# Towards a Boosted Route Planner Using Individual Mobility Models

Riccardo Guidotti[1,2](✉) and Paolo Cintia[1,2]

[1] KDDLab, University of Pisa, Largo B. Pontecorvo, 3, Pisa, Italy
{riccardo.guidotti,paolo.cintia}@di.unipi.it
[2] KDDLab, ISTI-CNR, Via G. Moruzzi, 1, Pisa, Italy
{riccardo.guidotti,paolo.cintia}@isti.cnr.it

**Abstract.** Route planners generally return routes that minimize either the distance covered or the time traveled. However, these routes are rarely considered by people who move in a certain area systematically. Indeed, due to their expertise, they very often prefer different solutions. In this paper we provide an analytic model to study the deviations of the systematic movements from the paths proposed by a route planner. As proxy of human mobility we use real GPS traces and we analyze a set of users which act in Pisa and Florence province. By using appropriate mobility data mining techniques, we extract the GPS systematic movements and we transform them into sequences of road segments. Finally, we calculate the shortest and fastest path from the origin to the destination of each systematic movement and we compare them with the routes mapped on the road network. Our results show that about 30–35 % of the systematic movements follow the shortest paths, while the others follow routes which are on average 7 km longer. In addition, we divided the area object of study in cells and we analyzed the deviations in the flows of systematic movements. We found that, these deviations are not only driven by individual mobility behaviors but are a signal of an existing common sense that could be exploited by a route planner.

## 1 Introduction

Route planners are systems which help users selecting a route between two locations. When providing directions, web and mobile mapping services generally suggest the shortest route. Popular route planning system such as Google Maps, Open Street Maps etc. generate diverging directions using powerful libraries of roads and road attributes [16]. However, they often ignore both the time at which a route is to be traveled and, more important, the preferences of the users they serve. Since cities are becoming crowded and jammed, smart route planning are gathering an increasing interest. In such a context, a route planner which takes into account users' preferences [8], and which exploits the crowd expertise w.r.t urban mobility in order to identify the best route, can be more desirable and helpful than an ordinary route planner [6].

A route planner which exploits individual mobility models to improve the planning will have a real advantage from these models only if the users do not


D. Bianculli et al. (Eds.): SEFM 2015 Workshops, LNCS 9509, pp. 108–123, 2015.
DOI: 10.1007/978-3-662-49224-6_10

follow the shortest path in their systematic movements but deviate from them. Consequently, the target of this work is twofold. The first one is to understand and estimate how much the systematic movements of a user are different from the shortest paths between the origin and destination locations. The intuition is that a user which lives and acts in a certain territory do not automatically select the shortest path. This can happen for many reasons: e.g. traffic conditions, road quality, for passing close to the cheapest petrol station, for avoiding roads with control of speed etc. However, independently from the reasons, if there is a divergence between the systematic route with origin point $o$ and destination point $d$, and the shortest route from $o$ to $d$ suggested by a route planner, then also other users could benefit from this kind of knowledge which comes from individual expertise on a certain area. This lead to the second and main target: a boosted route planner that, when is possible, proposes as alternative to the shortest path a route which is frequently followed by someone. This planner would be a route planner coming from the wisdom of the crowd in mobility.

By exploiting *individual mobility profile* models [15] and *trajectory map-matching* [5] for a set of users in Pisa and Florence province, we retrieved the systematical movements of the users, named *routines*, and we mapped these routines along a road network. By calculating the shortest path from the origin $o$ to the destination $d$ of each routine with an ordinary route planner we obtained the movements a user would have followed when there is not expertise of the area. Then we compared the routines with the corresponding shortest path. Thanks to this analysis we are able to *(i)* quantify how much human mobility differs from the shortest path and, on the other hand how good can be an approximation of human mobility made with the shortest paths, *(ii)* at which level appears the divergence between the routine and the shortest path w.r.t. origin/destination, and *(iii)* which are the road intersections, areas and flows of movements in which users mobility detaches more in comparison with the shortest paths.

Our experiments show that about 30–35% of the routines follow the shortest paths, while the others follow routes which are on average 7 km longer. In addition, 20 % of the routines deviate at the very beginning from the suggested paths. Despite these differences, 60 % of the route returned by the planner would belong to the individual mobility profiles. Consequently, even if the analyzed drivers follow routines quite similar to the routes suggested by a route planner, they deviate from them not to minimize the travel distance but for some other unknown reasons. Finally, we discovered a sort of collective "common sense": when moving from a certain origin to a certain destination nearly all the drivers deviate in the same area. This indicates that different users which systematically drive along the same roads develop similar individual mobility behaviors.

Preliminary techniques are illustrated in Sect. 2. In Sect. 3 we propose our analytic model. We show in Sect. 4 the results of our analysis. In Sect. 5 we summarize some related works on route planning. Finally, Sect. 6 reports a summary of the contributions of the paper and possible future works.

## 2 Preliminaries

Movements are usually performed by people in specific areas and time instants. These people are called *users* or *drivers* and each movement is composed by a sequence of spatio-temporal points $(x, y, t)$ where $x$ and $y$ are the coordinates, while $t$ is the time stamp. We call *trajectory* the movements of a user described by a sequence of spatio-temporal points:

**Definition 1 (Trajectory).** *A* trajectory *$m$ is a sequence of spatio-temporal points $m = [(x_1, y_1, t_1), \ldots, (x_n, y_n, t_n)]$ where the spatial points $(x_i, y_i)$ are sorted by increasing time $t_i$, i.e., $\forall 1 \leq i \leq k$ we have $t_i < t_{i+1}$*

The set of all the trajectories traveled by a user $u$ makes her *individual history*:

