REGULAR ARTICLE   **Open Access**

CrossMark

Unraveling pedestrian mobility on a road network using ICTs data during great tourist events

Chiara Mizzi[1], Alessandro Fabbri[1], Sandro Rambaldi[1], Flavio Bertini[1], Nico Curti[1], Stefano Sinigardi[1], Rachele Luzi[1], Giulia Venturi[1], Micheli Davide[2], Giuliano Muratore[2], Aldo Vannelli[2] and Armando Bazzani[1*] (iD)

*Correspondence:
armando.bazzani@unibo.it
[1]Physics and Astronomy
Department, University of Bologna,
Bologna, Italy
Full list of author information is
available at the end of the article

Abstract

Tourist flows in historical cities are continuously growing in a globalized world and adequate governance processes, politics and tools are necessary in order to reduce impacts on the urban livability and to guarantee the preservation of cultural heritage. The ICTs offer the possibility of collecting large amount of data that can point out and quantify some statistical and dynamic properties of human mobility emerging from the individual behavior and referring to a whole road network. In this paper we analyze a new dataset that has been collected by the Italian mobile phone company TIM, which contains the GPS positions of a relevant sample of mobile devices when they actively connected to the cell phone network. Our aim is to propose innovative tools allowing to study properties of pedestrian mobility on the whole road network. Venice is a paradigmatic example for the impact of tourist flows on the resident life quality and on the preservation of cultural heritage. The GPS data provide anonymized georeferenced information on the displacements of the devices. After a filtering procedure, we develop specific algorithms able to reconstruct the daily mobility paths on the whole Venice road network. The statistical analysis of the mobility paths suggests the existence of a travel time budget for the mobility and points out the role of the rest times in the empirical relation between the mobility time and the corresponding path length. We succeed to highlight two connected mobility subnetworks extracted from the whole road network, that are able to explain the majority of the observed mobility. Our approach shows the existence of characteristic mobility paths in Venice for the tourists and for the residents. Moreover the data analysis highlights the different mobility features of the considered case studies and it allows to detect the mobility paths associated to different points of interest. Finally we have disaggregated the Italian and foreigner categories to study their different mobility behaviors.

PACS Codes: 89.75.-k; 89.65.-s; 89.40.-a

Keywords: GPS data; Big tourist events; Pedestrian mobility; Statistical physics of human mobility

1 Introduction

The fast development of Information Communication Technologies (ICT) offers new opportunities for the realization of innovative analytical tools in the framework of big data analytics, which is one of the future challenges of the Complexity Science [1, 2]. Various authors have considered the large georereferenced datasets on individuals mobility, studying the statistical laws at the base of human movements [3–8] and the dynamic properties of human mobility in urban contexts [9, 10]. The aim of this research activity in the framework of Complex Systems Physics is to provide to stakeholders new knowledge tools to improve the sustainability of the mobility demand in future cities [8, 11–13]. The use of ICT datasets to study the human mobility is considered part of the road-map toward the realization of smart cities [2]. The big data science has certainly provided new tools to cope with complex problems of modern cities [14], however it has some intrinsic criticalities, that have given rise to a debate on the possibility of realizing smart cities [15]. The main questions are how to control the lack of information in the data and the representativeness of the datasets, that are strictly related to the use of new technologies. On one hand, data analytics is developing new statistical methods to extract the relevant information from large data sets [16]. On the other hand the possibility of using different data sources and the fast spreading of ICT in the population could reduce the bias in the data sample.

