




Individual mobility deep insight using mobile phones data

Chiara Mizzi¹, Alex Baroncini¹, Alessandro Fabbri¹, Davide Micheli³, Aldo Vannelli³, Carmen Criminisi³, Susanna Jean³ and Armando Bazzani^{1,2*} 

*Correspondence:

armando.bazzani@unibo.it

¹Laboratory of City Science, Physics and Astronomy Department, University of Bologna, Bologna, Italy

²INFN—sezione di Bologna, Bologna, Italy

Full list of author information is available at the end of the article

Abstract

The data sets provided by Information and Communication Technologies have been extensively used to study the human mobility in the framework of complex systems. The possibility of detecting the behavior of individuals performing the urban mobility may offer the possibility of understanding how to realize a transition to a sustainable mobility in future smart cities. The Statistical Physics approach considers the statistical distributions of human mobility to discover universal features. Under this point of view the power laws distributions has been extensively studied to propose model of human mobility. In this paper we show that using a GPS data set containing the displacements of mobile devices in an area around the city Rimini (Italy), it is possible to reconstruct a sample of mobility paths and to study the statistical properties of urban mobility. Applying a fuzzy c-means clustering algorithm, we succeed to detect different mobility types that highlight the multilayer structure of the road network. The disaggregation into homogeneous mobility classes explains the power law distributions for the path lengths and the travel times as an overlapping of exponential distributions, that are consistent with a maximum entropy Principle. Under this point of view it is not possible to infer other dynamical properties on the individual mobility, except for the average values of the different classes. We also study the role of the mobility types, when one restricts the analysis to the an origin-destination framework, by analyzing the daily evolution of the mobility flows.

Keywords: Human mobility; Mobile phone data; Path length and duration distributions; Multiscale road network

1 Introduction

Urban mobility is a key issue for the sustainability of the future smart cities [1] and the project of a digital twin for the city. The development of a Mobility as a Service [2] that offers different solutions to satisfy the citizen mobility demand has not only to cope with the problem of building new infrastructures for innovative transportation means, but also to understand the individual behavior in realizing the mobility demand in order to promote the transition to a smart urban mobility. This goal requires to collect and analyze dynamical data sets on the individual behavior and to build predictive models able to simulate the individual dynamics in future scenarios [3]. From one hand one has to consider the re-

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

restrictions due to the privacy laws (especially in Europe) so that the availability of data sets is limited and it is necessary to develop a statistical physics approach with a partial knowledge of the microscopic dynamics. On the other hand the predictability of traffic models is compromised by the chaotic and stochastic properties of traffic dynamics. The possibility offered by the new Information and Communication Technologies (ICT) allowed TIM, one of the leader telecommunication companies in Italy with $\simeq 30\%$ of penetration in the mobile phone population, to collect GPS data on the position of mobile phones in an anonymous way, for a sample of the population present in a large area (MDT data set [4–6]). In this way it is possible to reconstruct a sample of individual mobility paths in a dynamical way [7]. The ICT data sets have been extensively used to understand the statistical universal laws of human mobility [8–13] and the scaling laws of human mobility have been proposed as emergent properties to understand the features of individual behaviors [14–18]. However, previous papers have studied GPS dynamical data sets for the private car mobility [19–21] on an urban road network suggesting an exponential distribution for the path lengths and the travel times and pointing out the existence of a conserved ‘mobility energy’ on average [22]. The existence of conservation laws for human mobility is also discussed in the paper [23]. Therefore the power law distributions observed for long displacements [24], that correspond to the absence of characteristic spatial scales in how people travel [11] can be explained by the availability of different transport networks that introduce a non-homogeneity in the space [25–29]. The presence of different transportation networks is a property of the considered area and should introduce the same scaling laws for all the individuals. Then, the multilayer structure of the road network may be detected if it is possible to disaggregate the mobility data into homogeneous classes. Another question is if the mobility on the different transport networks still contains information on the origin-destination mobility or it is dominated by a random-like mobility. Finally, it is not clear the relevance of these results in the case of urban mobility, where the spatial scales are limited by the city dimension, and if a statistical physics approach would allow to infer the dynamical properties of individual behavior: i.e. how individuals choose their mobility paths or change their paths in case of under different traffic conditions or how they behave performing multimodal mobility [30]. The research in urban mobility aims to define a roadmap towards a sustainable mobility, so that it is a key issue to model the individual behaviors to reduce the congestion impact and to predict the impact of new mobility infrastructures or the effect of new mobility policies. It is important to understand the data quality required to cope with these problems: i.e. how effective can be a statistical physics approach to understand the individual behavior or if we really need to get the paths of a great number of individuals in each city with all the privacy issues that entails.

In the paper we address these problems using the MDT data set collected during August 2020 in the area of Rimini province, the Italian city on the Adriatic sea well known for the tourist activities. The data contain information on the mobility both of the residents and the visitors. Even if during the summer 2020 tourist activities were affected by the consequences of the COVID19 epidemic (the presence of foreign tourists was greatly reduced), there were more than 4×10^4 visitors who arrived in Rimini from the Emilia Romagna region, and $\simeq 5 \times 10^4$ visitors who arrived in Rimini from the rest of Italy with respect to a resident population of 150,000 inhabitants (see Fig. 1 in the Additional file 1).

Since the private car is the preferred transport means to reach the tourist locations in Italy, the traffic and the traffic problems in the area are particularly relevant.

The MDT data set contains GPS quality data collected from the smartphone population present in the area without any particular bias, but there is an intrinsic difficulty to check the representativeness of the data sample. If a monitored mobile phone performs an activity connecting to the communication network, it is possible to reconstruct the dynamics of a mobility path by georeferencing the data on the road network. We show how the application of a fuzzy c-means clustering (FCM) algorithm [31] turns out to be effective in classifying different types of mobility, that can be identified as a slow mobility, an urban mobility, an extra urban mobility and a highway mobility. By using the paths associated to the different classes, it is possible to detect the road sub-networks that show the expected non homogeneous structure of the whole road network. We perform a statistical analysis of the path lengths and travel time distributions to show that a power law provides a better interpolation of the distributions, in particular of the path length distribution even if the spatial dimension is limited. However, this behavior is an emergent property of the superposition of exponential distributions since the corresponding distributions of the single mobility classes are exponential-like and collapse if one normalizes each variable by the average value of each class. In the case of the path length, the average values characterize the different mobility types. On the contrary the average values for the travel times are very similar suggesting that travel time could be used as a measure of the *mobility energy*. Under this point of view, one can apply a maximum entropy (ME) Principle [32] since the exponential distribution is the distribution that maximizes the uncertainty on the system state using the Information Entropy constrained by existence of average values of the observables. This means that the statistical distributions computed in a stationary (or almost stationary) situation, do not allow to infer dynamical properties of the individuals, except for the existence of preserved quantities in average. Indeed the assumption that individuals move independently in a random way performing paths of a given average length in each road subnetworks, is consistent with the empirical statistical distributions computed using the MDT paths. Since it is certainly true that each individual moves according to an origin-destination (OD) logic, the empirical exponential distributions for the path lengths and travel times in the case of a homogeneous mobility, imply that there is no other spatial or time scale beyond the mean values. For a better understanding of the problem we have analyzed the role of the different mobility types when one considers a specific OD mobility demand. More precisely, we have studied how the different mobility types may explain the observed mobility between the coastal area, where most of the tourist activities are concentrated, and the inland. This OD-mobility involves both people resident in the Rimini, people coming from the nearby small towns and the visitors arriving in Rimini for the mid-August holidays. The dynamical character of the MDT data set highlights the daily mobility oscillations suggesting a correlation between the mobility demand, the tourist activities time schedule and the traffic conditions. Our results show that the choice of slow mobility is mainly affected by the distance from the destination, whereas the travel time and the traffic problems do not discourage the use of private cars. We also study the attractiveness dependence from the destination distance by means of the complementary cumulative distribution of the trip lengths. The trip frequency at a given distance can be related to the existence of a visitation law, useful to quantify the attractiveness of an area. In a recent work [33], the existence of a gravitational-like law to predict the mobility