**Definition 2 (Individual History).** *Given a user $u$, we define the* individual history *of $u$ as the set of traveled trajectories denoted by $H_u = \{m_1, \ldots, m_k\}$.*

### 2.1 Individual Mobility Profiles

It is possible to extract the systematic movements of a user $u$ by following the profiling procedure proposed in [15]. This approach groups the trajectories using a clustering algorithm equipped with a *distance function* defining the concept of trajectory similarity:

**Definition 3 (Trajectory Similarity).** *Given two trajectories $m'$ and $m''$, a trajectory distance function* dist *and a distance threshold $\varepsilon$, we say that $m'$ is* similar *to $m''$ ($m' \sim m''$) iff $\mathrm{dist}(m', m'') \leq \varepsilon$.*

The result is a partitioning of the original dataset from which the *clusters* with few trajectories and those containing noise are filtered out. Finally, the *representative trajectory* are extracted from the remaining clusters. These representative trajectories are called *routines* and the set of routines is called *mobility profile*:

**Definition 4 (Routine and Mobility Profile).** *Let $H_u$ the individual history of a user $u$, $ms$ a minimum size threshold,* dist *a distance function and $\varepsilon$ a distance threshold. Given a grouping function $\mathcal{M} = group(H_u, ms, \varepsilon, \mathrm{dist})$, such that $\mathcal{M} = \{M_1 \ldots M_k\}$ where $M_i \subset H_u$, we define a* routine *$r_i$ as the medoid trajectory of a group $M_i$. The set of routines extracted from $\mathcal{M}$ is called* mobility profile *and is denoted by $P_u = \{r_1 \ldots r_k\}$.*

A *mobility profile* describes an abstraction in space and time of the systematic movements: the user's real movements are represented by a set of trajectories delineating the generic paths followed. Moreover, the exceptional movements are ignored due to the fact they will not be part of the profile. Figure 1 depicts an example of mobility profile extraction. We name $getmedoids(\mathcal{M})$ the function that takes in input the output of $group()$ and returns the routines, i.e. the medoid trajectories $\{r_1 \ldots r_k\}$ of the groups in $\mathcal{M}$ describing the mobility profile $P_u$.

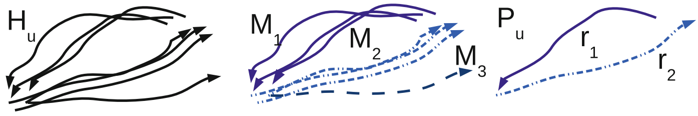


**Fig. 1.** The user *individual history* ($H_u$), the clusters identified by the grouping function ($M_1, M_2, M_3$) and the extracted *individual routines* ($P_u = \{r_1, r_2\}$) forming the *individual mobility profile*.

### 2.2 Trajectory Map Matching

A trajectory $m$ coming from GPS or GSM dataset generally does not contain the relative raversed road network segments. Such enrichment might not be straightforward, especially when raw trajectory data have a high sampling rate. This lack of information can be restored by means of some *map matching* techniques. We adopted the *gravity model* [5] as method to match each single trajectory point to the road segment it belongs to:

**Definition 5 (Gravity Force Attraction).** *Given a point $p_i$ and a set of road segments describing the road network $S = \{s_1, \dots, s_r\}$ where $s_j = \{p_{start}, p_{end}\}$, we define the* gravity force attraction *of a segment $s_j$ for a point $p_i$ as:*

$$GFA(p_i, s_j) = w_{(p_i,s_j)} = w^d_{(p_i,s_j)} \cdot w^\theta_{(p_i,s_j)}$$

*where* $w^d_{(p_i,s_j)} = 1 - \frac{\mathrm{dist}(p_i,s_j)}{\sum\limits_{s_k \in S} \mathrm{dist}(p_i,s_k)}$, $w^\theta_{(p_i,s_j)} = 1 - \frac{\mathrm{ang}(p_i,r_j)}{\sum\limits_{s_k \in S} \mathrm{ang}(p_i,s_k)}$, dist *is the euclidean distance between a point and a segment, and* ang *is the absolute difference between the direction of the point and the direction of the segment.*

This model can be applied over the whole road network segments. However, in real applications the set of segments $S$ to be considered can be very large. For this reason, it is possible to use a *nearest neighbor* approach and consider only a subset $S_k \subset S$ containing the $k$ segments closest to a given point.

Given a GPS trajectory $m = \{p_1, \dots, p_n\}$ and a set of road segments $S$, it is possible to assign each point $p_i$ to the segment with the most powerful force $\bar{s_j} = \sigma(p_i, S, k) = argmax_{s_j \in S_k}(GFA(p_i, s_j))$. The Gravity Model adopted has also been used to estimate the traveltime of each matched road segment; once every trajectory point have been matched, the typical travel time of a segment $s$, given $P$ the set of points matched to $s$, is defined as $\frac{\sum p_i \in P speed(p_i) * GFA(p_i,s)}{\sum p_i \in P GFA(p_i,s)}$. Road network travel times have been estimated from a dataset composed by 9.8 millions car travel.

**Definition 6 (Trajectory Map Matching).** *Given a trajectory $m$ and a set of road segments $S$, we refer to $m^*$ as the trajectory $m$ on the road segment network $S$, i.e. the points of $m^*$ belong to the segments in $S$:*

$$m^* = mapmatch(m, S, k)$$

*where* $m^* = [p^*_1, \dots, p^*_n] = [\bar{s_1}, \dots, \bar{s_{n-1}}]$ *and* $[p^*_i, p^*_{i+1}] = \bar{s_j} = \sigma(p_i, S, k)$

Thus, $m$ can be transformed in the map matched version $m^* = [p_1^*, \ldots, p_n^*]$ containing points which belong to the road segments $S$, where $p_1^*, \ldots, p_n^*$ maximize the attractions with $p_1, \ldots, p_n$, i.e. $m^*$ is the best representation of $m$ on $S$. Note that it is sufficient to set $k$ greater than one to guarantee the denominators be different to zero. A refinement is needed to obtain the map matched trajectory $m^*$, i.e. a path must be added for each couple of points which are not directly connected. This is a common case when low sampled GPS data are involved: in this scenario, a GPS point every ∼90 s is recorded. To find such path followed by the driver, we used a Time-Aware heuristic as described in [4]. This map-matching method takes the GPS travel time between the two consecutive GPS point as input and returns the path connecting the two points that better fit the input travel time. It is worth to consider that the road network is a directed graph, thus including and correctly recognizing one way segments.