The governance of the mobility demand generated by big tourist flows is becoming a key issue for the quality of life in the historical Italian cities, that will become even worse in the next future due to globalization processes. On one hand the frailty of the cultural heritage is incompatible with presence of big tourist flows, on the other hand the daily life of residents is heavily conditioned by the presence of tourists. However the relevance of tourism economy advises against restriction policies that would limited a priori the tourists number. The data collection on individual mobility is preliminary to the development of any dynamic model which simulates and, hopefully, forecasts the pedestrian flows. Moreover during big tourist events, the critical crowding conditions require a specific analysis to point out the spatial and dynamic features of the observed mobility in relation with the structure of the road network and attractiveness of the points of interest [17, 18]. The individual behavior is a key issue to understand the emergent properties of crowd dynamics [19]. The historical centre of Venice is a paradigmatic case study, both for the predominantly pedestrian character of the venetian mobility and the unique features of its monuments, that attract large crowds of visitors during all the year. The historical city of Venice has a surface of 6.7 km^2 (see Additional file 1) and $\simeq$55,000 inhabitants. This value can grow up to double during big tourist events. In this paper we cope with the problem of understanding how the pedestrian flows moved on the road network of the Venice historic centre during the Carnival of Venice 2017 (from 23/2/2017 up to 02/03/2017) and during the *Festa del Redentore* (from 14/7/2017 up to 16/7/2017). Our approach is based on two datasets provided by the Italian mobile phone company TIM [20] containing GPS (Global Positioning System) data about a relevant sample of mobile devices, with a ID number which changes every 24 hours. The collection of GPS data is possible by means of the new technologies that are currently being developed by NOKIA (Geosynthesis system) and the data provide anonymized GPS positions of a device each time certain types of network activities are on. We have introduced restrictive conditions to identify an individual mobility path to reduce the errors due to lacking of GPS data when the mobile device is in an idle condition ,and we have considered the problem of the
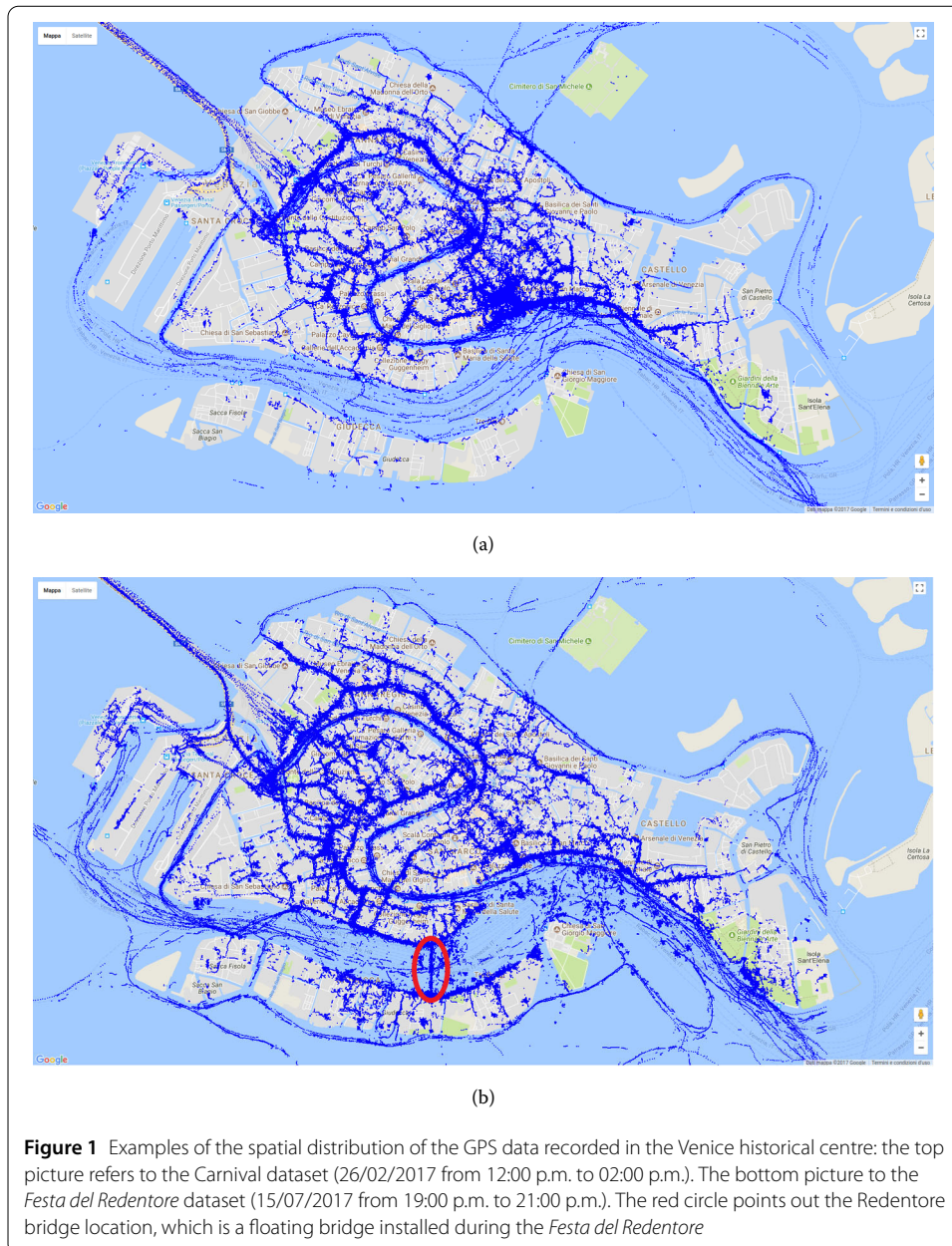
representativeness of our data sample performing a direct measure of the pedestrian flow on a bridge. Even if we cannot achieve a final answer to this problem we are confident that the expected diffusion of the ICT will improve the quality of the GPS datasets collected using mobile devices. The choice of studying the mobility during big tourist events was made for two reasons: on one hand we take advantage from the presence of many people to increase the penetration of the mobile device sample to reconstruct pedestrian mobility, on the other hand there is a specific request to study the venetian mobility during such events, since the municipality has proposed to limit the tourist presence. Moreover, at the moment, the development of counting systems for measuring the tourist flows in Venice is under discussion and the actual numbers are estimated using average data from transportation means. We are aware that an exhaustive understanding of the pedestrian mobility in Venice certainly requires further studies, which consider a long period of data collection (not available at the moment), but in this paper we limit ourselves to face with the problem of how to extract relevant information on pedestrian mobility from the GPS datasets. Both the chosen events attract big crowds of tourists, but they present different features besides the fact that the Carnival takes place in winter and the *Festa del Redentore* in summer, during the evening of 15 July: Carnival is a typical tourist festival with several scheduled events distributed throughout the city (even if the main attractions are in San Marco square), whereas *Festa del Redentore* is a religious festivity very important to the Venetians which attracts many people arriving from the Venice district to attend to the fireworks along the Giudecca Canal. For these reasons we expect differences in the observed mobility in the two case studies, that have to be pointed out by our analysis. The distribution of devices detected by the phone cells network has been used to measure the spatial activity patterns [21–24] or to estimate the evolution of crowding into different areas of a city [25–27]. The study of the mobility through the reconstruction of the device trajectories on a road network, requires a precision of few meters in the device location that is characteristic of GPS data. In previous works [6, 28] we have studied the private vehicle mobility on urban road networks using GPS dataset recorded for insurance reasons that contains information on the vehicle trajectories at a scale of 1 km or 30 sec, on a sample of $\simeq 3\%$ of the Italian vehicle population. In this work we apply the same methodologies to the GPS data recorded from mobile devices. After a preliminary data analysis to select the devices that have provided a suitable amount of GPS data, our approach is based on algorithms able to associate a daily mobility path to each device. The main difficulties are the occasional character of the mobile device activities that prevent the data collection at a fixed spatial scale and the signal losses mainly due to the narrow roads in Venice. To avoid the possible introduction of biases in our analysis, we prefer to follow a big data approach reducing drastically the numerousness of device sample and only reconstructing the mobility paths that satisfy well defined reliability criteria. As a consequence our approach is not able to detect critical crowding situations localized on the road network, but it succeeds to highlight the dynamic features of pedestrian mobility during the considered events. In particular the presented results refer to days of 26/02/2017 (Carnival Sunday) and 15/07/2017 (Redentore day) during which the presence of tourists was particularly relevant. We have then checked the penetration of the sample by comparing the estimated pedestrian flows aggregated at each hour, with a direct measure performed by volunteers on the *Redentore* bridge, that is crossed by a large amount of people due to the presence of fireworks show on the Giudecca Canal.

The main results of the paper is the emergence of a diffusion-like relation between the covered distance and the elapsed time $s \propto t^{\alpha}$ with $\alpha \simeq 1/2$ and the existence of preferred mobility connected subnetworks of the whole road network able to take into account the majority of the observed mobility [29]. In the first case we suggest the existence of a travel time budget [30, 31] for the pedestrian mobility in Venice and we introduce the concept of rest times during the individual mobility, that could play an important role in the construction of dynamic models for tourist flows. In the second case our results highlight that the existence of mobility subnetworks can simplify the monitoring and controlling problem of the tourist flows and help the definition of models. Thanks to the information in the datasets we can also distinguish between Italian and foreign visitors and point out the existence of different mobility paths for the two categories.

The paper is organized as follows: in the second section we describe the main features of the datasets and we give an estimate of the sample penetration; in the third section we describe the algorithms to reconstruct the mobility paths and we perform the study of statistical properties of the observed mobility during the considered events; in fourth section we highlight the mobility connected subnetworks that emerge from the aggregation of the mobility paths and we discuss the difference in the mobility paths between Italians and foreigners and in the mobility driven by different attraction points; the conclusive remarks are reported in the last section.

2 The datasets

The dataset used in this study has been provided by the Italian mobile phone company TIM and contains georeferenced positions of tens of thousands anonymous devices (e.g. mobile phones, tablets, etc. . . .), whenever they performed an activity (e.g. a phone call or an internet access) during eight days from 23/2/2017 up to 02/03/2017 (Carnival of Venice dataset), and from 14/7/2017 up to 16/7/2017 (*Festa del Redentore* dataset). According to statistical data, 66% of the whole Italian population has a smartphone [32] and TIM is one the greatest mobile phone company in Italy whose users are $\simeq$30% of the whole smartphone population. The datasets refer to a geographical region that includes an area of the Venice province, so that it is possible to distinguish commuters from sedentary people and the different transportation means used to reach Venice. Each valid record gives information about the GPS localization of the device, the recording time, the signal quality and also the roaming status, which in turns allow to distinguish between Italian and foreigners. More details on the dataset collection techniques are reported in Additional file 1. The devices are fully anonymized and not reversible identification numbers (ID) are automatically provided by the system for mobile phones and calls within the scope of the trial; the ID is kept for a period of 24 hours. During each activity a sequence of GPS data is recorded with a 2 sec. sampling rate and the collection stops when the activity ends. As matter of fact during an activity most of people reduce their mobility except if they are on a transportation mean, so that the dataset contains a lot of small trajectories that have to be joined to reconstruct the daily mobility. After a filtering procedure (see next subsection) these data provide information on the mobility of a sample containing 3000–4000 devices per day. Since the presences during the considered events were of the order of 10^5 individuals per day, as reported by the local newspapers [33] , we estimate an overall penetration of our sample of 3–4%. Figure 1 shows an example of the distribution of the GPS data recorded in the Venice historical centre. In the sequel we illustrate in details