demand has been verified using large mobility data sets. Assuming an uniform people distribution in the inland (this is a reasonable assumption for the Po Valley area in Italy), we infer that the empirical distribution for the trip lengths is consistent with a weak power law dependence $l^{-\eta}$ with $\eta \simeq 1$ for the probability to observe a path of length l toward the coast, which is consistent with the gravitational-like law.

The paper is organized as follows: in the second section we describe the main features of the MDT data set; in the third section we present the main results of the paper using the possibility of disaggregating the mobility paths into different classes by a fuzzy c-means clustering that highlight the multilayer structure of the road network. Then we explain the power law distribution for urban mobility by applying the ME Principle to the detected mobility classes. In the fourth section we study the relevance of mobility classes for a specific OD mobility and some conclusive remarks are reported in the last section.

2 The MDT data set

Taking advantage from the TIM MDT technology to record the GPS positions of mobile devices in an anonymous way [4–6] when they are connected to the telecommunication network, it is possible to detect individual paths in a large area. The data are related to a sample of mobile phones connected using the Long-Term Evolution (4G-LTE) wireless broadband communication recording the GPS positions at fixed time intervals and give the GPS position each 5 seconds of the mobile device during its activity. An anonymous *id* is associated to the activity, so that the same mobile phone can be localized by different *id*. The MDT data set used in the paper refers to the data collected from 7 to 17 August 2020 in an area of the Rimini province containing the Rimini municipality (see Fig. 1). We have estimated the sample penetration by comparing the MDT activities by comparing the number of different activities and the total number of mobile phones present in the area during the mid-August weekend estimated using another data set provided by TIM-Olivetti that counts how many mobile phones are connected to the telecommunication network in the census areas (ACE) of Rimini each 15 minutes (TIM is one of the main telecommunication companies in Italy with a coverage of $\simeq 30\%$ of the market). Under the assumption that the effects of repeated activities is not relevant at a time scale of 15 min., the estimated penetration of the MDT data set is $\simeq 8\%$ of the total mobile phones. The details of the analysis are reported first section of the Additional file 1 (cfr. Figure 2). The restriction mobility policies due to the COVID-19 epidemic, reduces the number of foreigners present in Rimini, but many Italian people visited Rimini during the mid-August holidays. Using the TIM-Olivetti data set, it is possible to estimate the presences of almost 10^5 visitors in Rimini, with respect to a resident population of $\simeq 1.5 \times 10^5$ people (cfr. Figure 1 of the Additional file 1). Then, we expect to detect also some effect due to the visitor mobility in the MDT data set.

Each activity consists in a sequence of GPS points at a time sampling of 5 sec. but we have specific algorithms to select the activities that can be associated to a mobility path. These algorithms have been developed for another MDT data set in Venice [7] and they allowed to reconstruct the mobility paths on the road network of the Opens Street Map cartography [34] (more details are reported in the Sect. 2 of the Additional file 1). The MDT data set contains $\sim 1.8 \times 10^6$ activities per day, but the amount of records that can be correctly georeferenced reduces to $\simeq 1.5 \times 10^6$ per day. In Fig. 1 we show the georeferenced data distribution, that shows a concentration of points not only inside the urban

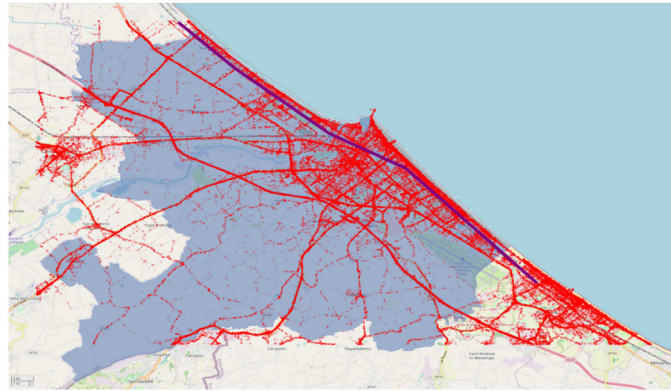


Figure 1 The grey area corresponds to the Rimini municipality, whose linear extension of the coast is $\simeq 25$ km (the spatial scale is reported on the left bottom angle) and the vertical axis is oriented along the North. The red dots show the distribution of the GPS mobile phone data set recorded in the considered area. We observe a concentration of points inside the urban settlements (in particular along the Adriatic sea coast Rimini and Riccione are the main cities), and along the main country roads. A small town (Santarcangelo di Romagna) is also visible in the inland on the left. The blue line corresponds to the railway that is a natural border between the coastal area, where most of the tourist activities are located, and the inland

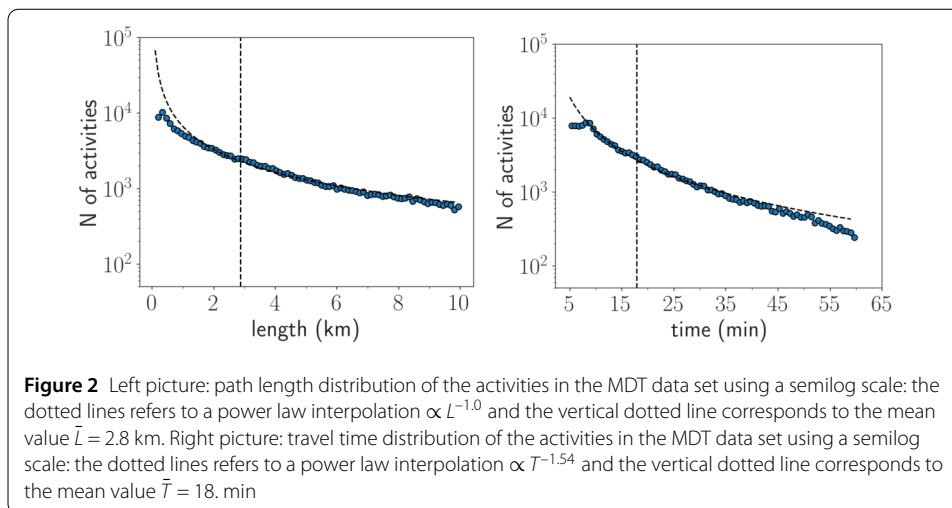
settlements, but also along the main country roads where many small towns are located. The filtering algorithms to detect the mobility paths provide $\simeq 40000$ paths per day.