## 3 Proposed Analytic Model

In the following we describe the analytic model adopted to discover how much the shortest/fastest path can approximate the systematic movements of a user, and how much a route planner could improve its performances by using the wisdom of systematic drivers.

Given a set of users $U$ and set of road segments $S$, for each user $u \in U$, we calculate the individual mobility profile

$$P_u = getmedoids(group(H_u, ms, \varepsilon, \mathrm{dist}))$$

Then for each routine $r_i \in P_u$, we map match the routine on the road network

$$r_i^* = mapmatch(r_i, S, k)$$

We name *map matched individual mobility profile* $P_u^* = \{r_1^*, \ldots, r_k^*\}$ the profile of a user $u$ containing the routines mapped on the road network.

We define a route planner

$$\bar{m} = routeplanner_{type}(o, d, S)$$

as a function which returns the best path $\bar{m} = [o, \bar{p}_2, \ldots, \bar{p}_{n-1}, d]$ w.r.t. the type of search $type \in \{s, f\}$ (where $s$ stands for *shortest* and $f$ stands for *fastest*) on the road segments $S$ where $o$ is the origin point and $d$ is the destination point.

Finally, for each routine $r_i^* = [o_i, \ldots, d_i] \in P_u^*$ we calculate the path returned by the route planner $\bar{r}_i = routeplanner_{type}(o_i, d_i, S)$ on the origin and destination. We indicate with $\bar{P}_u^{type} = \{\bar{r}_1, \ldots, \bar{r}_k\}$ the *shortest/fastest individual mobility profile* of a user $u$ containing the paths returned by the route planner.

Summing up, given a set of users $U$ and their individual history $H_u \forall\, u \in U$, and the road network segments set $S$ we obtain:

1. $P_u \;\forall\; u \in U$ with the *Mobility Profiles* step as result of the application of *group*() and *getmedoids*() using $H_u$ for each $u \in U$;

2. $P_u^* \ \forall\ u \in U$ through the *Map Matching* step as result of the application of $mapmatch()$ for each $r_i \in P_u$, $\forall\ u \in U$;
3. $\bar{P}_u^{type} \ \forall\ u \in U$ by means of the *Route Planner* step as result of the application of $routeplanner()$ on the origin and destination points $o_i, d_i$ of for each $r_i^* \in P_u^*$ , $\forall\ u \in U$.

Figure 2 shows the steps of the analytic mobility model. In the next section we will observe the differences between $P_u^*$ and $\bar{P}_u^s$, $\bar{P}_u^f$. We remark that the *shortest path* is the path which minimizes the distance, while the *fastest path* is the path which minimizes the travel time.

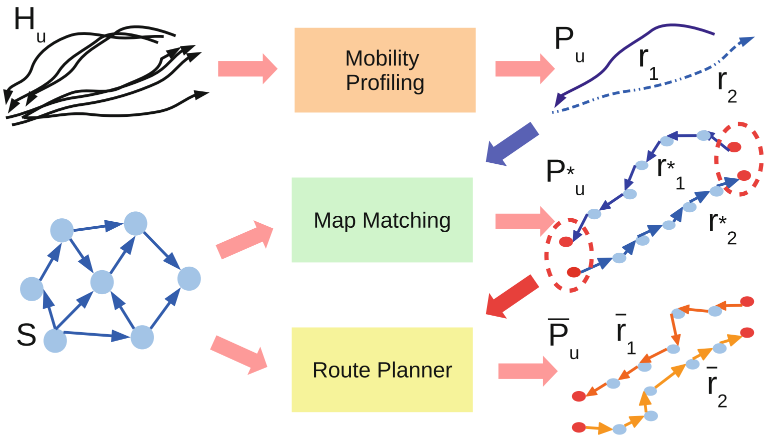


**Fig. 2.** Steps of the analytic mobility model. Input: individual history $H_u$, road network segments set $S$. Output: individual map matched mobility profile $P_u^*$, individual shortest/fastest mobility profile $\bar{P}_u^{type}$. $P_u$ is calculated by using the *Mobility Profiling* functions. Then, the *Map Matching* module produces $P_u^*$ by using the routines in $P_u$. Finally, $\bar{P}_u^{type}$ is obtained by using the *Route Planner* on the origin and destination points (highlighted in the red dotted circles) of the routines in $P_u^*$.

## 4 Experiments

In the following we evaluate how much systematic users described by their map matched individual mobility profile $P_u^*$ deviate from the shortest and fastest routes contained in the shortest mobility profile $P_u^s$ and fastest mobility profile $P_u^f$ for the provinces of Pisa and Florence. Moreover we analyze which are the nodes on the road network $S$, the areas and the flows more affected by deviations.

### 4.1 Dataset

As a proxy of human mobility, we use real GPS traces collected for insurance purposes by *Octo Telematics S.p.A*[1]. This dataset contains 9.8 million car travels

[1] http://www.octotelematics.com/it.

