# Understanding individual routing behavior

Antonio Lima,[1] Rade Stanojevic,[2] Dina Papagiannaki,[2] Pablo Rodriguez,[2] and Marta C. Gonzàlez[1]

[1] *Civil and Environmental Engineering Department, Massachusetts Institute of Technology*
[2] *Telefonica Research, Barcelona*

Knowing how individuals move between their places is fundamental to advance our understanding of human mobility[1], improve our urban infrastructure [2] and drive the development of transportation systems. Current route choice models that are used in transportation planning are based on the widely accepted assumption that people follow the minimum cost path [3], despite little empirical support. Fine-grained location traces collected by smart devices give us today an unprecedented opportunity to learn how citizens organize their travel plans into a set of routes, and how similar behavior patterns emerge among distinct individual choices. Here we study 92,419 anonymized GPS trajectories describing the movement of personal cars over an 18-month period. We group user trips by origin-destination and we find that most drivers use a small number of routes for their routine journeys, and tend to have a preferred route for frequent trips. In contrast with the cost minimization assumption, we also find that a significant fraction of drivers' routes are not optimal. We present a spatial probability distribution that bounds the route selection space within a ellipse, having the origin and the destination as focal points, characterized by high-eccentricity independent of the scale. While individual routing choices are not captured by path optimization, their spatial bounds are similar, even for trips performed by distinct individuals and at various scales. These basic discoveries can inform realistic route choice models that are not based on optimization, having an impact on several applications, such as infrastructure planning, routing recommendation systems and new mobility solutions.

## I. INTRODUCTION

The high urban population density [4] poses new critical challenges in designing the cities of the future. Among those, traffic congestion is one of the most pressing issues. Under increasing mobility demand, the intricate task of improving existing infrastructure to allow swift mobility in the city requires special efforts. Technology can be used to collect data about humans interacting with their built environment. Converting unstructured data into knowledge requires specialized methods that extract meaningful information about individual preferences.

In the previous decade we have learned valuable aspects of human mobility, mainly from large scale data mined from mobile phone networks. Individuals' visit patterns are highly predictable, presenting unique and slow exploration habits [1, 5–9]. Mobile phone traces still remain too coarse, both in space and in time, and are unsuitable to investigate details of human choices. On the other hand, the rapidly increasing popularity of devices equipped with location sensors offers unprecedented possibilities to study individual mobility at an finer-grained level. This new lens enriches our understanding of human behavior, and allows us to examine each movement in detail and to better comprehend routing decisions, at the root of vehicular traffic.

Route choice modeling is the process of estimating the number of vehicles using a link in the road network through a model and it is a fundamental step of transportation forecasting [2]. Given some knowledge of travel demand, models associate individuals to the path they are likely to follow during their journey. Urban travel demand has traditionally been estimated by up-scaling travel diary surveys [10] and, more recently, through analysis of mobile phone data [11–15]. Route assignment techniques are based on the widely accepted assumption that individuals choose the route that minimizes a cost, usually distance, travel time and/or fuel consumption. The true utilization of a road link is assumed to be similar to that obtained under deterministic user equilibrium (UE), or Wardrop's equilibrium [3]. In stochastic user equilibrium (SUE) [16, 17] a random component is added in the expected travel times, in order to introduce heterogeneity in the routes and to represent travelers' preferences unknown to the modeler. The sets of feasible routes are either obtained by the two methods described above: deterministic shortest path and stochastic shortest path. More recently, probabilistic approaches and constrained enumeration algorithms have also been used to this purpose. In probabilistic approaches [18] a network link is chosen depending on its distance from the shortest path, according to a generalized cost function. Enumeration methods [19, 20] rely on the assumption that travelers choose routes according to behavioral rules other than the minimum cost path. However, empirical results have shown that users choose multiple routes over origin-destination pairs, reporting that most choices deviate significantly from the shortest time path [20–23]. Detours can happen for several reasons, like picking up or dropping off a passenger, having a short break at a favourite place or avoiding unpleasant areas (because of high traffic, crime, aesthetic reasons, etc.). While investigating the reason behind each detour might be a daunting task, we would first instead try to quantify how often these detours occur and how large they are. The hypothesis that we would like to check is that, regardless of the reason behind the detour, a clear physical limit dictates whether a possible deviation that is being taken