To check the capacity of the MDT data set to reproduce the real mobility in the area, we have computed the hourly traffic flows along the main country roads average on the considered period by using the dynamics of the reconstructed paths. Then we have compared the estimated traffic flows with the real traffic flows recorded by magnetic coils along three roads. The results are reported in the Additional file 1 (cfr. Figure 5) and measure a penetration of the MDT traffic flows of $\geq 1.5\%$ of the total traffic flows, but with a quite good reproduction of the daily evolution of the real traffic flows. Then, despite of the low penetration of the MDT sample, the reconstructed paths contain relevant information on the mobility observed in the area and they allow to perform a statistical physics approach to study the mobility features.

The data sets analysed in the paper are not publicly available since they are subjected to data use agreements, but the aggregated data to generate all the figures are available on request from the corresponding author.

2.1 Statistical properties of the mobility paths

To understand the statistical properties of the observed mobility, we start from the path length and the travel time distributions (see Fig. 2) that have been also considered by several authors. Since we expect that the MDT data set contains the contribution of different types of mobility, we consider the problem of studying the effect on the behavior of the statistical distributions. Various papers suggest the existence of power law interpolations [8, 14, 21] for these mobility distribution functions. In Fig. 2 a power law interpolation $\rho(x) \propto x^{-\alpha}$ is proposed for both the path length and travel time distributions with exponents $\alpha = 1.0$ (for the path length) and $\alpha = 1.54$ (for the travel times) [24]. The power law turns out to be more effective than an exponential interpolation for both the distributions, even if the limited spatial and temporal extension of our data (one order of magnitude for both the path length and the travel time) does not allow to draw a definitive conclusion,



and the tail of the travel time distribution seems to decay exponentially. However the multimodal distribution for the average velocity of the reconstructed detected paths (see Fig. 3 (left)) clearly suggests that the MDT data set contains different types of mobility. Different authors [25–29] suggest that the power law distributions of the urban mobility could be considered an emergent property of the urban mobility explained by the heterogeneity of the transportation networks present in the area, and not to a specific property of the individual dynamics: i.e. it is not justified to relate the mobility power law distributions to the individual dynamics since they are explained by the inhomogeneity of the urban space. This approach follows the theoretical framework explained in the paper [35] and it is consistent with the results of the paper [11]. The exponential distribution plays a central roles in Statistical Physics since it can be derived from a ME Principle [32] for a system in a equilibrium state. As described in the papers [3, 19] the ME Principle assumes that the stationary distribution $\rho(x)$ where x is a microstate, maximizes the Information Entropy

$$S[\rho] = -k \int \rho(x) \ln \rho(x) dx \quad (1)$$

(k is a suitable constant that can be set to $k = 1$) with the constraint of the existence of an average conserved quantity. The justification of the ME Principle lies in the unique properties of the definition (1) as a measure of the amount of uncertainty represented by the probability distribution $\rho(x)$. The maximum entropy distribution turns out to be the most probable distribution for the considered quantity consistent with the existence of the constraint in the average of the conserved quantity. From a physical point of view, the entropy is a state function that can be correctly defined when a statistical system is in equilibrium state, however the stochastic thermodynamics [36] has shown the possibility to extend the entropy concept to stochastic dynamical systems and to study the relaxation process to a stationary state using the entropy production. The application of the ME Principle to human mobility assumes the possibility of describing the mobility as a random walk in a homogeneous space with a cost function associated to the path length (or the travel time). In this case all the individuals move as independent particles and the probability $\rho(L, t)$ to

observe a path of length L after a time t is described by the Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} = c \frac{\partial \rho}{\partial L} + D \frac{\partial^2 \rho}{\partial L^2}, \quad (2)$$

where c is the cost of a path per unit length and time and D is the diffusion coefficient in the considered area. If $c = 0$ (i.e. there is no cost for a path length) then the distribution $\rho(L, t)$ is a Gaussian function whose variance increases as $2Dt$. A direct computation using the definition (1) provides the stationary solution

$$c \frac{\partial \rho}{\partial L} + D \frac{\partial^2 \rho}{\partial L^2} = 0 \quad \Rightarrow \quad \rho(L) \propto \exp\left(-\frac{c}{D}L\right)$$

that coincides with the extremality condition of the functional (1) with the constraint

$$\int_0^\infty L \rho(L) dL = \bar{L} = \frac{D}{c},$$

where \bar{L} is the average value. The existence of a mobility cost has been considered by various authors after the seminal paper [22], and the complexity of the individual behavior implies that the stochastic effects in the urban mobility are more relevant than the deterministic effects, due to the existence of common origin-destination areas. This does not mean that individuals behave as random particles, but that the uncertainty of the individual dynamics implies that the statistical properties of the urban mobility are the same as the individuals move as random particles. The possibility of assuming a stationary condition can be justified when we consider a time average over a sufficiently long period. Therefore the exponential distribution is explained by a ME Principle when one considers a homogeneous mobility, even if each individual moves according to a best path strategy toward his destinations. Under this point of view, we remark that the understanding of the statistical distributions does not allow to infer any prediction on the system evolution when external perturbations introduce non-equilibrium conditions and the properties of individual dynamics become relevant.

If different transportation networks are present, the homogeneity assumption is no more justified, since the mobility cost may depend on the transportation network: for pedestrian mobility the cost is directly the energy required by the walking activity, for the urban mobility the cost may be represented by the time spent in traffic and for the extra-urban mobility the cost could be related to the path lengths. However if it is possible to disaggregate the urban mobility into different homogeneous mobility classes to which the ME Principle can be applied, the power law distributions can be explained by an overlapping of exponential distributions [20] and they characterize the degree of non-homogeneity of the urban space and how individuals use the different transportation network (in average) to realize their mobility demand.

In the next section we will show that disaggregating the data set into homogeneous mobility classes allows to justify the previous argument.

3 Detecting different mobility types

To highlight homogeneous mobility classes, we apply a fuzzy c-means clustering algorithm [31] to the MDT data set with a soft threshold. The original algorithm has been

Table 1 Percentage of the selected paths for each class defined by the fuzzy c-means clustering, average values for the path length, the travel time, the velocity and the sinuosity of the paths: the statistical error on the parameter values is implied in the last digit

Class	% of selected paths	$L_m(km)$	T_m (min.)	V_m (km/h)	Sinuosity
0	19.4%	0.99	18.8	4.4	0.74
1	22.7%	2.95	14.5	16.4	0.78
2	43.7%	6.64	13.3	36.0	0.88
3	14.2%	14.37	7.8	113.1	0.97

modified to classify the MDT paths [37]. The algorithm proves efficient to classify almost all the reconstructed trajectories ($\simeq 96\%$ of the selected paths) (more details on the application of fuzzy c-means clustering algorithm to the MDT paths are reported in the Sect. 3 of the Additional file 1). The classification procedure is based on the following four features of each path:

- The average velocity defined by the ratio of path length (i.e. the sum of the distance between consecutive points) and travel time;
- The maximum velocity: the maximum among the instantaneous speeds measured using two consecutive georeferenced points;
- The minimum speed of path: the minimum among the instantaneous speeds;
- The sinuosity: the ratio of the Euclidean distance between the first and last record of a path and the length measured as the sum of all record distances (this value can be a maximum of 1.0 and indicates how curvy the path is: the smaller the sinuosity, the more tortuous is the chosen path).