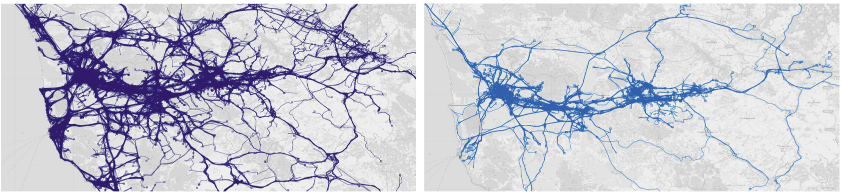

**Fig. 3.** (Left) A sample of the considered trajectories in Pisa province. (Right) Mobility profiles extracted in Pisa province.

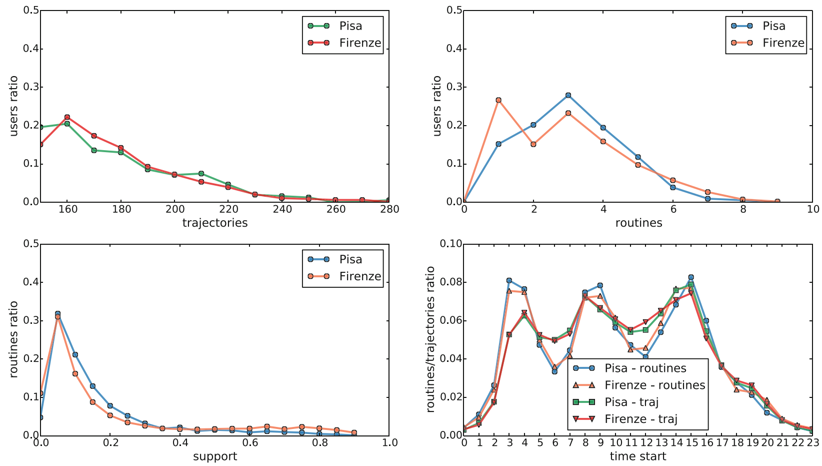


**Fig. 4.** Distributions of number of trajectories (top - left), number of routines (top - right), routine relative support (bottom - left), trajectories and routines starting time (bottom - right).

performed by about $160,000$ vehicles active in a geographical area focused on Tuscany (Italy) in a period from 1st May to 31st May 2011. Figure 3-left depicts a sample of the considered trajectories. In our analysis we split geographically the dataset in provinces to consider the fact that each area has its type of mobility with characteristics depending on the surface, on the topology and on the number of inhabitants. In this paper we present the results obtained for the provinces of Pisa and Florence. A user is analyzed in one province if at least one of his/her trajectories passes through that province. In particular we analyzed a subset of $3,000$ representative users which have traveled along a total of about 500.000 trajectories. The individual history $H_u$ represents our input data.

### 4.2 Mobility Profiles Analysis

To perform the *Mobility Profiling* step, we used as profiling function $profile()$ the clustering algorithm Optics [1], and as distance function $dist()$ a function which compares the points distances along the trajectories (or an interpolation of

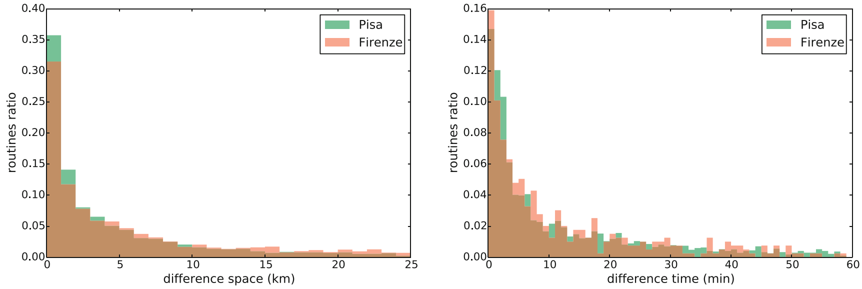


**Fig. 5.** (Left) Space difference distribution in km between the routines in $P_u^*$ and the corresponding routines in $\bar{P}_u^s$. (Right) Time difference distribution in minutes between the routines in $P_u^*$ and the corresponding routines in $\bar{P}_u^f$.

them) and returns the average of these comparisons. In order to obtain sound and reliable routines we performed some preliminary tests to set the best parameters to extract the mobility profiles $P_u$. We choose $\varepsilon = 500$ m and $ms = 8$ since a routine is a movement that must be repeated a significant number of time during a month. Figure 3-right depicts an example of profile extracted in Pisa province modeling the users' systematic movements.

In Fig. 4(top) we can observe the distributions of the number of trajectories and number of routines per user (left and right respectively). All the users selected have more than 150 trajectories and most of them has 160 with an average of about 200 trajectories. Most of the individual mobility profiles $P_u$ contain $1 - 4$ routines. The average length of a routine is about 8.87 km ($\pm$ 8.96 km of standard deviation), while the average duration is about 20 min ($\pm$ 12 min standard deviation). In Fig. 4(bottom - left) we can observe that most of the routines have a relative support of 0.2 of the trajectories. This means for example that given a user with 160 trajectories and a routine with support equals to 0.2, then that routine is supported by about 30 trajectories, i.e. a trajectory per day on average in the observation period. Finally, the starting time distributions of trajectories and routines is depicted in Fig. 4(bottom - right). Note how the starting time distribution of the routines, more than the starting time distribution of the trajectories, follows a clear M-shape pattern. This highlight how the routines capture the systematic movements from home to work in the morning and from work to home in the afternoon.