(a)

(b)

Figure 1 Examples of the spatial distribution of the GPS data recorded in the Venice historical centre: the top picture refers to the Carnival dataset (26/02/2017 from 12:00 p.m. to 02:00 p.m.). The bottom picture to the *Festa del Redentore* dataset (15/07/2017 from 19:00 p.m. to 21:00 p.m.). The red circle points out the Redentore bridge location, which is a floating bridge installed during the *Festa del Redentore*

the results of our approach for the Sunday 26/2/2017 during Carnival and for Saturday 15/2/2017 during the *Festa del Redentore* that were particularly crowded days.
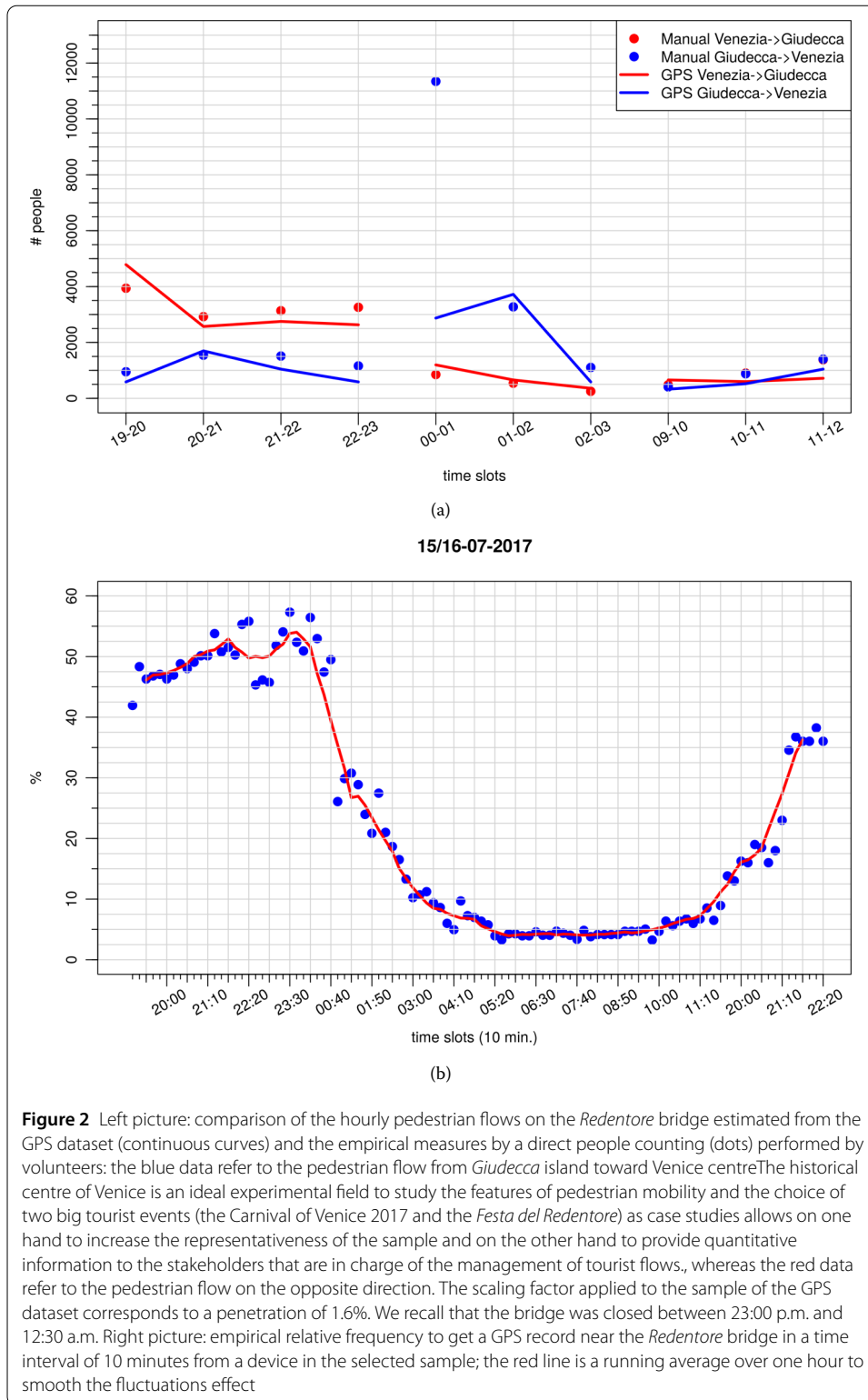
2.1 Data filtering procedure

We perform a filtering process on the datasets to select the devices that give information suitable to study the daily mobility on the Venice road network. We have extracted the Venice road network from Open Street Map database [34] using a filtering procedure to neglect small open arcs and a fusion procedure to join consecutive arcs. Moreover we have added the ferryboat lines to georeference correctly people in the public means. We have compared the extracted road network with the official cartography of Venice municipality (http://smu.insula.it/ Ramses project). The Carnival and the *Festa del Redentore* datasets

contain respectively $\simeq 10^6$ and $\simeq 1.8 \times 10^6$ georeferenced records in the Venice historic centre. We aggregate the GPS data of each device-ID to downsample the data by starting from an initial position (pivot point) and by computing the geodesic distance on the road network with the successive points associated to the same ID. When the distance overcomes a fixed threshold (we choose a threshold of 50 m) we keep the new point and restart the procedure using the new point as pivot. In this way the number of valid positions is reduced respectively to $\simeq 60 \times 10^3$ in the Carnival dataset and to $\simeq 90 \times 10^3$ in the *Festa del Redentore* dataset. Each selected GPS point is then located in the nearest arc of the road network within a maximal distance of 60 m including the ferryboat lines; points that cannot be attributed to any arc according to this criterion, are discarded. The positioning procedure further reduces the valid points down to $\simeq 50 \times 10^3$ in the Carnival dataset and down to $\simeq 80 \times 10^3$ in the *Festa del Redentore* dataset. These positions allow to get dynamic information both on the most used paths on the road network, and we have a measure of the elapsed time between two successive positions, that could point out the main points of interest.