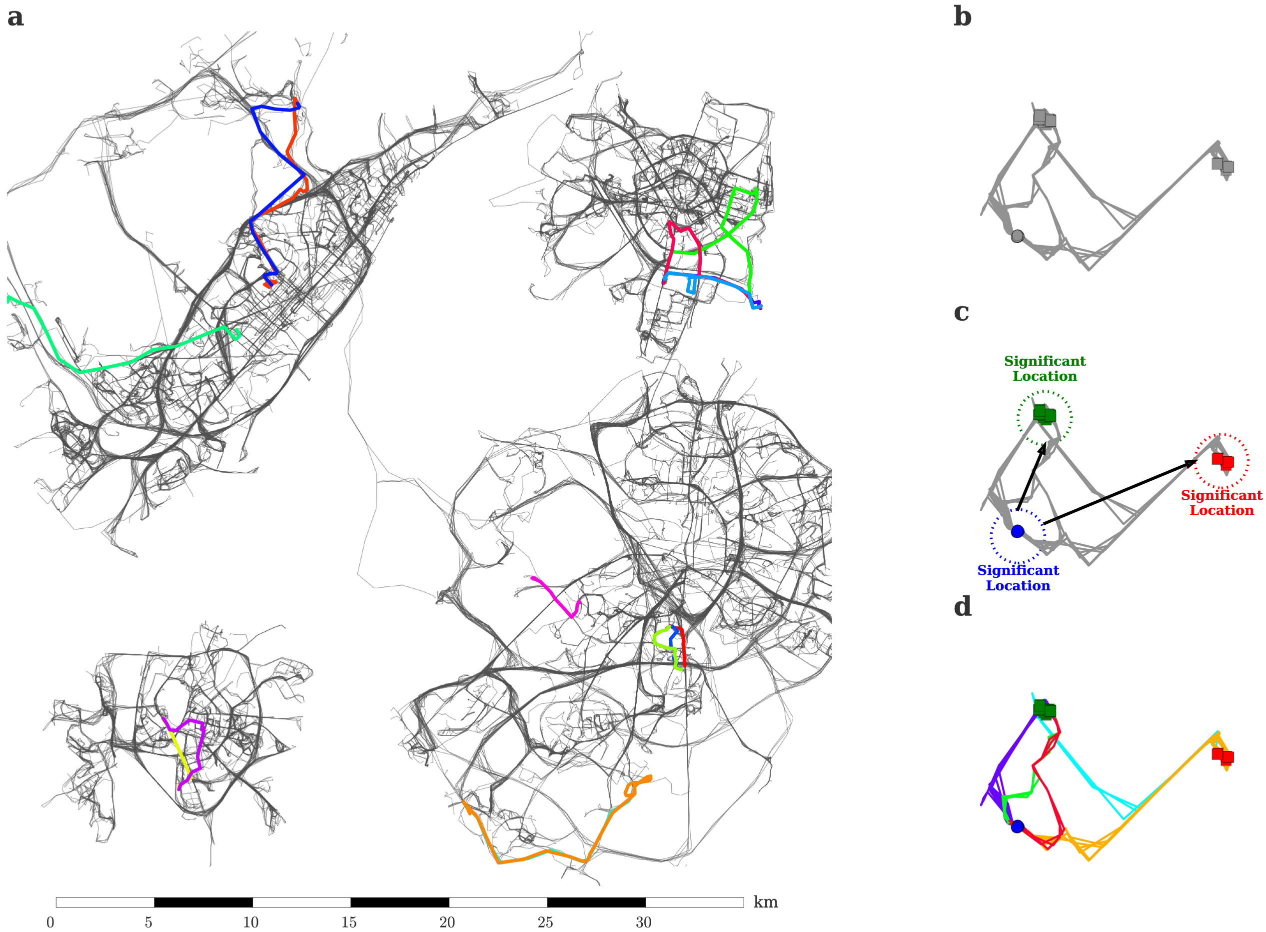


FIG. 1: **From trajectories to route choices.** (a) A sample of the trajectories analyzed from the four cities, shown in gray, outline their urban road networks. Colored trajectories spanning between the same pair of points represent seven routine trips. In each routine trip, a colored line represents a distinct route choice. (b) A set of trajectories belonging to a car. Each trajectory starts at the circle marker and ends at a square marker. (c) By clustering the endpoints of the trips we find three significant places. Two routine trips are shown with a solid black arrow. (d) We finally discover, for each routine trip, the different route choices performed by the driver. In this example one routine trip has three route choices (purple, green, red), the other has two (cyan, orange).

in consideration by the driver will be ultimately taken or ignored by the driver. In other words, how to quantify these deviations in a set of universal rules in order to be able to synthesize them? Doing that would allow us to inform probabilistic and enumeration route choice models. Namely, given the daily routine of individuals in different cities with long term observations, how can we generate heterogeneous yet feasible rules related to alternative route choices?

To that end, we use GPS traces generated by 526 private cars over an 18-month period, and explore how their routing behavior unfolds in four cities. We investigate how many routes they use and how often they use each of them. We also consider whether these routes are shortest paths and evaluate how far they typically go. We finally give a spatial characterization of routing behavior. These findings can be used in existing or new models of route choice.