### 4.3 Deviation Analysis

As first experiment we analyzed the deviation in term of space difference from the routines in $P_u^*$ to those in shortest path $\bar{P}_u^s$, and the deviation in term of time difference from the routines in $P_u^*$ to those in fastest path $\bar{P}_u^f$. In particular, for each user $u \in U$ analyzed, for each routine in $r_i^* = \{o_i, \dots d_i\} \in P_u^*$, we calculated the difference in length with the corresponding route in $\bar{P}_u^{\{s,f\}}$, i.e. the route $\bar{r}_i$ which starts in $o_i$ and ends in $d_i$. Note that the following results

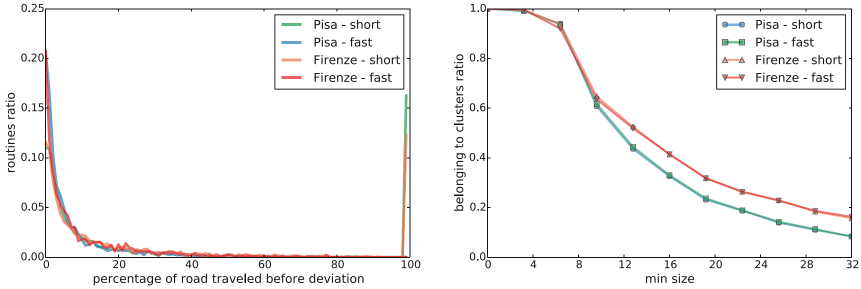


**Fig. 6.** (Left) Distribution of the percentage of road traveled before the routine deviates from the shortest/fastest path. (Right) Ratio of shortest and fastest routes belonging to the clusters of the corresponding routines by varying the *minsize* parameter.

**Table 1.** median, average and standard deviation of the space difference (km), time difference (min) and relative percentage of road traveled before the deviation (*pbd*).

| | Short - space diff | | | Fast - time diff | | | Short - pbd | | | Fast - pbd | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | med | avg | std | med | avg | std | med | avg | std | med | avg | std |
| Pisa | 02.31 | 07.16 | 13.56 | 07.42 | 26.92 | 58.13 | 07.07 | 25.14 | 35.52 | 07.96 | 23.19 | 32.33 |
| Florence | 03.64 | 10.22 | 18.45 | 07.31 | 19.06 | 29.90 | 02.97 | 07.58 | 13.54 | 01.05 | 01.58 | 21.58 |

are biased by the route planner used: by applying different route planners the shortest and fastest path obtained could be different.

In Fig. 5 we can observe the space and time differences distributions. With respect to the shortest path (left in the figure), in both dataset there is a consistent set of routines with space difference equals to zero. This indicates that 30–35 % of the routines (for Pisa and Florence respectively) follow the shortest path suggested by the route planner. The remaining routines differentiate on average of 7 km (see Table 1). On the other hand, in Fig. 5(right) none of the routines follows exactly the fastest path. Just few routines, i.e. the 10 %, follow the fastest routes with less than a minute of difference. All the others differentiate consistently (20 min on average Table 1). In addition, we observed that 15 % of the drivers in Pisa and 10 % of the drivers in Florence have the individual mobility profile exactly equal to the shortest mobility profile ($P^*_u = \bar{P}^s_u$). On the contrary, none of the user has all the routines equal to the fastest path, i.e. $P^*_u = \bar{P}^f_u$

In Fig. 6 left is reported the percentage of road traveled before the deviation (*pbd*), both for Pisa and Florence. It is obtained by observing after how much $r^*_i$ deviates from $\bar{r}_i$ after the start point $o_i$ (for $\bar{r}_i \in \bar{P}^s_u$ and $\bar{r}_i \in \bar{P}^f_u$). We can notice how 20 % of the systematic movements deviate from the shortest/fastest paths at the very beginning. The distribution is a long tailed power law with average percentage before deviation of 7 % and 3 % for Pisa and Florence respectively (see Table 1). Furthermore, how already observed, there is a consistent subset of routines (12–15%) which do not deviate from the shortest path. This does not occur for the fastest path.

Finally, we studied the percentage of shortest/fastest movements which would have belonged to the clusters by varying the *minsize* (*ms*) parameter (Fig. 6 right). We calculated for each user $u \in U$ the trajectory distance (using the same distance function *dist* applied for the clustering) between the short/fast paths $\bar{r}_1 \dots \bar{r}_k$ and the trajectories belonging to the corresponding cluster $M_1 \dots M_k$. For $minsize = 8$ (the value used for the clustering), 60 % of the movements returned by the route planner would have belonged to the clusters in both shortest and fastest path. This indicates that the movements returned by the route planner are similar enough to the trajectories belonging to the cluster to be considered part of them. This fact is quite interesting if we consider that the space and time difference between routines and suggested routes are in some cases not negligible, and that the routines generally deviate not far from the origin point.

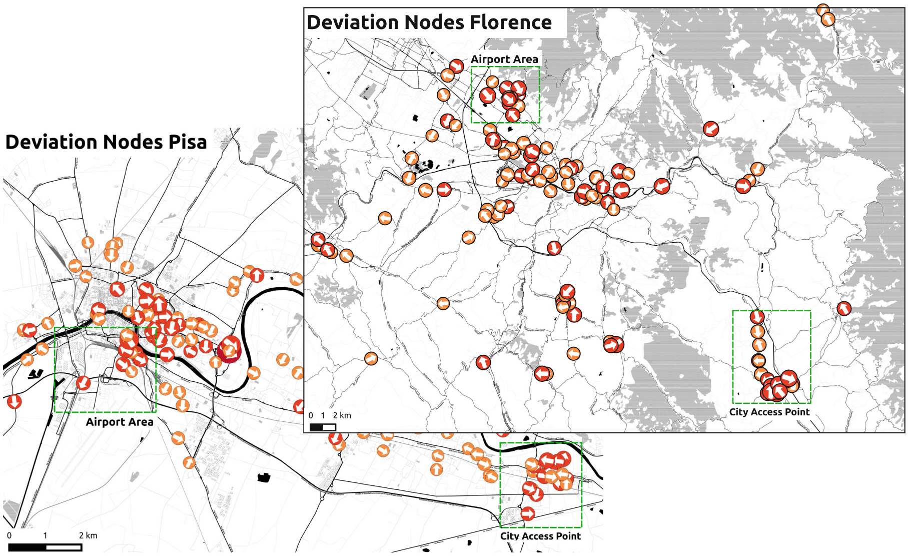


**Fig. 7.** Deviation nodes supported by with at least 100 deviations.

The conclusion is that systematic drivers generally deviate from the routes suggested by a route planner at the very beginning of their movements, and that in general they do not optimize their travel time but try to minimize the travel distance. However, even the drivers deviate from the short/fast routes, these routes are in many cases very similar to the routines systematically followed.