2.2 Sample penetration estimate

We have performed a direct check for the representativity of the considered sample on the spatial scale of a single road. In particular, we compare the pedestrian flows estimated using GPS data with the pedestrian flows directly measured by volunteers on the *Redentore* bridge. The campaign of measures was organized by CORILA [35] and the data were collected each 15 minutes by using people count devices. The *Redentore* bridge is a floating bridge on the Giudecca Canal (see Fig. 1 and the map in Additional file 1). The bridge has a length of $\simeq 300$ m and it was opened from 7:00 p.m. of 15/07/2017 for all the night, except during the firework show between 23:00 p.m. and 12:30 a.m. To estimate the pedestrian flow across the bridge we have counted the mobile devices that leave two GPS signals at opposite sides of the bridge during the considered time interval slot, so that we distinguish between the two crossing directions. The results of the direct measures are reported in Fig. 2 (above picture) together with the estimated pedestrian flows scaled according to a penetration of $\simeq 1.6\%$ for our sample. This result is obtained by means of a best fit of the direct measures with 20% average error (excluding the flow measured at the reopening of the bridge after midnight). The reduced sample penetration with respect to the 5% expected, is probably due to the small spatial scale of the bridge that requires a coincidence of two GPS signals from the same device at the opposite sides of the bridge in a short time interval. Indeed, we expect that the variability of the device activity rate reduces the sample penetration as it is shown in Fig. 2 (bottom picture), where we have computed the probability that a device located in an area near the bridge leaves GPS signal in a time interval of 10 minutes. We remark as the activity rate of the devices changes drastically form 23:30 p.m. to 12:30 a.m. The estimated flows allow to reproduce with good accuracy the evolution of the empirical observations except for a single point between midnight and 01:00 a.m. when the bridge was reopened after the firework show. A big pedestrian flow was recorded between 12:30 a.m. and 01:00 a.m. that is not detected by the GPS dataset. A possible explanation is that the mobile devices activity in the area is dropped down after the fireworks (cfr. Fig. 2 (bottom picture)). Probably, after the firework show, most of people were mainly interested in crossing the bridge towards the Venice centre. The direct people counting points out a net pedestrian flow towards the Giudecca island of 8000

15/16-07-2017

(a)

(b)

Figure 2 Left picture: comparison of the hourly pedestrian flows on the *Redentore* bridge estimated from the GPS dataset (continuous curves) and the empirical measures by a direct people counting (dots) performed by volunteers: the blue data refer to the pedestrian flow from *Giudecca* island toward Venice centreThe historical centre of Venice is an ideal experimental field to study the features of pedestrian mobility and the choice of two big tourist events (the Carnival of Venice 2017 and the *Festa del Redentore*) as case studies allows on one hand to increase the representativeness of the sample and on the other hand to provide quantitative information to the stakeholders that are in charge of the management of tourist flows., whereas the red data refer to the pedestrian flow on the opposite direction. The scaling factor applied to the sample of the GPS dataset corresponds to a penetration of 1.6%. We recall that the bridge was closed between 23:00 p.m. and 12:30 a.m. Right picture: empirical relative frequency to get a GPS record near the *Redentore* bridge in a time interval of 10 minutes from a device in the selected sample; the red line is a running average over one hour to smooth the fluctuations effect

people during the opening of the bridge up to 23:00 p.m. and a net flow of 14,000 people in the opposite direction, after the bridge reopening (some people arrived the island by ferryboat). The GPS dataset estimates correctly the incoming flow, but underestimates the
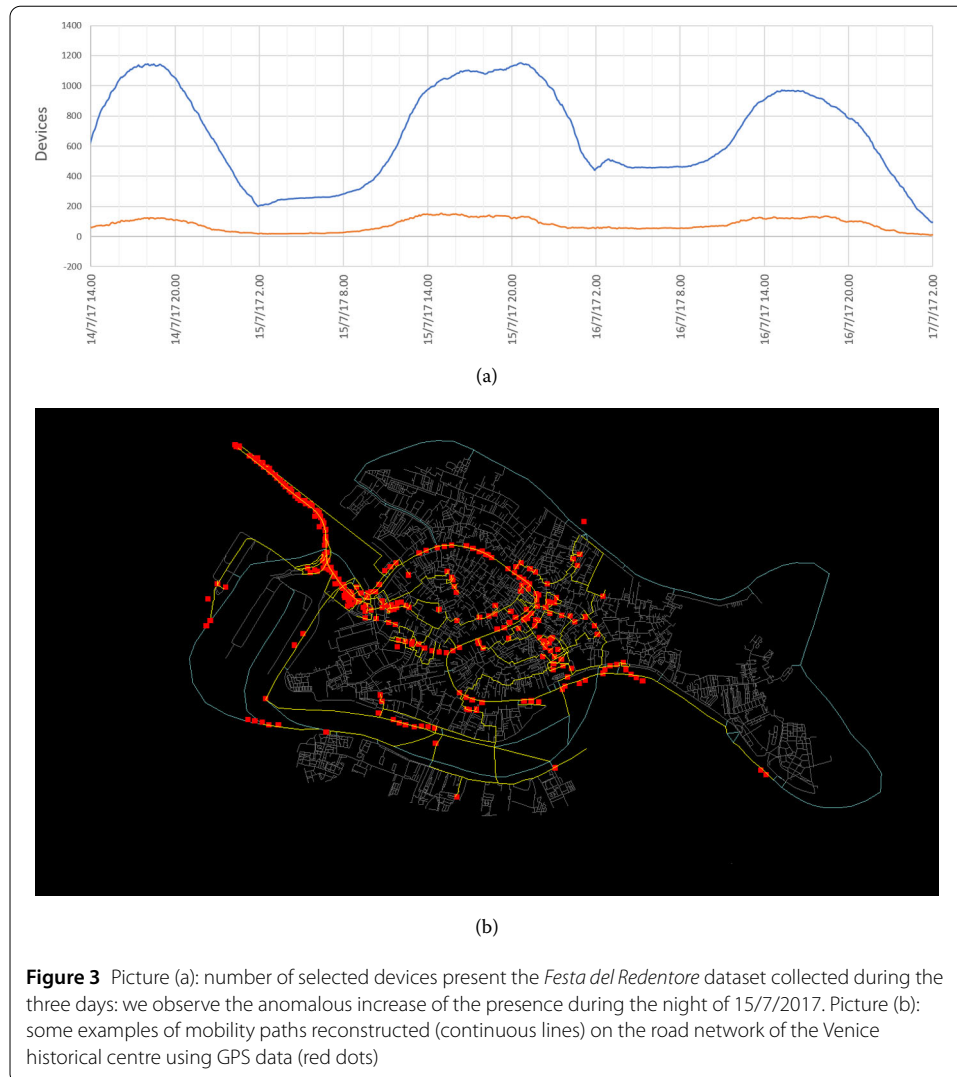
outgoing flow with an error of approximately 8000 people. This estimate can be consistent if the device activity at the bridge were reduced by a factor 3 in the time interval from 12:30 a.m. to 01:00 a.m. The comparison with empirical observations suggests that the selected device sample recovers its representativity during the night. On our opinion, the fact that the selected sample could fail to detect localized crowded situations can be the consequence of two causes. On one hand we have selected the device sample maximizing the possibility to reconstruct the daily mobility on the road network and not to detect crowded situations. On the other hand ,since the GPS data are only recorded when the device performs an activity, there are necessary further studies to understand how people use ICT devices in crowded situations.