## II. RESULTS

Firstly, we convert the unstructured sequences of time-referenced positions coming from GPS devices into a meaningful set of locations, trips and route choices [35, 36], as shown in Fig. 1 a. We describe a *trajectory* as a finite sequence of $(t, x)$ tuples, where $t$ represents a time value and $x$ a location vector. The source and the destination of the trajectory are the first and last point of the sequence, respectively. We call *significant place* a geographic region which a person goes to several times. Significant places detected in this study have a diameter smaller than 500 m, which is compatible to choosing parking spots in the proximity of a destination (see Supplementary Material). Several trips performed by a user between the same pair of significant places define together a *routine trip*. Finally, depending on how spatially similar are the trajectories of a routine trip, they can be grouped in one or more *route choices* (see an illustration in Fig. 1 b-d and more details in *Materials and Methods*).

The first question we would like to answer is: how many different routes do drivers use in their routine trips? In Fig. 2 a we plot the histogram of the number of routes used for each routine trip. The histograms are surprisingly similar among diverse cities. Independently of the urban settlement under consideration, different individuals prefer to use a limited number of routes, and a third of them use only one route. This is a noticeable result, considering that these trips span over an 18-month period. We can safely conclude that users organize their routine trips only through a few preferred route options, where the number of choices follows a log-normal distribution with parameters $\mu = 0.71$ and $\sigma = 2.22$. The log-normal distribution, linked to a multiplicative random process, is ubiquitous in social science [37] and has been found also in the distribution of single-mode distance trips [38]. In this case, it may arise from the set of unknown random variables that determine individual route choices.

Next, for routine trips that have used more than one route, are some of them chosen more often than others? In order to answer this question, we use a normalized Gini coefficient $G_n$, corrected to have meaningful values when the number of routes is small (see *Materials and Methods*). A value close to 0 (maximum equality) suggests that routes are evenly used. A value close to 1 (maximum inequality), suggests that the user is strongly biased towards one route for that routine trip, and that the alternate ways have been used seldom. In Fig. 2 d we plot, for routine trips that have at least two route choices, the normalized Gini, computed on the number of times that the route has been used. In general, routine trips have high values of the Gini coefficient with a median value of 0.6, suggesting that people tend to have a dominant route (Fig. 2 d). Moreover, a mild correlation between the Gini values and the number of trips made suggests an adaptation process: when an individual repeats a journey more than 20 times, a preferred route tends to dominate their route choices. By contrast, we found both the number of routes and $G_n$ to be uncorrelated with the most common time and day of the week of the routine trip.

Finally, what are the characteristics of a dominant route? Previous research assume that drivers prefer routes minimizing some cost function, directly connected to travel time, fuel consumption or distance. We compare the routes taken by the user with the routes suggested by a popular online routing service. The service provides up to three alternative routes, accounting for expected travel times and traffic conditions. In order to compare these recommended optimal routes to the routes actually chosen, we measure the maximum distance between a user's GPS positions and the recommended path (see Materials and Methods for further information). In Fig. 2 b we show the distribution of these distances, in four cases: when comparing only the top optimal route to the most used route; when comparing the optimal route to all a user's routes; when comparing the three suggested optimal routes to the dominant user route; and, finally, when comparing all suggested routes to all a user's routes. In the last three cases only the pairs of routes that deviate the least are considered. In about 53% of the cases, the dominant route chosen by the user is not the first optimal choice. For about 34% of the user routines none of the routes are compatible with the optimal choices, indicating that preferred routes do not minimize the travel cost. A previous study at a smaller scale had also found similar results that reject the shortest-path assumption [39].

Next, our goal is to determine how far away individuals are willing to go while undertaking their trip. To that end, we study the probability density function $\Phi(x, y)$ of the route locations, normalized with respect to the source and the destination. We transform trajectories to a common reference frame of coordinates for all trajectories. The goal is to see how paths unfold and how far they usually go from their endpoints, regardless of their geographic position and of the trip length. In Fig. 3 b we see that most of the deviations are small with respect to the source-destination endpoints. In particular, we find that the majority of the positions recorded are contained within an area of elliptic shape, having as the two foci the first and last point of the trip (Fig. 3 c). This result suggests that, while individuals commonly take detours due to personal preferences or characteristics of the street network [15, 40], these detours are well bounded. The emergence of an elliptical shape is not surprising. Keeping in mind that an ellipse is the loci of the points $P$ such that the sum of the distances to the the two focal points $F_1$, $F_2$ is constant ($d(F_1, P) + d(F_2, P) = a$), this result shows the detour that people are willing to take is bounded. Trips that require larger detours are rare, as they are unlikely to be undertaken, or they might be split into two distinct trips.