### 4.4 Towards a Boosted Route Planner

Before presenting the analysis of this section we remark that routines are movements repeated many times (on average 15 times) during the observation period. Thus, if drivers systematically deviate from what is supposed to be the shortest (or the fastest) path there should be a valid reason. Given a user moving for the first time in a certain area, it could be better for him/her to follow the routines described by "expert driver" instead of the routes suggested by a route planner.

A route planner could be boosted by exploiting the knowledge given by the individual mobility models. Such a route planner should consider various information: *(i)* the road intersections where the systematic drivers deviate more, *(ii)* the areas where those intersections are concentrated, and *(iii)* the main flows of movement containing deviations. In the following we analyze these three factors to understand their impact and which are their possible uses. Due to lack of space in the following we focus the analysis only on the deviation of the routines against the shortest path.

We refer to the road intersections as *deviation nodes*. They correspond to the first nodes in the set of road segments $S$ from which the routines in $P_u^*$ deviate from the route in $\bar{P}_u^s$. To count the number of deviations, instead of considering only the number of routines, we weighted each routine $r_i^* \in P_u^*$ with the number of trajectories that support it. In Fig. 7 we can observe the deviation nodes in which there are at least 100 trajectories which deviate. The darker and the bigger is a marker, the higher is the number of deviations performed by the routines on that node. As expected, for both cities, the highest numbers of deviation nodes appear into the city center. This confirms the fact that in the city is very difficult to follow the shortest paths. Moreover, in both cities we can observe some particular areas not in the city center (those highlighted in the green dotted squares) with an high number of deviations. They correspond in both cases *(i)* to the main access points to/from the city center, and *(ii)* to the roads close to the airports. This is a signal that these areas are probably affected by consistent traffic and the systematic users which have to pass through them prefer longer but less stressful routes.

To analyze the deviations' areas we divided the territory using a grid with cells of 2.5 km of radius. The heatmap of the deviations is shown in Fig. 8. The darker is a cell, the higher is the number of trajectories which support the routines deviating there. For these images no filters are applied. The first insight is that the users acting in province of Florence have an active role even in the mobility of Pisa but the viceversa is not true. Indeed, most of the cells with more deviation in Pisa occur also in the Florence heatmap. From the intersection of the two images emerges that most of the systematic deviations take place along the main road between Pisa and Florence (named SGC Fi-Pi-Li) with a concentration in the area around Empoli. This probably happens because most of the people living in Empoli, which is in province of Florence, go systematically to Pisa for working. For example, instead of following SGC Fi-Pi-Li that is an highway but has a lot of traffic, many drivers could prefer as alternative the road SS67 which runs along SGC Fi-Pi-Li but has much more turns and is not an highway. In Fig. 9(left) we report the distribution of the number of cells per routines' deviations. It is a power low distribution indicating that there are few cells where most of the systematic users decide to take alternative routes. Those are the cells that more than the others the boosted root planner should consider when suggesting the routes which exploit the wisdom of the crowd.

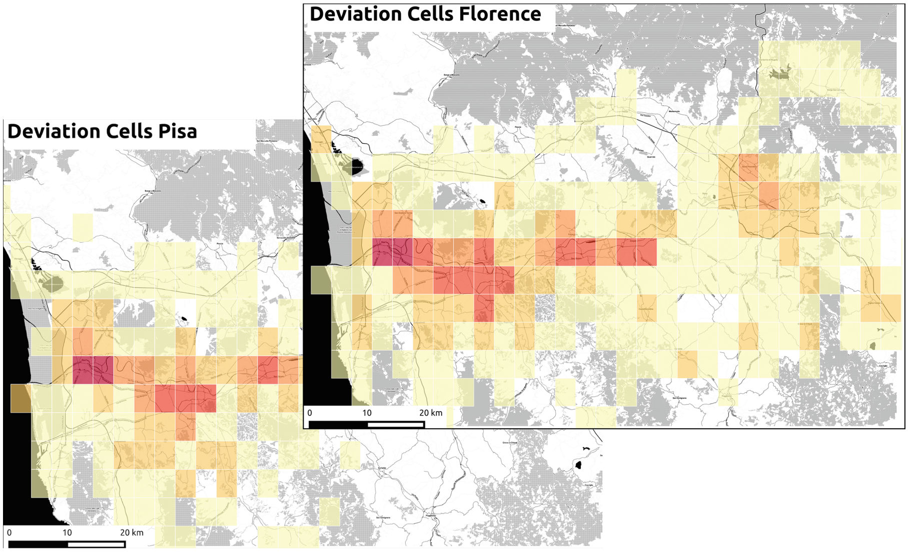


**Fig. 8.** Heatmap of the deviation cells.

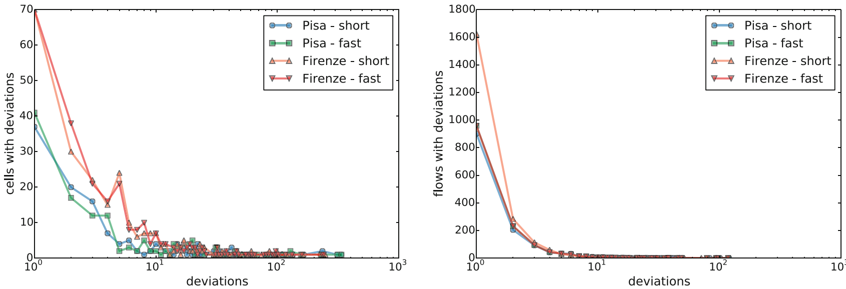


**Fig. 9.** Distributions of the number of cells with deviations (left), and of the number of flows with deviation (right).