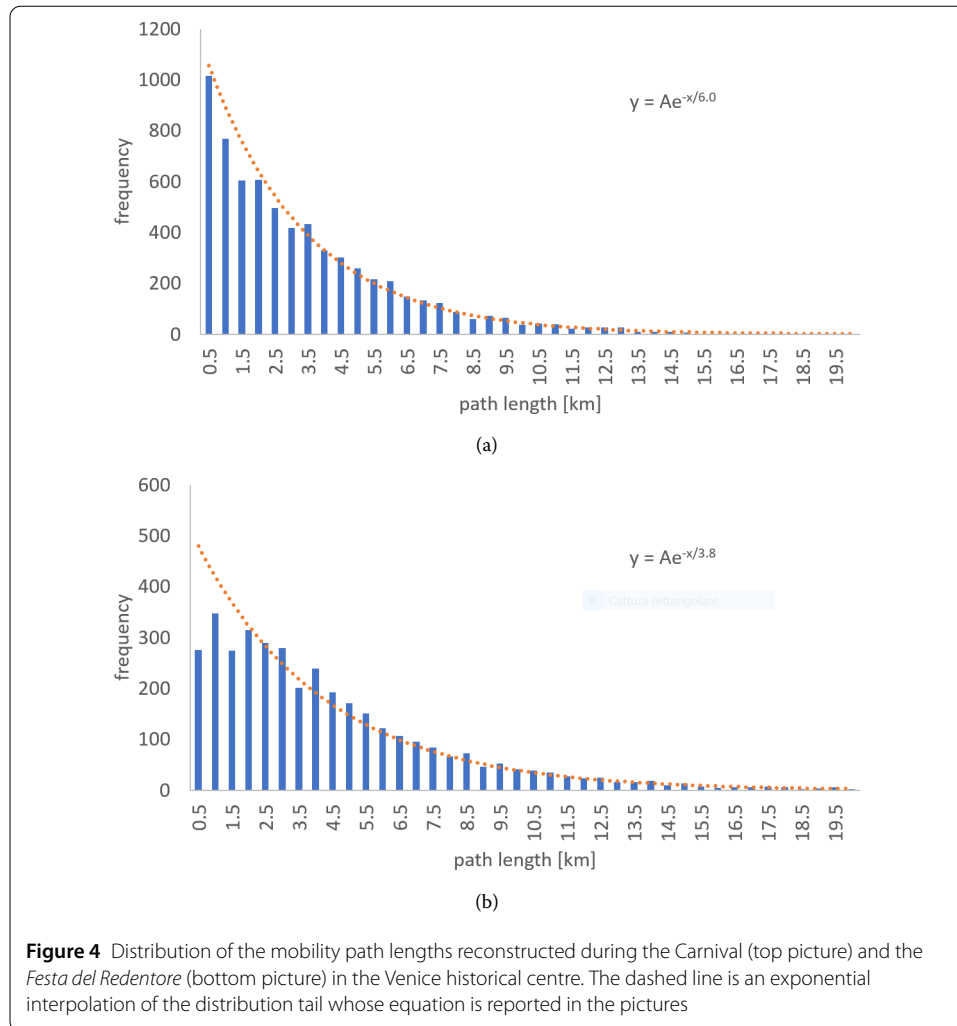## 3 Mobility paths reconstruction on the road network

The procedure of mobility path reconstruction considers separately the land mobility and the water mobility since the two mobility networks have different features, so that it is necessary to check carefully the transitions from one network to the other. To create a mobility path, we connect two successive points left by the same device using a best path algorithm on the road network with a check on the estimated travel speed to avoid unphysical situations and discarding the paths whose velocity is clearly not consistent with the typical pedestrian velocity (or ferryboat velocity). To end a land path and to start a water path, we require that at least two successive points of the same device are attributed to a ferryboat line by the localization algorithm. In the case of a single point on a ferryboat line, we force the localization of this point on the nearest road on the land. An example of daily mobility paths are shown in Fig. 3 (bottom). This procedure allows to reconstruct the daily mobility of $\simeq$4000 devices for the Carnival dataset and 5000 devices for the *Festa del Redentore* dataset. However some devises leave a very low number of points (less than 3) that are not enough to study their mobility, and other devices show an anomalous mobility paths which crosses a very high number of roads (more than 200). In such a case we consider outliers these paths, that could be associated to people performing particular activities in Venice, which are not related to the tourist or citizen mobility. Finally, we succeed to reconstruct the daily mobility of $\simeq$2800 (resp. $\simeq$3600) different devices per day for the Carnival dataset (resp. for the *Festa del Redentore* dataset), so that the representativeness of the mobility sample is estimated between 2.8 $\div$ 3.6%. In Fig. 3 (top) we show the measured number of moving devices detected in the historical centre of Venice, whose mobility paths have been correctly reconstructed by the algorithms during the *Festa del Redentore*: the figure refers both to the land and water mobility and clearly shows the circadian rhythm of the presences with a peak during the evening of 15/7/2017 in occasion of the firework show.

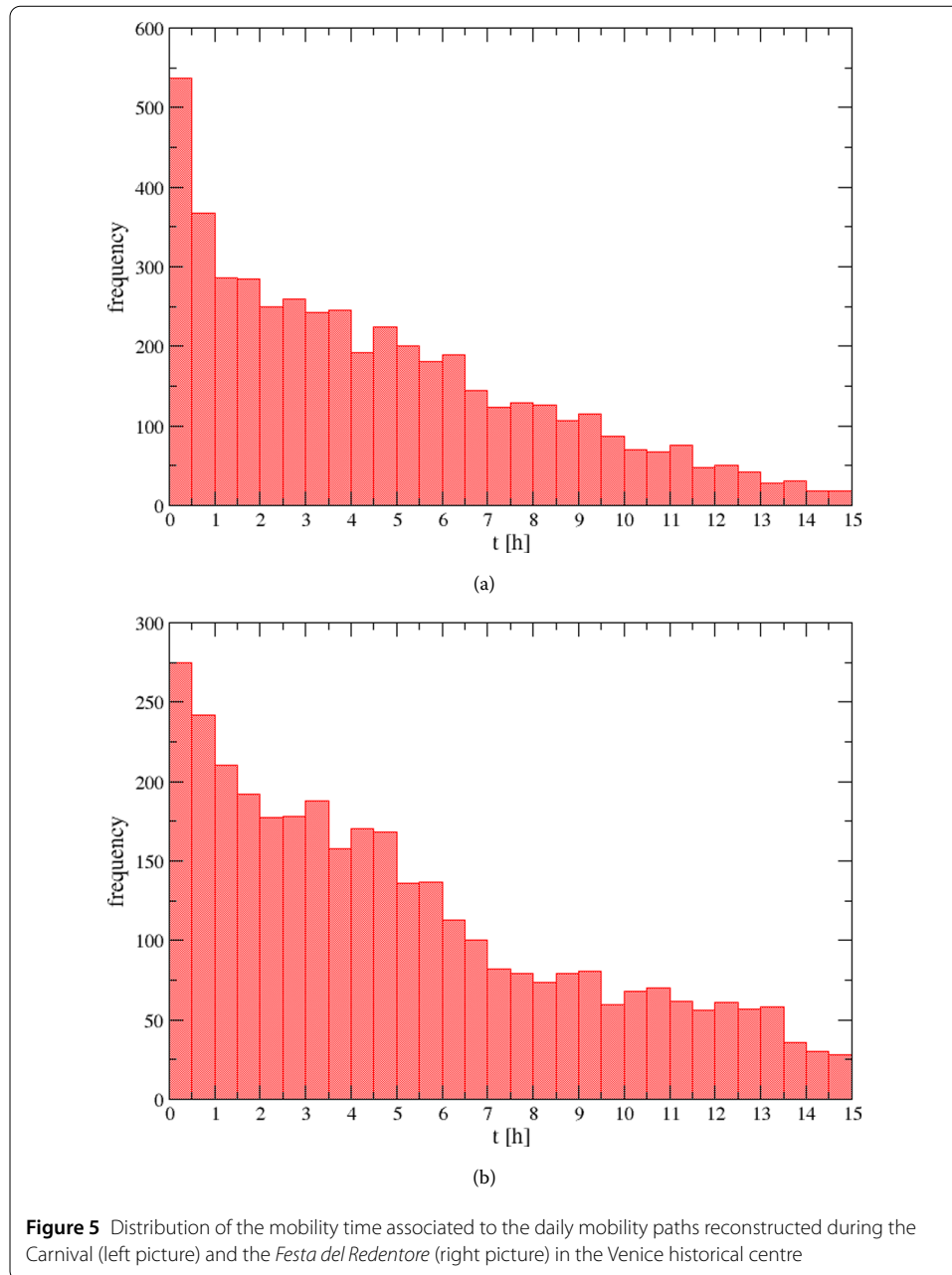### 3.1 Statistical properties of mobility paths

