parkes. On Non-Cooperative Location Privacy: A Game-Theoretic Analysis. In *Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS'09)*, pages 324–337, 2009.
- [15] GeoLife GPS Trajectories. <http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/>.
- [16] P. Golle and K. Partridge. On the Anonymity of Home/Work Location Pairs. In *Proceedings of the 7th International Conference on Pervasive Computing (Pervasive'09)*, pages 390–397, 2009.
- [17] B. Hoh and M. Gruteser. Protecting Location Privacy Through Path Confusion. In *Proceedings of the 1st International Conference on Security and Privacy for Emerging Areas in Communication Networks (SECURECOMM'05)*, pages 194–205, 2005.
- [18] S. Kindra, W. Thomas, and W. Keith. Comparing GPS Receivers: A Field Study. *URISA Journal*, 18:19–23, Feb. 2006.
- [19] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. SLAW: A Mobility Model for Human Walks. In *Proceedings of the 29th Annual IEEE International Conference on Computer Communications (INFOCOM'09)*, 2009.
- [20] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong. Mobile Data Offloading: How Much Can WiFi Deliver? In *Proceedings of the 6th ACM International Conference on emerging Networking Experiments and Technologies (CoNEXT'10)*, 2010.
- [21] J. Lin, G. Xiang, J. I. Hong, and N. Sadeh. Modeling People's Place Naming Preferences in Location Sharing. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp'10)*, 2010.
- [22] M. Gonzalez and C. Hidalgo and A. Barabasi. Understanding Individual Human Mobility Patterns. *Nature*, 453(7196):779–782, June 2008.

- [23] J. Meyerowitz and R. Roy Choudhury. Hiding Stars with Fireworks: Location Privacy through Camouflage. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking (MobiCom'09)*, pages 345–356, 2009.
- [24] Mobile Phone Payments. http://money.cnn.com/galleries/2011/technology/1101/gallery.mobile_payments/index.html.
- [25] Mobile Social Networking Set for Growth. <http://www.emarketer.com/Article.aspx?R=1006514>.
- [26] Apple Plans to Link Smart Phones to Credit Cards. <http://www.newsweek.com/2011/01/28/apple-plans-to-link-smart-phones-to-credit-cards.html>.
- [27] A. J. Nicholson and B. D. Noble. Breadcrumbs: Forecasting Mobile Connectivity. In *Proceedings of the 14th Annual International Conference on Mobile Computing and Networking (MobiCom'08)*, 2008.
- [28] POI data. <http://poi.gps-data-team.com/>.
- [29] POI and Map Data. <http://downloads.cloudmade.com/>.
- [30] M.-R. Ra, J. Paek, A. B. Sharma, R. Govindan, M. H. Krieger, and M. J. Neely. Energy-Delay Tradeoffs in Smartphone Applications. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (MobiSys'10)*, 2010.
- [31] Study: Americans Reading Less Than They Used To. <http://www.npr.org/templates/story/story.php?storyId=16739654>.
- [32] SCVNGR. <http://www.scvngr.com/>.
- [33] SCVNGR Goes Global, Unleashes Zombie Horde on Businesses. <http://personalmoneystore.com/moneyblog/2010/11/02/scvngr-global-zombies-google-places/>.
- [34] SCVNGR Unleashes Zombie Horde Through Social Check-In Feature. <http://www.wired.com/magazine/2010/10/scvngr-unleashes-zombie-horde-through-social-check-in-feature/>.
- [35] P. Stuedi, I. Mohomed, and D. Terry. WhereStore: Location-Based Data Storage for Mobile Devices Interacting with the Cloud. In *Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond (MCS'10)*, 2010.
- [36] K. P. Tang, J. Lin, J. I. Hong, D. P. Siewiorek, and N. Sadeh. Rethinking Location Sharing: Exploring the Implications of Social-Driven vs. Purpose-Driven Location Sharing. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp'10)*, 2010.
- [37] E. Toch, J. Cranshaw, P. H. Drielsma, J. Y. Tsai, P. G. Kelley, J. Springfield, L. Cranor, J. Hong, and N. Sadeh. Empirical Models of Privacy in Location Sharing. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp'10)*, 2010.
- [38] E. Toch, J. Cranshaw, P. Hankes-Drielsma, J. Springfield, P. G. Kelley, L. Cranor, J. Hong, and N. Sadeh. Locaccino: A Privacy-Centric Location Sharing Application. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing (UbiComp'10)*, 2010.
- [39] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci. Measuring Serendipity: Connecting People, Locations and Interests in a Mobile 3G Network. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (IMC'09)*, 2009.

- [40] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci. Taming User-Generated Content in Mobile Networks via Drop Zones. In *Proceedings of the 31st Annual IEEE International Conference on Computer Communications (INFOCOM'11)*, 2011.
- [41] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding Mobility Based on GPS Data. In *Proceedings of the 10th ACM International Conference on Ubiquitous Computing (UbiComp'08)*, 2008.
- [42] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining Interesting Locations and Travel Sequences from GPS Trajectories. In *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*, 2009.



Centre de recherche INRIA Rennes – Bretagne Atlantique
IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399