We defined a flow as a triple of cells ($origin$, $deviation$, $destination$) where $origin$ is the cell origin of the routine, $deviation$ is the cell where $r_i^*$ deviates from $\bar{r}_i$, and $destination$ is the ending cell of the routine. In Fig. 10 we can observe the flows containing the routines supported by at least 100 trajectories. Through this approach we can observe the main flows along with most of the drivers deviate from the shortest paths. We can observe how in Pisa province there are various flows of entrance to and exit from the city center. The flow with more deviations (the purple biggest arrows) are just under the city center starting from the airport area up to the suburbs. They are surrounded by a large number of in-coming and out-coming flows. We remark that in many cases the deviation from the shortest path appears at the very beginning of the movement. Thus the flows reported mainly highlight the part of the movement after the deviation.

Some deviation flows do not have a mutual reverse flow of the same importance. For these cases the deviation is more evident only in one direction. On the other hand, in province of Florence, the flows in the city center are on average shorter than those outside. In addition, the biggest flows are present in the airport area (big green arrow in the center) and close to the exit of the highways (big blue arrow bottom right and big aqua green arrow in the center). Figure 9(right) shows the distribution of the number of flows per routines' deviations. Similarly to the cells, the distribution is long tailed indicating a small set of flows where many routines deviates from the shortest/fastest path. A route planner having this kind of knowledge should recommend paths which run along these flows and are similar to the individual routines. Indeed, by applying appropriate weights on the road network segments in $S$ the route planner could provide solutions boosted by the routes systematically followed by expert drivers.

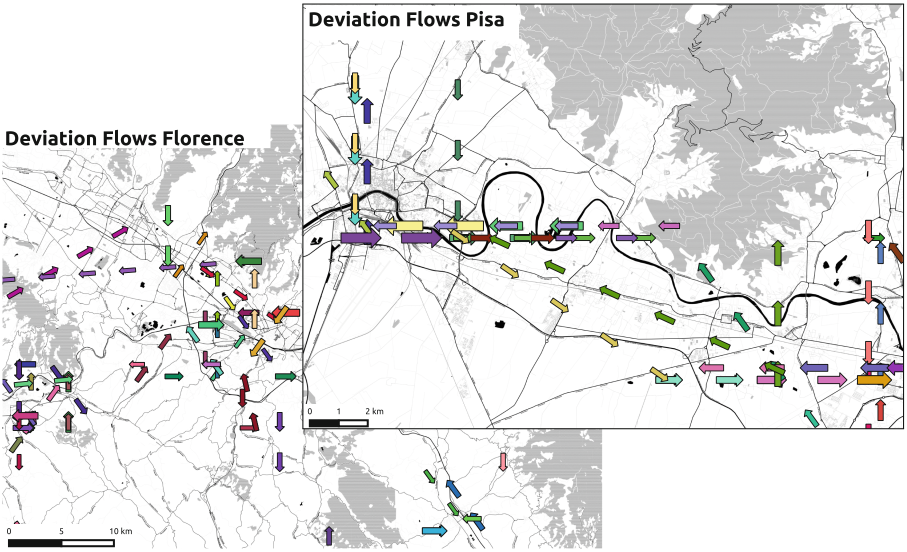


**Fig. 10.** Deviation flows supported by with at least 100 deviations.

Finally, we analyzed the difference between the flows described above and the flows built using only origins and destinations. In other words given a origin-destination flow (*origin*, *destination*) how many flows (*origin*, *deviation*, *destination*) pass through the same *deviation*? We name this indicator *flow similarity in deviation*. This value give us a hint of how much a certain deviation is stable along a flow. A flow similarity in deviation of X% indicates the percentage of (*origin*, *deviation*, *destination*) flow on the number of origin-destination flows (*origin*, *destination*) which pass through the same *deviation* cell. E.g. given the following origin-destination flows $\{A \rightarrow B, X \rightarrow Y\}$ and the flows $\{A \rightarrow C \rightarrow B, A \rightarrow C \rightarrow B, A \rightarrow D \rightarrow B, X \rightarrow Z \rightarrow Y, X \rightarrow Z \rightarrow Y\}$, then the percentage of flow difference is 80 %. In our dataset of Pisa and Florence we obtained the following results: *Pisa*: 83 % (short), 78 % (fast), *Florence*: 87 % (short), 85 % (fast). These high percentages are a clear signal that the deviation along the various flows are not a matter of individuals, but that are known and subscribed from the majority of the drivers. It is a sort of “common sense” which surprisingly emerges at collective level even though all the mobility models used in the proposed analysis are individual.

## 5 Related Work

Route planners are designed to provide information about the possible journeys in a certain area. Generally route planners refer to means of transportation which are either private or public. However, the application prompts a user to input an origin and a destination and it recommends some routes which are considered to be the best for that query.

Route planners generally use some smart variations of well known shortest path algorithms to search a graph of nodes (modeling access points to the network) and edges (modeling links between nodes) [8]. Different cost weights such as distance, cost etc. can be associated with edges and nodes. However, it is generally quite difficult to plan high quality routes [11]: *(i)* the notion of “route quality” is different from person to person, and *(ii)* available route networks rarely contain all the information needed for proposing the best route (e.g. traffic information, road quality etc.). Thus, even though the search can be optimized w.r.t. different criteria, e.g. the shortest, the fastest, the cheapest [13] and even the happiest ones [14], there is not guarantee that the route provided will be considered “the best” by the majority of the users.