The algorithm identifies four classes numbered from 0 to 3, that we are able to interpret as a slow mobility (a mix of pedestrian and bike mobility), the urban traffic mobility, the extra-urban traffic and the highway traffic. In the Table 1 we report the average statistical values characterizing each class, together with the percentage of the selected paths belonging to the class and the average sinuosity index. We remark that average velocity and the sinuosity are used by the fuzzy *c*-mean algorithm and take different values for the different classes, whereas the average path length and the travel time are emergent properties of the detected classes. We remark as the velocity is the main feature that characterizes the mobility type: indeed the first class 0 contains paths with an average velocity 4.4 km/h that is a typical velocity for a *slow mobility*, whereas the other three classes have average velocities that can be associated to private car mobility in different contexts. The average velocity of 16.5 km/h of the class 1 is typical of an *urban traffic* and the class 3 velocity is clearly related to the *highway traffic* that is present in the area. The class 2 has an average velocity of 36 km/h and we associate this class with the *extra-urban traffic*. The percentage associated to the different classes measures the relevance of each classes to explain the mobility in the considered area: the first two classes mainly refer to the urban mobility, that is divided between the slow mobility and the urban traffic in a similar percentage. The extra-urban mobility represents the largest class suggesting that is probably due to the mobility demand from the inland toward the coastal area. Finally the highway traffic (the smallest class) is mainly due to traffic crossing the area.

The average path length changes among the four classes consistently with our interpretation. An average path length of 1 km for the slow mobility is a typical spatial scale for a mixed pedestrian and bike mobility. The private cars are used to perform relative small paths in a urban context, and a value of $\simeq 3$ km is consistent with the urban path length

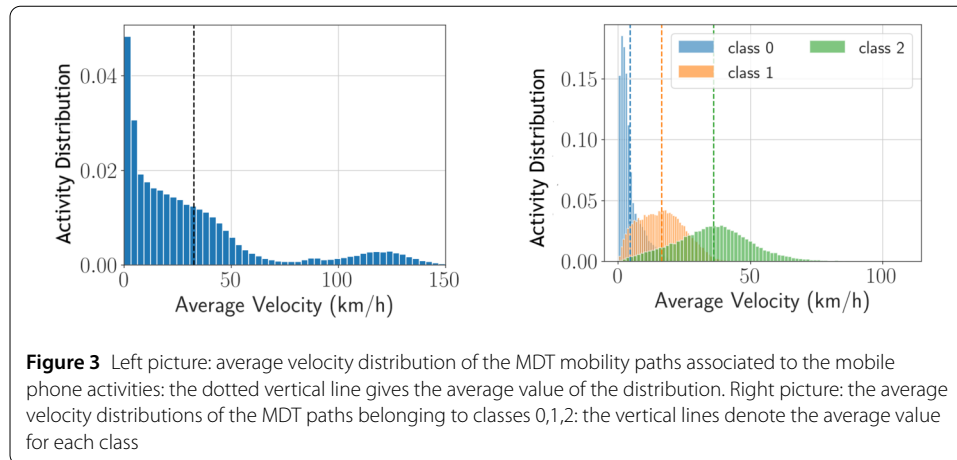
obtained using different data sets [19]. Finally, we observe that the path length of the highway traffic is of the same order of the spatial dimension of the considered area, so that the corresponding paths could be part of longer paths crossing the area. Regarding average travel times T_m , we observe that for the highway traffic, the travel time and the average path length satisfy the relation $T_m V_m \simeq L_m$ denoting a small variability in the path velocity according to the interpretation that the highway traffic is crossing the area. Considering the first three classes, the travel time has a limited variation, suggesting that it could be related to a mobility cost function [19, 22, 38]: the individuals seem to organize their mobility using a typical time budget and this fact could also influence the decisions to use of different transportation means. Of course, in the case of the slow mobility there is a proportionality between the travel time and the fatigue during the mobility, but also driving in congested condition can introduce a fatigue effect. We observe that for the mobility classes 0,1,2 the ratio L_m/T_m is always smaller than the average velocity V_m , that could be explained assuming that the fluctuations in the average velocity are mainly due to the travel time fluctuations δT . For a given length L_m we have

$$V_m = \left\langle \frac{L_m}{T_m(1 + \delta T/T_m)} \right\rangle \simeq \frac{L_m}{T_m} \left\langle 1 - \frac{\delta T}{T_m} + \left(\frac{\delta T}{T_m} \right)^2 \right\rangle,$$

where $\delta T/T_m$ are the relative travel time fluctuations (i.e. $\langle \delta T \rangle = 0$). Then we get the relation

$$V_m \simeq \frac{L_m}{T_m} \left(1 - \frac{\langle \delta T^2 \rangle}{T_m^2} \right) < \frac{L_m}{V_m} \quad (3)$$