The mobility paths provide dynamic information on how people realize their mobility demand on the road network during the considered events. The elapsed time between two successive GPS data is used to attribute a displacement velocity that of course is affected by the rest times at any point of interest. We remark that we have not a start and end point of each single trip, but only a sampling of the whole daily mobility of a device, since the GPS data are recorded only in conjunction with an activity: for example the elapsed time between two successive points may be affected by a stop for shopping. A dynamic model

Figure 3 Picture (a): number of selected devices present the *Festa del Redentore* dataset collected during the three days: we observe the anomalous increase of the presence during the night of 15/7/2017. Picture (b): some examples of mobility paths reconstructed (continuous lines) on the road network of the Venice historical centre using GPS data (red dots)

to simulate the pedestrian dynamics on the Venice road network based on the individual dynamics has to includes tracts covered at constant velocity and breaks due to the presence of points of interests, crowded situations or to recover from the walking fatigue. We consider some statistical properties of the reconstructed mobility paths to check if they are consistent with other statistical laws suggested by the analysis of mobility datasets in urban contexts [4, 6, 9, 10]. In Fig. 4 we report the daily path length distribution for both the considered datasets: the average mobility lengths are 3.1 km and 4.3 km respectively for the Carnival and the *Festa del Redentore* datasets. The differences between the two distributions may be explained both by the effect of weather conditions [36] (the Carnival takes place on winter whereas the *Festa del Redentore* is celebrated during summer) and by the different organization of the two events. The Venice Carnival is an ensemble of events spread on the historical centre even if San Marco square is always the main attractive location, whereas the *Festa del Redentore* is celebrated in the area near *Giudecca* Canal between the Giudecca island and the *Riva degli Schiavoni*. Therefore one expects a mobility more influenced by an origin destination character during the *Festa del Redentore* than during the Carnival of Venice. The path distribution in Fig. 4 refers only to pedestrian

Figure 4 Distribution of the mobility path lengths reconstructed during the Carnival (top picture) and the *Festa del Redentore* (bottom picture) in the Venice historical centre. The dashed line is an exponential interpolation of the distribution tail whose equation is reported in the pictures

mobility since we have excluded all the mobility paths with a tract on a ferry line. This criterion is satisfied by 2/3 of the devices in our sample, whereas the remaining 1/3 performs a mixed mobility. We propose an exponential interpolation of the path length distribution for both the datasets (cfr. dashed lines in Fig. 4) and we observe as the exponential interpolation overestimates the short paths in the *Festa del Redentore* according to the existence of a great origin destination component. Assuming the existence of an average characteristic pedestrian velocity, the path length can be interpreted as a *mobility energy* distribution in agreement with a Maxwell-Boltzmann distribution [6] and it is consistent with the concept of travel time budget proposed in other studies of urban mobility [30, 31]. The exponential decaying defines two different characteristic lengths, 3.0 km for the Carnival dataset and 3.8 km for the *Festa del Redentore* dataset and it suggests the propensity of the individuals to perform a greater mobility in the last case. In both cases these distances are probably greater than the typical pedestrian mobility in a city, but they reflect the average walking distance in the historical centre of Venice, where the pedestrian mobility is prevalent. Short mobility paths are overestimated by the exponential distribution since one has to cover a minimal distance to satisfy the mobility demand. The presence

Figure 5 Distribution of the mobility time associated to the daily mobility paths reconstructed during the Carnival (left picture) and the *Festa del Redentore* (right picture) in the Venice historical centre

of short daily mobility paths could also be related to the use of the public transportation system.

To understand the statistical features of the observed mobility we also consider the mobility time distribution associated to the mobility paths, computed as the elapsed time between the first and the last recorded GPS position of a device in the area of interest (see Fig. 5). The mobility time is the sum of the travel times and the rest times. The distribution tail can be affected by the device activity during the night not directly related to the mobility. It is a reasonable assumption that if an individual has spent more than 8 h in Venice then he has a relevant probability to spend also the night in Venice: indeed to spend more than 8 h in Venice living outside, one has to add a commuting time between 1 and 2 h and

to consider the possibility to take lunch and dinner in Venice, that could be quite expensive. The exponential interpolation is less justified in this case due to the increased effect of the rest times with respect to the mobility times, and we derive a dynamic model for the relation between the mobility path lengths and the mobility time.

3.2 Dynamic properties of the mobility paths

Let us consider an ensemble of individual moving on the road network, we define the average moving velocity $v(t)$