Various effort in different directions have been made to improve route planning applications. In particular, personalized route services able to deal with individual users preferences have been investigated recently. For example in [12] complex users preferences were modeled into a route planner by means of the fuzzy set theory. In [9] the authors provided improved individual route plans for Dublin inhabitants by exploiting both historical data and estimated traffic flows. Still according to an estimation of future travels obtained by mining public transport data, in [7] were recommended personalized tickets for London public transport network. Another framework for personalized trip recommendations considering user preferences and temporal properties was proposed in [10]. In [16] were introduced real-time information coming from GPS-equipped taxi together with historical data for an improved route planner which uses traffic conditions and driver behavior for selecting the best path. Finally, a multi-modal journey planner can consider at the same time various means of transport and minimize the uncertainty of catching a certain means [3], or it can provide for the same journey personalized public and private transportation solutions [2].

## 6 Conclusion

In this work we analyzed the deviation of the systematic movements from the shortest and fastest paths suggested by a route planer on a set of drivers in Pisa and Florence provinces. We found that systematic drivers deviate from the routes suggested by a route planner at the very beginning of their movements, and that they generally try to minimize the travel distance more than the travel time. Moreover, we observed that the shortest paths are in many cases very similar to the systematic movements from which they deviate. Through our analytic model we were able to select the areas and the flows with the highest number of systematic deviation. We discovered that given a flow from an origin $o$ to a destination $d$ nearly all the users which systematically move from $o$ to $d$ deviate in the same area. Our analysis shows that, for some unknown reasons, the traveled systematic movements give to the drivers a feeling that their route is better than the shortest or fastest paths suggested by a route planner. This kind of knowledge can be exploited by a route planner which can weight the cost on the edges with the number of supported trajectories instead of with the length or with travel time. Following this approach, a user which travels for the first time in a certain area could be helped in selecting the route by the wisdom of the drivers which systematically pass there. Also a city manager could gain worth information from our analysis. Indeed, he/she could favor the cars circulation along the routes followed by systematic drivers and improve the others which are in fact not exploited enough.

**Acknowledgements.** This work has been partially supported by the European Commission under the SMARTCITIES Project n. FP7-ICT-609042, PETRA. We thank Roberto Trasarti and Mirco Nanni for the help that lead to the creation of this paper.

## References

1. Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J.: Optics: ordering points to identify the clustering structure. In: ACM SIGMOD, vol. 28. ACM (1999)
2. Botea, A., Braghin, S., Lopes, N., Guidotti, R., Pratesi, F.: Managing travels with petra: the rome use case. In: ICDE. IEEE (2015)
3. Botea, A., Nikolova, E., Berlingerio, M.: Multi-modal journey planning in the presence of uncertainty. In: ICAPS (2013)
4. Cintia, P., Nanni, M.: An effective time-aware map matching process for low sampling GPS data. Technical report cnr.isti/2015-TR-011
5. Cintia, P., Trasarti, R., Cruz, L., Costa, C., de Macedo, J.A.F.: A gravity model for speed estimation over road network. In: 2013 IEEE 14th International Conference on Mobile Data Management (MDM), vol. 2, pp. 136–141. IEEE (2013)
6. Giannotti, F., Nanni, M., Pedreschi, D., Pinelli, F., Renso, C., Rinzivillo, S., Trasarti, R.: Unveiling the complexity of human mobility by querying and mining massive trajectory data. VLDB J. **20**(5), 695–719 (2011)
7. Lathia, N., Capra, L.: Mining mobility data to minimise travellers' spending on public transport. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1181–1189. ACM (2011)

8. Letchner, J., Krumm, J., Horvitz, E.: Trip router with individualized preferences (trip): incorporating personalization into route planning. In: Proceedings of the National Conference on Artificial Intelligence, vol. 21, p. 1795. AAAI Press/MIT Press, Menlo Park/Cambridge/London (1999, 2006)
9. Liebig, T., Piatkowski, N., Bockermann, C., Morik, K.: Predictive trip planning-smart routing in smart cities. In: EDBT/ICDT Workshops, pp. 331–338 (2014)
10. Lu, E.H.-C., Chen, C.-Y., Tseng, V.S.: Personalized trip recommendation with multiple constraints by mining user check-in behaviors. In: ICAGIS, pp. 209–218. ACM (2012)
11. McGinty, L., Smyth, B.: TURAS: a personalised route planning system. In: Mizoguchi, R., Slaney, J.K. (eds.) PRICAI 2000. LNCS, vol. 1886, p. 791. Springer, Heidelberg (2000)
12. Mokhtari, A., Pivert, O., Hadjali, A.: Integrating complex user preferences into a route planner: a fuzzy-set-based approach (2009)
13. Pelletier, M.-P., Trépanier, M., Morency, C.: Smart card data use in public transit: a literature review. Transp. Res. Part C Emerg. Technol. **19**(4), 557–568 (2011)
14. Quercia, D., Schifanella, R., Aiello, L.M.: The shortest path to happiness: recommending beautiful, quiet, and happy routes in the city. In: Conference on Hypertext and Social Media, pp. 116–125. ACM (2014)
15. Trasarti, R., Pinelli, F., Nanni, M., Giannotti, F.: Mining mobility user profiles for car pooling. In: KDD. ACM (2011)
16. Yuan, J., Zheng, Y., Xie, X., Sun, G.: Driving with knowledge from the physical world. In: KDD, pp. 316–324. ACM (2011)