In order to further investigate this hypothesis and formally quantify the detours, we calculate two quantities for each trip: the geodesic distance between source and destination $f$; and $a$, the maximum value of the sum of the distance to the source and to the destination from any points along the path taken by the user. Finding these values is equivalent to identifying an idealized ellipse that fully contains all the paths taken by the driver. The eccentricity of the ellipse $e = f/a$ indicates how far from the geodesic this path goes. In the unlikely case where the endpoints lie on the same straight street and the driver takes the shortest route, $f = a$, the eccentricity takes the maximal value of 1, and the ellipse degenerates into a straight line. At the other extreme, a value of eccentricity close to 0 indicates that the path taken is very far from the endpoints, the ellipse tends to look like a circle in the target space and the two endpoints are close to each other compared to the path taken by the driver while moving between them.

Generally the straight route is not a viable option, because of physical obstacles. Drivers deviate from that

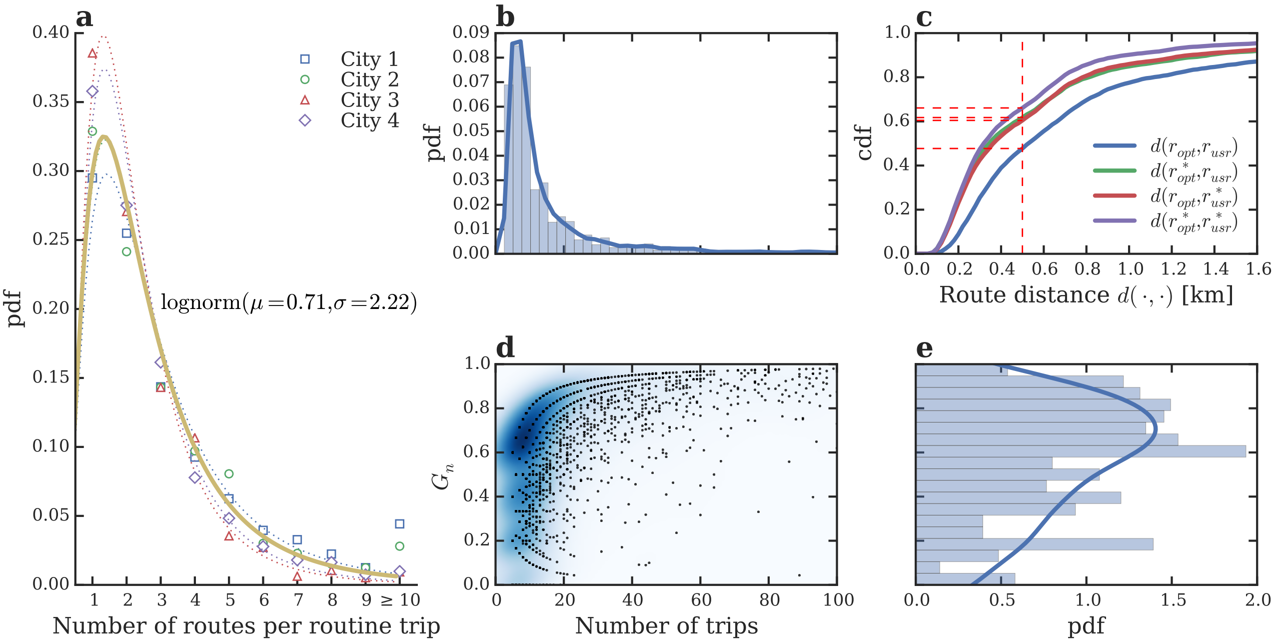


FIG. 2: **Individual patterns of route choice.** (a) The distribution of the number of routes used for a routine trip. For most routine trips this number is low, despite the fact that these trips span over a period of up to 18 months. The markers show the empirical histograms about routine trips grouped by city. The solid curve shows the best lognormal fit, obtained on aggregated data generated in all four cities. (b) The probability density distribution of number of trips performed in a routine trip; the solid line is a kernel density estimation. (c) Maximum point distance between the optimal route $r_{opt}$, as suggested by the on-line routing service, and the favorite user route $r_{usr}$. For the other three curves we consider all the alternative routes returned by the service and all the routes ever used by the driver, choosing for each element the route that deviates the least from its counterpart, respectively $r^*_{opt}$ and $r^*_{usr}$. Noticeably, 34% of the routes chosen are not any of the shortest paths and over 53% of the preferred routes are not optimal. (d) The number of trips performed during a routine journey versus the normalized Gini coefficient related to how many times each route choice is used. The two quantities show a weak correlation (Pearson $r = 0.48$, $p = 4.2e - 255$). The more a driver travels between two locations, the more likely it is for them to have a route of preference. (e) The probability density distribution of the normalized Gini coefficient $G_n$; the solid line shows a kernel density estimation.