$$\frac{d\langle s \rangle}{dt} = v(t),$$

where $\langle s \rangle$ is the average path length corresponding to a mobility time t. In Fig. 6 we show the result of an interpolation of the empirical relation between $\langle s \rangle$ and t by means of a power law
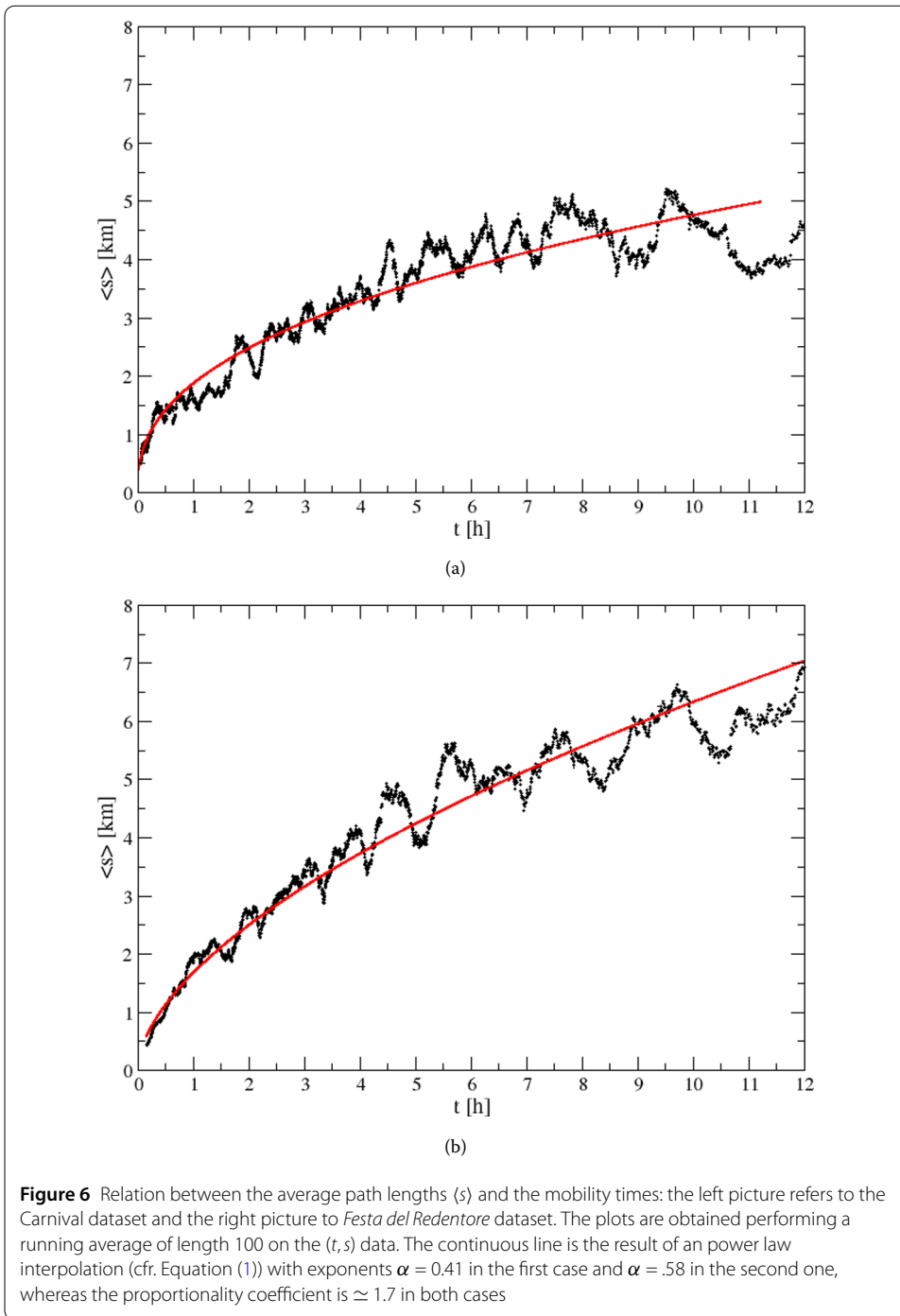
$$\langle s \rangle = ct^{\alpha}, \tag{1}$$

where c is a suitable constant. In normal conditions, the pedestrian dynamics is performed at a constant velocity v_0, with a stochastic variation among individuals, and a linear relation $s = v_0 t_w$ is expected where t_w is the walking time. The statistical law $\langle s \rangle \propto t^{\alpha}$ with $\alpha < 1$, where we average on the path lengths corresponding to a given mobility time t implies that the rest times, defined by the difference $t - t_w$, increase as a function of t. Therefore the relation (1) simulates a fatigue effect of individuals during pedestrian mobility. We remark that it is difficult to relate this effect to crowding conditions in the road network unless one could compute a fundamental diagram [37] for the pedestrian dynamics in the Venice road network. On our opinion this is possible, but it requires a dataset that includes a long period of observations. The interpolation of the empirical data gives an exponent $\alpha = 0.41$ in the case of the Carnival dataset and $\alpha = 0.58$ in the case of the *Festa del Redentore* dataset. This difference suggests a less effective mobility during the Carnival than during the *Festa del Redentore*, probably due to the weather conditions in winter, but also by the many activities that could attract the attention of people. To relate the empirical observations with a microscopic dynamic model, we propose a relation between the walking time t_w and the mobility time t of the form

$$dt_w = \frac{\alpha\, dt}{(1 + t/\tau)^{1-\alpha}}, \tag{2}$$

where τ is a fatigue scale time for pedestrian mobility and $\alpha > 0$ measures the efficiency of the mobility: $\alpha \to 1$ is the most efficient mobility when space and time are proportional. The relation (2) implies that if $t < \tau$ the mobility time practically coincides with the walking time, whereas the walking time reduces to a small fraction of the mobility time when $t \gg \tau$ as fast as $\alpha \ll 1$. For a typical visit of 6 h in the Venice historical centre, the formula (2) implies that the walking time fraction is $t^{\alpha}\tau^{1-\alpha} \simeq 2.5$ h for a fatigue time scale $\tau \simeq 1$ h. A simple calculation gives

$$s = t^{\alpha} v_0 \tau^{1-\alpha}\left[\left(1 + \frac{\tau}{t}\right)^{\alpha} - \left(\frac{\tau}{t}\right)^{\alpha}\right] \simeq \bar{v}_0 \tau^{1-\alpha}\alpha t^{\alpha}, \quad t \gg \tau$$

Figure 6 Relation between the average path lengths ⟨s⟩ and the mobility times: the left picture refers to the Carnival dataset and the right picture to *Festa del Redentore* dataset. The plots are obtained performing a running average of length 100 on the (t, s) data. The continuous line is the result of an power law interpolation (cfr. Equation (1)) with exponents $\alpha = 0.41$ in the first case and $\alpha = .58$ in the second one, whereas the proportionality coefficient is $\simeq 1.7$ in both cases

so that one recovers Eq. (1)

$$\langle s \rangle = \frac{\bar{v}_0}{\alpha} \tau^{1-\alpha} t^{\alpha}. \tag{3}$$

We remark that the relation (3) is singular when $\alpha \to 0$ (i.e. there is no mobility). Moreover the validity of Eq. (2) for long times t is questionable since they can be affected by the device activities at home, hotels or restaurants. The numerical interpolation provides the

value
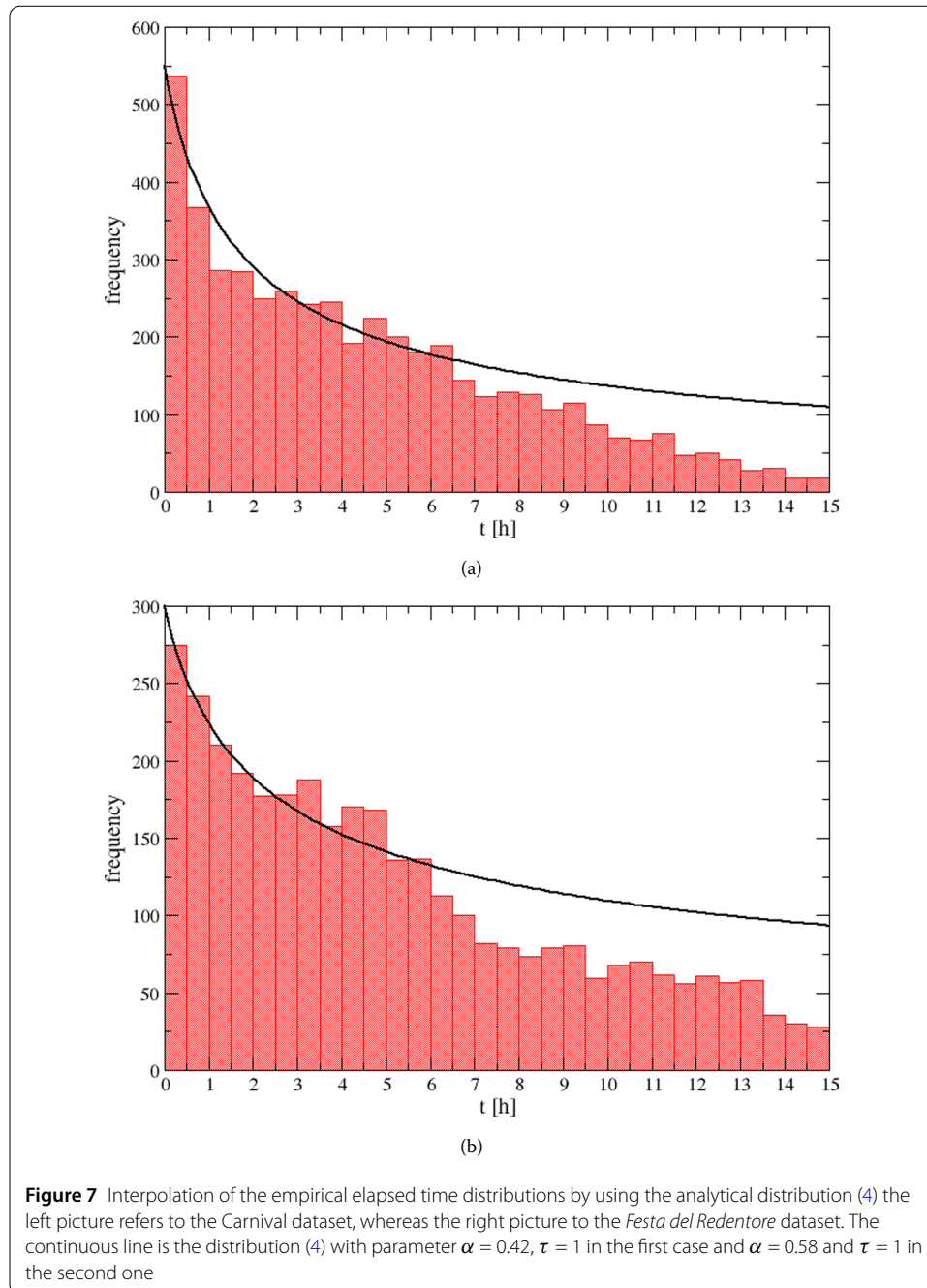
$$\bar{\nu}_0 \tau^{1-\alpha} \simeq 1.7$$