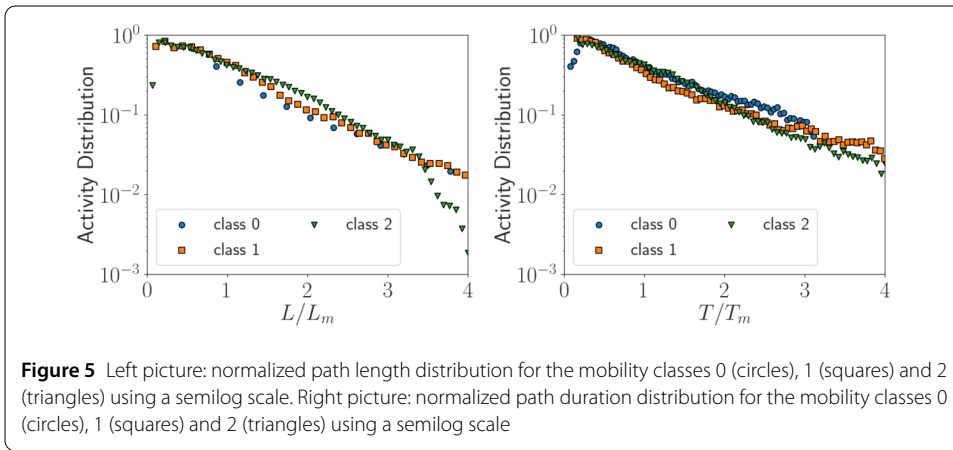
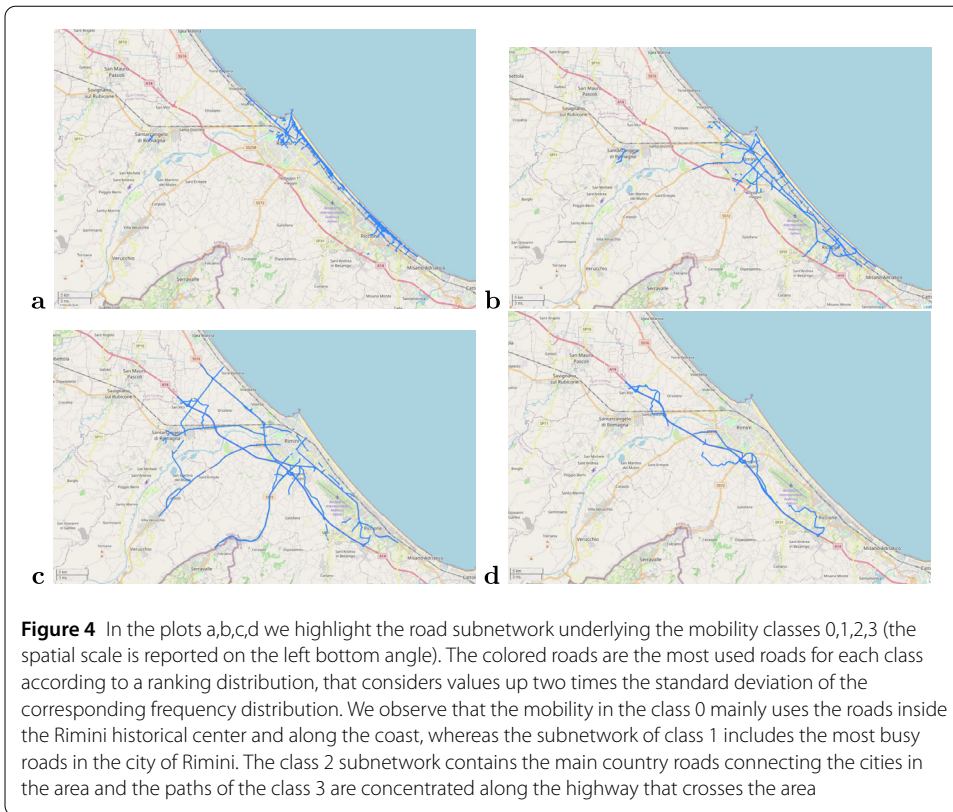
The travel time fluctuations could be related both to the individual heterogeneity and to traffic conditions in the cities. Indeed, for the urban mobility the reduction factor is ≤ 0.75 that suggests a greater relevance of the fluctuations in the traffic dynamics due to the heterogeneity in the population or to congestion effects. In Fig. 3 (right), we show how the average velocity for the detected the MDT mobility paths can be recovered by using the overlap of velocity distributions for the classes 0,1,2, if we neglect the highway traffic. We observe that as the distribution of the slow mobility (the class 0) is very narrow due to the obvious upper limit in the pedestrian velocity, with a right tail that probably corresponds to the bike mobility. The urban traffic mobility (class 1) is a bell shaped skewed distribution, which distinguish the left part (near zero) that is probably affected by the traffic rules and the congestion effects. Finally the extraurban traffic (class 2) provides a symmetric bell shaped distribution. The results shown in Fig. 3 suggest that the selected classes represent a homogeneous mobility performed in the area and the increasing of the sinuosity parameter for the different classes (see Table 1) could be related to the existence a multilayer structure in the road network [39]. To check the previous conjecture and to highlight the road sub-networks underlying each mobility class, we have georeferenced the mobility paths on the road network and applied an algorithmic procedure that considers a ranking distribution of the roads according to the number of crossing paths for each class, and connects all the roads in the ranking up to a given threshold, defined using the standard deviation of the frequency distribution of the crossing paths [7]. In Fig. 4 we show the four road sub-networks associated to the mobility classes. The differences among the road networks of the classes are clear. The sub-network of the class 0 contains



the roads along the coast where the tourist activities are concentrated and the roads in the Rimini historical center where there are traffic restriction rules. The sub-network of the class 1 is mainly composed by the most busy urban roads that rule the traffic mobility around the historical center and toward the coastal area, whereas the road sub-network of the class 2 contains the country roads of the area. This subnetwork can explain both the mobility due to the visitors that arrive to the coast by car and the mobility due to the commuters. Finally, we remark as that the paths of the last class are mainly concentrated along the highway and they probably correspond to a mobility crossing the area.

3.1 Power law distributions and the maximum entropy Principle

To understand the power law interpolations shown in Fig. 2 as the consequence of the existence of a multilayer structure for the road network [27, 40], we decided to restrict our analysis to the first three classes, 0,1,2. To explain the mobility statistical distribution of Fig. 2, we consider if each mobility class can be interpreted as a statistical system in a maximal entropy equilibrium state, so that the path length and the travel time distributions are exponentially distributed according to the existence of a *mobility energy* [20, 22]. This is certainly not true for the highway mobility since the paths are mainly concentrated along a single road (the highway). According to this approach, the average path lengths L_m and the average travel time T_m of the classes 0,1,2 (see Table 1) should completely characterize the corresponding distribution functions. Therefore, if one considers the normalized variables L/L_m and T/T_m for each class, the corresponding distributions should collapse to a single exponential distribution function $\rho(x) \propto \exp(-x)$. The normalized distributions are plotted in Fig. 5, where a collapse is indeed observed for the three classes. The main differences are observed for short or long activities but this is expected since short paths and short travel times are certainly depressed for the extra-urban mobility and the slow mobility respectively. Furthermore the long path number is affected by the finite extension of the considered area. We also observe as the collapse of the average travel time distributions seems to hold for a larger interval, suggesting that travel time is the best candidate for the definition of a mobility cost. The results of Fig. 5 are consistent with our assumptions and they should explain the power law interpolation of the distributions shown in Fig. 2. A heuristic argument to derive a power law distribution is the following: if the probability



of observing a path of length L is the overlap of exponential functions one gets

$$\rho(L) \propto \sum_{k \geq 1}^{k_{\max}} \exp\left(-\frac{L}{L_k}\right), \tag{4}$$

where the exponential measures the cost of the path L when performed in the k -class and L_k^{-1} is the cost for unit length. If there exists a spatial multilayer structure in the system, there should exist a scaling law for the path cost of the different classes

$$L_k^{-1} = k^a L_{\max}^{-1} \quad a > 0$$

Table 2 Average values for the path length, duration, velocity distributions and sinuosity of the selected paths for the mobility classes 0,1,2 considering the OD mobility from hinterland to the coastal area

Class	$L_m(km)$	T_m (min.)	V_m (km/h)	Sinuosity
0	1.77	33.9	4.4	0.75
1	4.26	27.8	12.4	0.76
2	8.79	25.8	23.3	0.82

We remark that this scaling is certainly not valid for small L since one has also to consider the convenience to use a transportation network (for example the car is not convenient for short trips due to parking problem). Then, by approximating the sum with an integral it is possible to estimate the L -dependence in the validity interval of the distribution (4)

$$\rho(L) \propto \int_1^{k_{\max}} \exp\left(-k^a \frac{L}{L_{\max}}\right) dk = a \left(\frac{L}{L_{\max}}\right)^{-1/a} \int_{L/L_{\max}}^{u_{ast}} u^{(a-1)/a} \exp(-u) du,$$