idealized shortest path according to the underlying street network and personal routing preferences. While these two phenomena are hard to treat, we find that routing detours are well approximated by an ellipse with high values of eccentricity (Fig. 3 d). Large deviations are rare; we speculate they are caused by intermediate destinations that the driver intends to reach before the final destination (e.g. giving a ride to somebody and dropping them off). Interestingly, the value of the eccentricity does not change considerably with distance between the endpoints (Fig. 3 e), suggesting that, in an urban setting, the space of the routing alternatives is proportional to the effective distance traveled. Whether this result also holds for trips at longer distances, such as inter-city journeys, is to be investigated in future analyses.

## III. DISCUSSION

We have discovered a set of behavioral rules that capture individual behavior in an urban environment. They are independent of the urban layout and were obtained by methods that are agnostic of the underlying street network. The rules establish the basic ingredients of realistic route-choice models. Once a travel plan is established for a user, a dominant route must be assigned. This choice should be spatially bounded within an elliptic shape of high eccentricity, as observed in the experimental distribution, opportunely scaled so that the origin and the destination are the foci of the ellipse. Although the choice can be driven by a distance/cost function from the main axis of the ellipse, it does not have to be deterministically chosen as the path that minimizes a travel cost, as we have seen this does not typically reflect personal routing choices. Finally, individuals could choose alternate routes, within the ellipse, with probability inversely proportional to how often the person travels between the endpoints.

A new science of cities is emerging [44], heavily fueled by the massive data generated by numerous sensors, inherently interdisciplinary, motivated by the need to improve people's lives and counteract the negative effects of the increasing urban population (such as traffic congestion and pollution, to name the most urgent). The find-