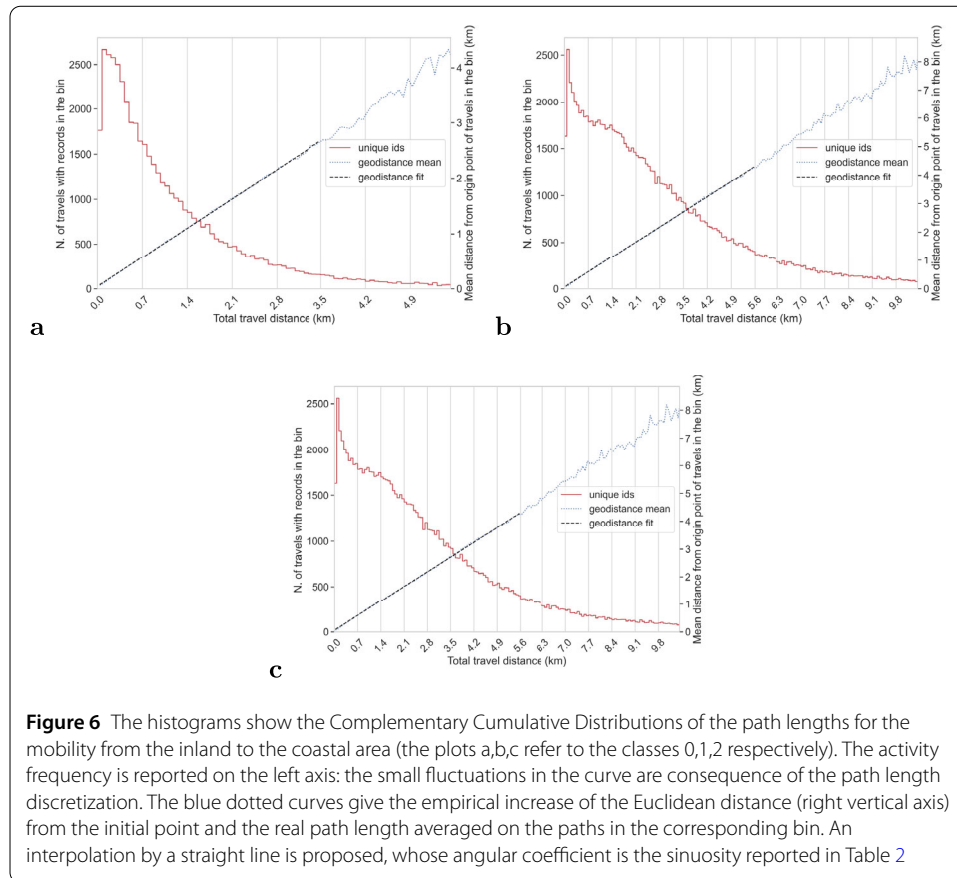
where the last integral is almost independent from L when $L_{\max} > L$. Finally we get

$$\rho(L) \propto \left(\frac{L}{L_{\max}}\right)^{-1/a} \quad (5)$$

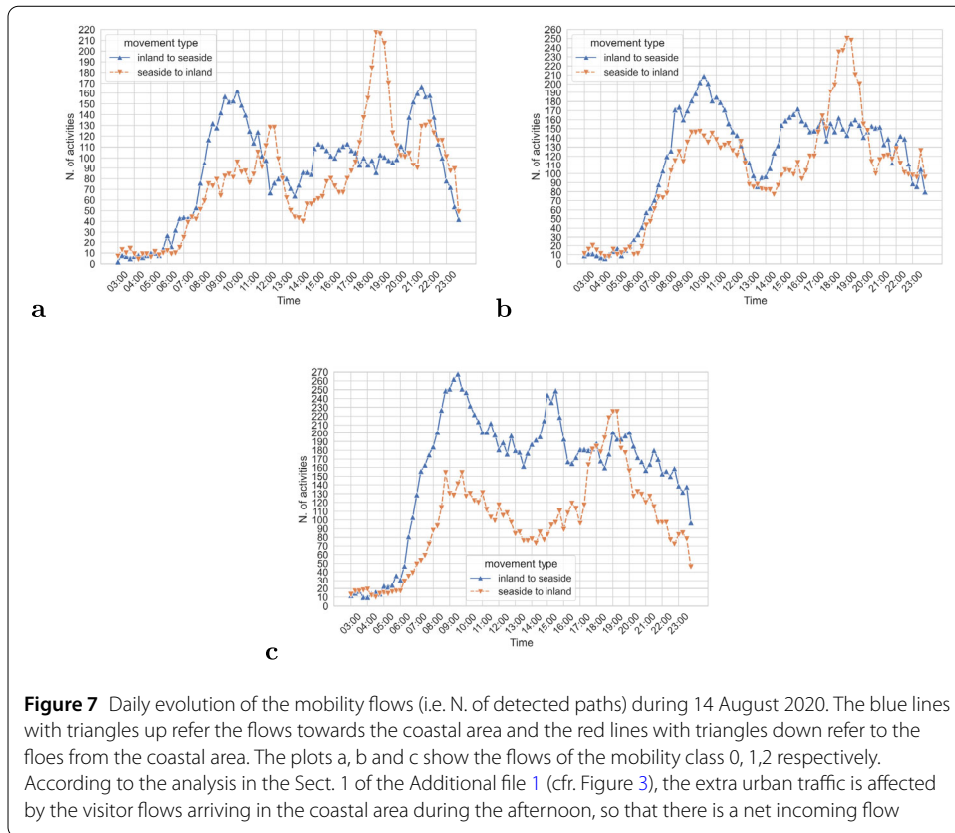
Comparing the empirical distribution shown in Fig. 2 we infer the exponent $a = 1$ for the path length distribution that means a relation among the average path lengths of the classes reported in Table 1 which would imply a strong dependence of the path cost on the mobility class. In the case of travel time distribution, where the power law interpolation is less effective, one estimates $a = .65$ that implies a weak dependence of the travel time cost from the mobility class. This is consistent with the limited changes of T_m in Table 1. However a more in-depth analysis is necessary to prove the previous heuristic argument. The previous result could be interpreted as a lacking of complexity in the human mobility, but the real complexity of the human mobility could emerge in not stationary situations or when critical states appears and there are necessary data sets able to detect small scale dynamical phenomena on large areas.

4 Study of a specific OD mobility demand

The Statistical Physics approach can be used to study the relevance of the detected mobility classes in the realization of a specific mobility demand. At this purpose we consider the paths related to the OD mobility between the inland and the coastal area where are located the main tourist attractions. An ideal boundary to distinguish the inland from the coastal area is the railway that crosses the center of Rimini (see Fig. 1), and we select the mobility paths with a starting point on one side and the ending point on the other side of the railway. In the Table 2 we report the average values for the path length, travel time, velocity and the sinuosity for these paths. A comparison with the values reported in Table 1 shows as the average length is slightly increased for all classes since the OD character of the mobility reduces the presence of the small paths, and the average travel time is almost double. There is a direct proportionality between the path length and the travel time increase for the slow mobility since the average velocity is not changed, whereas we have a reduction of the average velocity for the private car mobility, suggesting that the efficiency of this class



to perform the OD mobility is considerably reduced even if the sinuosity value of the traffic mobility decreases. To study the possible causes of the efficiency reduction for the car mobility, in Fig. 6 we plot the Complementary Cumulative Distribution (CCD) of the path lengths for the different mobility classes. The number of paths for the slow mobility and the urban mobility is approximately the same consistently with the percentage reported in Table 1, suggesting that the slow mobility plays a relevant role in the OD mobility from the Rimini historical center to the coast, despite of the increase of the average path length. In the case of the slow mobility (cfr. Figure 6a), we see an exponential like behavior in the CCD, so that the path length distribution is consistent with the corresponding distribution plotted in Fig. 5, but with an increased average value. Conversely, for the urban traffic (Fig. 6b), after the initial sharp peak, that is probably an artifact due to how people use the mobile phones, a flat behavior is observed for trip lengths ≤ 1 km followed by a decreasing that is approximately linear up to a distance $\simeq 4$ km. The initial flat part means that the cars are little used for short trips, whereas, assuming that the attractiveness of the coastal area is constant in the urban area, the decrease observed in the CCD of urban traffic can be consequence of the decreasing of the population density as one leaves the city and the distance from the coast increases. We remark that 4 km is the linear dimension of Rimini and, when one considers path lengths ≥ 4 km, the CCD slope decreases, suggesting that the corresponding paths could belong to the extra urban class. In the case of the extra urban mobility, the CCD (Fig. 6c) shows a linear decreasing from $l = 1$ km up to $l = 16$ km, after the initial peak of short activities. We also observe that number of



paths is of the same order as for the slow and urban mobility, i.e. almost a factor 2 lower than the percentage of the extra urban mobility class with respect to the total mobility (cfr. Table 1). This could be a consequence that the distance reduces the visitation frequency of the coastal area. According to the results in the paper [33], for a linear decreasing CCD, one expects that the product lf , where f is the frequency of the visit of the population, is constant. We observe that the considered area can be divided into two zones: the coastal zone where the main cities (i.e. Rimini and Riccione) are located and the hinterland where the population is distributed in the small towns near the main country roads (see Fig. 1). This can be also seen from the georeferenced distribution of the initial and ending points of the paths (see Fig. 8 in the Additional file 1). Therefore the population density can be assumed constant for the extra urban mobility and the visitation frequency towards the coastal zone decays $f \propto l^{-1}$.

Finally, we have studied how the role of the different mobility types changes during the day to try to highlight the individual mobility choices: in Fig. 7 we show the evolution of the daily mobility demand each 15 min. during 15 August 2020. We observe as the mobility evolution has the expected daily behavior for all the classes, with an incoming flow in the morning and return flow late in the afternoon. The contribution of the slow mobility is similar to that of the urban traffic, whereas for the extra urban mobility there is a net incoming flux. A possible explanation can be based on the analysis of the mobile phone presences in the ACE areas along the coast reported in Fig. 3 of the Additional file 1 where using the TIM-Olivetti data set it is possible to disaggregate the presence into four categories: the residents, the commuters, the visitors from the Emilia–Romagna region and from the other Italian regions. The data show a clear increase of the visitors from the

Emilia–Romagna that could be the cause of the difference between the incoming and the returning flows.

It is interesting to remark some differences among the classes that may highlight individual behaviors. In the urban traffic, we have exchange flows between the city center and the coastal area during the morning, with a net incoming flows due to the summer activities. A similar behavior is observed for the extra urban mobility, but the peaks are more pronounced early in the morning. The slow mobility has an incoming flow in the morning and a well defined peak just before lunch. This peak is not present for then other mobility classes, suggesting that the slow mobility is preferred when one has to cope with parking problems. In the evening we have a peak in all the mobility classes probably due to the people coming back from the beaches, but the slow mobility shows other two peaks after dinner, that can be related to evening activities and the closure to traffic of the coastal area.

5 Discussion

The use of GPS quality data sets to study the urban mobility offers new perspectives for the modeling of mobility at a microscopic (i.e individual) level, but it has to face the problem of the privacy laws so that it could be difficult and expensive to collect big data sets containing information on the individual mobility in all the European cities. A possible approach is to apply the statistical physics techniques to extract universal features of urban mobility that hold for any city. In this paper we take advantage from the MDT data set that contains the GPS positions of a sample of mobile phones collected in the Rimini province during the summer 2020 (7-17 August). Rimini is one of the most famous city in Italy for its tourist activities and even if some social restriction policies due to the COVID-19 epidemic were still enforced, there have been many visitors from Italy, as shown by the TIM-Olivetti data set on the mobile phone presences in the first section of the Additional file 1. The MDT data set allows to reconstruct the dynamics of individual mobility paths on the road network when the duration of the mobile phone activity is sufficiently long. By comparing the reconstructed hourly traffic flows using the MDT paths and the data recorded by magnetic coils on three country roads, the penetration of the MDT sample was estimated $\geq 1.5\%$, and the daily changes of the traffic flows were well reproduced on average (see Sect. 2 in the Additional file 1). We have verified that even in an urban context the path length distribution and the travel time distribution can be reproduced by a power law as proposed by various authors at larger spatial scales [14–18]. However, the MDT paths can be disaggregated into homogeneous classes applying FCM clustering algorithm. The features of each class and the identification of the underlying road sub-networks allow to associate the detected classes to different mobility types, defined as the slow mobility, the urban traffic, the extra-urban traffic and the highway traffic. The statistical distributions of MDT paths in each class turn out to be exponential and tend to collapse in a single curve if one normalizes the variables by their average values. Therefore, an application of the ME Principle can explain the exponential distributions by the existence of a *mobility energy* and it allows to understand the power law distributions as the consequence of the multilayer structure of the road network, and not of specific dynamical features of individuals performing urban mobility. Under this point of view, from one hand it is possible to highlight the relevance of the different mobility types in the area, but from the other hand one cannot infer the individual behavior from the statistical distributions and no other spatial or temporal structures emerge except the average values.

To detect some features of the individual behavior we have considered a specific OD mobility (from the inland to the coast and vice versa). Our results show that the slow mobility can play a relevant role even when the path length increases probably because pedestrian and cycle paths are present and traffic restriction rules are applied in the area. Moreover, for the extra-urban mobility the results are consistent with the existence of a gravitational-like law [33] for the attractiveness of the coastal zone, where most of the tourist activities are located. The hourly changes of the different mobility types highlight that the slow mobility is favored by the limited availability of parking places and by the existence of pedestrian zones, whereas the private car is mainly used during the rush hours in the morning and in the evening by commuters and incoming visitors.

Our results prove that the existence of GPS quality data sets (like the MDT data set) on individual mobility can highlight the multilayer structure of the road network involved in the mobility, that is a property of the considered urban context, but it is not sufficient to understand the dynamics of individuals to make predictions when changes occur. A possible approach is to study transient situations using the results of the Non-Equilibrium Statistical Physics [36]. This issue requires data sets able to detect individual dynamics with a sufficient statistical penetration. Only in this way the complexity science could provide useful tools to stakeholders to develop policies for the realization of a sustainable mobility [41].

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-023-00431-4>.

Additional file 1. An additional file is available illustrating the results of the mobile phone presences using the TIM-Olivetti data set, studying of the penetration of the MDT paths with respect the traffic flows, giving some details on the path reconstruction algorithm using the MDT data set and the application of the fuzzy c-means algorithm to detect the different mobility classes and providing complementary information to the results of the main text. (PDF 9.4 MB)

Acknowledgements

Not applicable.

Funding

The research activity has been partially supported by the *Fondazione Uni. Rimini* (https://www.unirimini.it/chi_siamo.php).