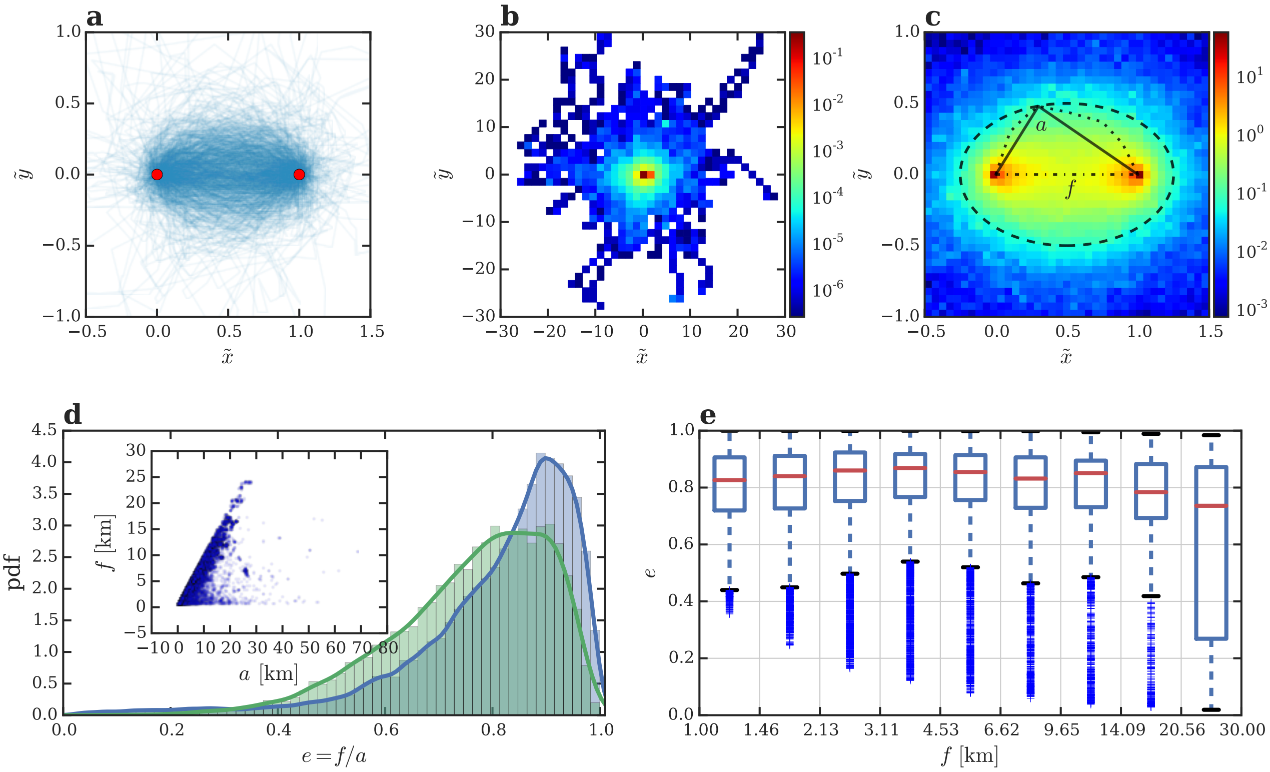


FIG. 3: **The boundaries of human routes.** Coordinates are projected to a cartesian coordinate system using the spatial reference system EPSG:2062. Each trajectory $(x, y)$ is roto-translated and scaled into $(\tilde{x}, \tilde{y})$ so that the source and destination of each trip are $(0, 0)$ and $(1, 0)$, respectively. (a) A 1% random sample of the trips, shown as partially transparent lines connecting consecutive $(\tilde{x}, \tilde{y})$ positions. (b) The probability density function $\Phi(\tilde{x}, \tilde{y})$ of the trajectory positions during a journey. Significant detours in all directions are uncommon but not unheard of. (c) 95% of the positions are within the region shown in the figure. The figure shows a sample trajectory, as a dotted line, the ellipse that fully contains it, as a dashed line, the focal distance as a dash-dot line and the major axis as a solid line. (d) The probability density function of the eccentricity of the ellipse containing each trip, shown in blue, and for comparison, the same quantity measured for the optimal trips, shown in green. While both groups of trajectories are characterized by high-eccentricity, optimal trips are slightly less eccentric than actual user trips, suggesting that the former deviate slightly more from the ideal origin-destination straight line. (e) The boxplot shows the values of eccentricity $e$ against the scale of the journey $f$. The median eccentricity value does change considerably at different scales.

ings generated by this urban science can be successfully used to design simple yet innovative solutions [45, 46] that can help cities of today turn into smart cities of tomorrow.

## Materials and methods

**GPS data**. The dataset contains information about the trajectories followed by 526 users in an undisclosed European country over a period of 18 months. The trajectories followed by the cars were collected by GPS devices installed on them. Each trajectory is composed by periodic location updates, taken every 60 seconds, starting when the driver turns the engine on, until it is turned off. We remove inconsistent data points that are collected when the number of satellites available is lower than 4 and we remove sudden GPS jumps that are inconsistent with average travel speeds higher than 110 km/h. All user IDs were given in anonymized form.

**Significant locations extraction.** We extracted each user's significant locations by clustering the starting and ending point of each trajectory. The geographic distance between points was computed using the Haversine formula. The clustering was performed using the mean shift algorithm. This clustering method detects groups of points that are dispersed around a center, according to a Gaussian distribution. By choosing the bandwidth parameter $\gamma = 0.025$ we find clusters of points that are distant from each other at most by 600 m. These points can be reasonably different parking spots used to to reach the same final destination, located at walking distance.

**Distance between trajectories** In order to compare trajectories, which in general are defined by an heterogeneous number of points, we use the Dynamic Time Warping (DTW) algorithm, traditionally used in speech recognition and shape analysis. Given two paths $A = [a_1, a_2, ..., a_N]$ and $B = [b_1, b_2, ..., b_M]$, specified as sequences of geographic points of different length, we first find an alignment such that the following recursive definition, for $i = 1 \ldots N-1$, $j = 1 \ldots M-1$, is minimized:

$$W(A_i, B_j) = d(a_i, b_j) + \min \begin{cases} W(A_{i+1}, B_{j+1}) \\ W(A_{i+1}, B_j) \\ W(A_i, B_{j+1}) \end{cases} \quad (1)$$

where $A_i$ and $B_j$ are subsequences containing all the elements $1 \dots i$ from $A$ and $1 \dots j$ from B, respectively; the element-wise distance $d$ is here considered to be the Haversine distance. The algorithm tries to match each point in $A$ with a point in $B$, taking into consideration the sequence order. Initially, the two starting points are associated; then the algorithm advances one of the two trajectories, or both, depending on which pair of points minimizes the element-wise distance; the algorithm proceeds until both end points are reached. Once the alignment is found between the two trajectories, we consider the maximum distance between all the matched pairs of points.

**Route detection** Clustering of trajectories in a routine trip is performed using the DBSCAN algorithm on the maximum distance in the DTW-aligned trajectories, obtained as previously described. This clustering method has the advantage of not needing to specify the number of groups. However, it is necessary to choose two parameters: $B$, the minimum number of trajectories necessary to form a route, and $\epsilon$, the maximum distance to consider an element part of the cluster. We set $B = 1$, so that a single different trajectory is considered as a distinct route choice. We obtained the best clustering results with a choice of $\epsilon = 0.5\,\text{km}$; such a value is reasonable, considering that a car traveling at 30 km/h covers that distance during the sample period of 60 seconds.