Abbreviations

ICT, Information and Communication Technologies; GPS, Global Positioning System; FCM, Fuzzy C-means Clustering; ME, Maximum Entropy; OD, Origin Destination; LTE, Long Term Evolution; CCD, Complementary Cumulative Distribution.

Availability of data and materials

The original data sets used in the paper are of public domain, but the aggregated data to perform statistical analysis are available on request from the corresponding author (A.B.). The fuzzy c-means clustering algorithm is in the Github repository <https://github.com/physcom/city-pro>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author contributions

DM, AV, CC, JS realized the data collection and made available the MDT data set. CM, AIB, AF proposed the methodologies and performed the data analysis. ArB conceived the work, supervised the data analysis and wrote the original draft. All authors have read and agreed to the published version of the manuscript.

Author details

¹Laboratory of City Science, Physics and Astronomy Department, University of Bologna, Bologna, Italy. ²INFN—sezione di Bologna, Bologna, Italy. ³Telecom-Italia SPA, Rome, Italy.

Received: 11 January 2023 Accepted: 25 November 2023 Published online: 08 December 2023

References

1. Batty M, Axhausen KW, Giannotti F et al (2012) Smart cities of the future. *Eur Phys J Spec Top* 214:481–518
2. Signorile P, Larosa V, Spiru A (2018) Mobility as a service: a new model for sustainable mobility in tourism. *Worldw Hosp Tour Themes* 10(2):185–200
3. Barbosa H, Barthelemy M, Ghoshal G, James RC, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Models and applications. *Phys Rep* 734:1–74
4. Scaloni A, Micheli D (2015) Estimation of mobility direction of a people flux by using a live 3G radio access network and smartphones in non-connected mode. In: *Proc. IEEE 15th Int. Conf. Environ. Elect. Eng.*, vol 1869
5. Micheli D, Diamanti R (2019) Statistical analysis of interference in a real LTE access network by massive collection of MDT radio measurement data from smartphones In: *Photonics & Electromagnetics Research Symposium—Spring (PIERS-Spring)*, vol 1906
6. Micheli D, Muratore G, Vannelli A et al (2021) Rain effect on 4G LTE in-car electromagnetic propagation analyzed through MDT radio data measurement reported by mobile phones. *IEEE Trans Antennas Propag* 69(12):8641
7. Mizzi C et al (2018) Unraveling pedestrian mobility on a road network using ICTs data during great tourist events. *EPJ Data Sci* 7:44
8. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding human mobility patterns. *Nature* 454:779–782
9. Giannotti F, Nanni M, Pedreschi D et al (2011) Unveiling the complexity of human mobility by querying and mining massive trajectory data. *VLDB J* 20:695
10. Deville P, Linard C, Martin S et al (2014) Dynamic population mapping using mobile phone data. *Proc Natl Acad Sci USA* 111(45):15888–15893
11. Alessandretti L, Aslak U, Lehmann S (2020) The scales of human mobility. *Nature* 587:402–407
12. Zhao C, Zeng A, Yeung CH (2021) Characteristics of human mobility patterns revealed by high-frequency cell-phone position data. *EPJ Data Sci* 10:5
13. Mizzi C, Fabbri A, Colomobini G et al (2022) A survival model to explain the statistical properties of multimodal mobility. *J Stat Mech* 2022:023404
14. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439:462465
15. Song C, Koren T, Barabasi AL (2010) Modelling the scaling properties of human mobility. *Nat Phys* 6:818–823
16. Yanqing H, Jiang Z, Zengru D (2011) Toward a general understanding of the scaling laws in human and animal mobility *EPL. Europhys Lett* 96(3):38006
17. Noulas A, Scellato S, Lambiotte R et al (2012) Tale of many cities: Universal patterns in human urban mobility. *PLoS ONE* 7:e37027
18. Yan XY, Wang WX, Gao ZY et al (2017) Universal model of individual and population mobility on diverse spatial scales. *Nat Commun* 8:1639
19. Bazzani A, Giorgini B, Rambaldi S et al (2010) Statistical laws in urban mobility from microscopic GPS data in the area of Florence. *Journal of Statistical Mechanics: Theory and Experiment* 2010:P05001
20. Gallotti R, Bazzani A, Rambaldi S (2012) Towards a statistical physics of human mobility. *Int J Mod Phys C* 23(09):1–16
21. Liang X, Zheng X, Lv W (2012) The scaling of human mobility by taxis is exponential. *Phys A, Stat Mech Appl* 391(5):2135–2144
22. Kolbl R, Helbing D (2003) Energy laws in human travel behaviour. *New J Phys* 5:48
23. Alessandretti L, Sapiezynski P, Sekara V et al (2018) Evidence for a conserved quantity in human mobility. *Nat Hum Behav* 2:485–491
24. Wang BH, Wang XW, Han XP (2014) Correlations and scaling laws in human mobility. *PLoS ONE* 9(1):e84954
25. Liang X, Zhao J, Dong L et al (2013) Unraveling the origin of exponential law in intra-urban human mobility. *Sci Rep* 3:2983
26. Gallotti R, Bazzani A, Rambaldi S (2015) Understanding the variability of daily travel-time expenditures using GPS trajectory data. *EPJ Data Sci* 4:18
27. Gallotti R, Barthelemy M (2015) The multilayer temporal network of public transport in Great Britain. *Sci Data* 2:140056
28. Ding R, Ujang N, Bin Hamid H et al (2018) Detecting the urban traffic network structure dynamics through the growth and analysis of multi-layer networks. *Phys A, Stat Mech Appl* 503:800–817
29. Louail T, Lenormand M, Picornell M et al (2015) Uncovering the spatial structure of mobility networks. *Nat Commun* 6:6007
30. Gallotti R, Barthelemy M (2014) Anatomy and efficiency of urban multimodal mobility. *Sci Rep* 4:6911
31. Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. *Comput Geosci* 10(2–3):191
32. Jaynes ET (1957) Information Theory and Statistical Mechanics. *Phys Rev* 106(4):620–630
33. Schläpfer M, Dong L, O’Keeffe K et al (2021) The universal visitation law of human mobility. *Nature* 593:522
34. Homepage of the OpenStreetMap project (2021) <https://www.openstreetmap.org/>
35. Gheorghiu S, Coppens MO (2004) Heterogeneity explains features of “anomalous” thermodynamics and statistics. *Proc Natl Acad Sci USA* 101(45):15852–15856
36. Seifert U (2008) Stochastic thermodynamics: principles and perspectives. *Eur Phys J B* 64:423–431
37. Github repository. <https://github.com/physycom/city-pro>
38. Marchetti C (1994) Anthropological invariants in travel behavior. *Technol Forecast Soc Change* 47(1):75
39. Kalapala V, Sarwalani V, Clauset A, Cristopher Moore C (2006) Scale invariance in road networks. *Phys Rev E* 73:026130
40. Gallotti R, Bazzani A, Rambaldi S et al (2016) A stochastic model of randomly accelerated walkers for human mobility. *Nat Commun* 7:12600
41. Tran M, Draeger C (2021) A data-driven complex network approach for planning sustainable and inclusive urban mobility hubs and services. *Environ Plan B: Urban Anal City Sci* 48(9):2726–2742

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.