**Normalized Gini coefficient.** The Gini coefficient $G$ is a statistical index of dispersion of values, typically used in economics to quantify the inequality of income among people. Its value is bounded between $0 \leq G \leq 1 - \frac{1}{N}$, where $N$ is the size of the population; the coefficient is null for perfect equality and maximum for complete inequality. We use the Gini coefficient to quantify, for a routine trip, how similar the usage frequencies are among all the routes employed at least once by the user. In order to compare this index on routine trips with an heterogeneous number of routes $N$, which is typically small, we consider a variant of the Gini coefficient, normalized by the maximum value the Gini index $G$ obtainable with a number $N$ of routes:

$$G_n = \frac{G}{1 - \frac{1}{N}} \quad (2)$$

As a consequence $G_n = 1$ for perfect inequality, regardless of the number of elements considered.


### Acknowledgments

The research was partly funded by the World Bank, the Future Urban Mobility project of SMART, the Department of Transportation's grant of the New England UTC Y25, the MIT Portugal Program and the Center for Complex Engineering Systems at KAUST-MIT. A. Lima is funded by the Vest Scholarship.



[1] González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).

[2] Prato, C. G. Route choice modeling: past, present and future research directions. *Journal of Choice Modelling* **2**, 65–100 (2009).

[3] Wardrop, J. G. Some theoretical aspects of road traffic research. *Proc. Inst. Civ. Eng.* **1**, 325–362 (1952).

[4] Cohen, J. E. Human Population: The Next Half Century. *Science* **302**, 1172–1175 (2003).

[5] Brockmann, D., Hufnagel, L. & Geisel, T. The scaling laws of human travel. *Nature* **439**, 462–465 (2006).

[6] Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z. & González, M. C. Unravelling daily human mobility motifs. *J. R. Soc. Interface* **10**, 20130246 (2013).

[7] Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of Predictability in Human Mobility. *Science* **327**, 1018–1021 (2010).

[8] de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. & Blondel, V. D. Unique in the Crowd: The privacy bounds of human mobility. *Sci. Rep.* **3** (2013).

[9] Louail, T. *et al.* Uncovering the spatial structure of mobility networks. *Nat Commun* **6** (2015).

[10] Stopher, P. R. & Greaves, S. P. Household travel surveys: Where are we going? *Transport Res A-Pol* **41**, 367–381 (2007).

[11] Wang, P., Hunter, T., Bayen, A. M., Schechtner, K. & González, M. C. Understanding Road Usage Patterns in Urban Areas. *Sci. Rep.* **2** (2012).

[12] Alexander, L., Jiang, S., Murga, M. & González, M. C. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transport Res C-Emer* (2015).

[13] Toole, J. L., Colak, S., Alhasoun, F., Evsukoff, A. & Gonzalez, M. C. The path most travelled: Mining road usage patterns from massive call data. *ArXiv14030636 Phys.* (2014).

[14] Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R. & González, M. C. Analyzing Cell Phone Location Data for Urban Travel. *Transportation Research Record: Journal of the Transportation Research Board* **2526**, 126–135 (2015).

[15] Barthélemy, M. Spatial networks. *Physics Reports* **499**, 1–101 (2011).

[16] Daganzo, C. F. & Sheffi, Y. On Stochastic Models of Traffic Assignment. *Transportation Science* **11**, 253–274 (1977).

[17] Cascetta, E., Russo, F. & Vitetta, A. Stochastic User Equilibrium Assignment with Explicit Path Enumeration: Comparison of Models and Algorithms. In *Transportation Systems* (1997).

[18] Frejinger, E., Bierlaire, M. & Ben-Akiva, M. Sampling of alternatives for route choice modeling. *Transport Res B-Meth* **43**, 984–994 (2009).

[19] Prato, C. & Bekhor, S. Applying Branch-and-Bound Technique to Route Choice Set Generation. *Transportation Research Record: Journal of the Transportation Research Board* **1985**, 19–28 (2006).

[20] Levinson, D. & Zhu, S. A portfolio theory of route choice. *Transport Res C-Emer* **35**, 232–243 (2013).

[21] Jan, O., Horowitz, A. J. & Peng, Z.-R. Using Global Positioning System Data to Understand Variations in Path Choice. *Transp. Res. Rec. J. Transp. Res. Board* **1725**, 37–44 (2000).

[22] Selten, R., Chmura, T., Pitz, T., Kube, S. & Schreckenberg, M. Commuters route choice behaviour. *Games and Economic Behavior* **58**, 394–406 (2007).

[23] Nielsen, O. A., Daly, A. & Frederiksen, R. D. A stochastic route choice model for car travellers in the Copenhagen region. *Netw. Spat. Econ.* **2**, 327–346 (2002).

[24] Bailenson, J. N., Shum, M. S. & Uttal, D. H. The initial segment strategy: A heuristic for route selection. *Mem. Cognit.* **28**, 306–318 (2000).

[25] Hochmair, H. H. & Karlsson, V. Investigation of Preference Between the Least-Angle Strategy and the Initial Segment Strategy for Route Selection in Unknown Environments. In Freksa, C., Knauff, M., Krieg-Brückner, B., Nebel, B. & Barkowsky, T. (eds.) *Spatial Cognition IV. Reasoning, Action, Interaction*, no. 3343 in Lecture Notes in Computer Science, 79–97 (Springer Berlin Heidelberg, 2004).

[26] Brunyé, T. T., Mahoney, C. R., Gardony, A. L. & Taylor, H. A. North is up(hill): route planning heuristics in real-world environments. *Mem Cognit* **38**, 700–712 (2010).

[27] Brunyé, T. T. *et al.* Planning routes around the world: International evidence for southern route preferences. *Journal of Environmental Psychology* **32**, 297–304 (2012).

[28] Golledge, R. G., Smith, T. R., Pellegrino, J. W., Doherty, S. & Marshall, S. P. A conceptual model and empirical analysis of children's acquisition of spatial knowledge. *Journal of Environmental Psychology* **5**, 125–152 (1985).

[29] Manley, E. J., Addison, J. D. & Cheng, T. Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. *Journal of Transport Geography* **43**, 123–139 (2015).

[30] Manley, E. J., Orr, S. W. & Cheng, T. A heuristic model of bounded route choice in urban areas. *Transportation Research Part C: Emerging Technologies* **56**, 195–209 (2015).

[31] Brunyé, T. T. *et al.* Strategies for Selecting Routes through Real-World Environments: Relative Topography, Initial Route Straightness, and Cardinal Direction. *PLoS ONE* **10**, e0124404 (2015).

[32] Hochmair, H. & Frank, A. U. Influence of estimation errors on wayfinding-decisions in unknown street networks – analyzing the least-angle strategy. *Spatial Cognition and Computation* **2**, 283–313 (2000).

[33] Chorus, C. G., Arentze, T. A. & Timmermans, H. J. P. A Random Regret-Minimization model of travel choice. *Transportation Research Part B: Methodological* **42**, 1–18 (2008).

[34] Senk, P. Route Choice Under the Microscope. *Transportation Research Record: Journal of the Transportation Research Board* **2156**, 56–63 (2010).

[35] Giannotti, F., Nanni, M., Pinelli, F. & Pedreschi, D. Trajectory Pattern Mining. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '07, 330–339 (ACM, New York, NY, USA, 2007).

[36] Zheng, Y. & Zhou, X. *Computing with spatial trajectories* (Springer Science & Business Media, 2011).

[37] Limpert, E., Stahel, W. A. & Abbt, M. Log-normal Distributions across the Sciences: Keys and Clues On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *BioScience* **51**, 341–352 (2001).

[38] Kölbl, R. & Helbing, D. Energy laws in human travel behaviour. *New J. Phys.* **5**, 48 (2003).

[39] Zhu, S. & Levinson, D. Do People Use the Shortest Path? An Empirical Test of Wardrop's First Principle. *PLoS ONE* **10**, e0134322 (2015).

[40] Porta, S., Crucitti, P. & Latora, V. The network analysis of urban streets: a primal approach. *Environ. Plan. B Plan. Des.* **33**, 705 – 725 (2006).

[41] Newsome, T. H., Walcott, W. A. & Smith, P. D. Urban activity spaces: Illustrations and application of a conceptual model for integrating the time and space dimensions. *Transportation* **25**, 357–377 (1998).

[42] Schönfelder, S. & Axhausen, K. W. Activity spaces: measures of social exclusion? *Transport Policy* **10**, 273–286 (2003).

[43] Buliung, R. N. & Kanaroglou, P. S. Urban Form and Household Activity-Travel Behavior. *Growth Change* **37**, 172–199 (2006).

[44] Zheng, Y., Capra, L., Wolfson, O. & Yang, H. Urban Computing: Concepts, Methodologies, and Applications. *ACM Trans Intell Syst Technol* **5**, 38:1–38:55 (2014).

[45] Strano, E., Shai, S., Dobson, S. & Barthelemy, M. Multiplex networks in metropolitan areas: generic features and local effects. *J. R. Soc. Interface* **12**, 20150651 (2015).

[46] Domenico, M. D., Lima, A., González, M. C. & Arenas, A. Personalized routing for multitudes in smart cities. *EPJ Data Sci.* **4**, 1–11 (2015).