so that estimating $\bar{\nu}_0 \simeq 0.5$ m/sec as a typical average pedestrian velocity, one obtains the fatigue time scale $\tau \simeq 1$ h. This approach provides an analytical formula for the mobility time distribution once the distribution of $\langle s \rangle$ is known. Due to the great individual variability in the recorded mobility, the distribution of $\langle s \rangle$ is no more exponential and the approximation with a constant distribution is reasonable at this stage (see Additional file 1). Then one obtains a mobility time distribution of the form

$$p(t) \propto (1 + t/\tau)^{-(1-\alpha)}. \tag{4}$$

We remark that this distribution is not summable and we expect a validity for a limited time interval. In Fig. 5 we show the comparison between the empirical mobility time distribution and the analytical distribution (4). The parameters used in the interpolation are consistent with the interpolation shown in Fig. 7 with $\tau = 1$. We remark as the analytical law provides a quite good interpolation of the mobility time distributions with $t \in [0 : 6]$ h, whereas the distribution tail is still of exponential nature.
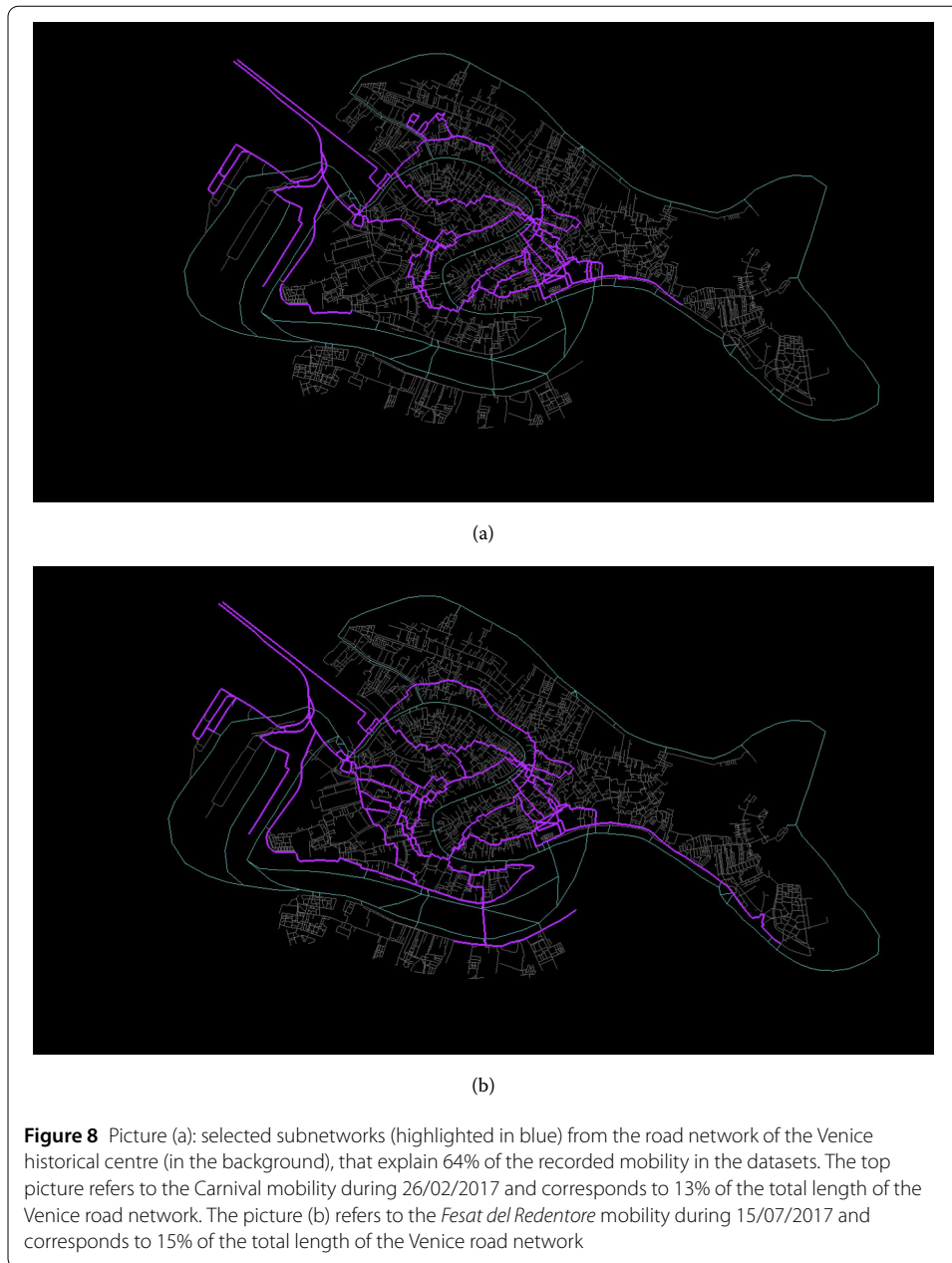
4 Pedestrian mobility network

The reconstruction of the mobility paths also allows to study how people perform their mobility on the road network. We consider the problem of determining the most used subnetwork of the Venice road network. The existence of mobility subnetworks could be the consequence of the peculiarity of Venice road network, where it is quite easy to get lost if you do not have a map. Therefore people with a limited knowledge of the road network move according to paths suggested by internet sites or following the signs on the roads. To point out a mobility subnetwork we rank the roads of Venice according to a weight proportional to the number of mobility paths passing through each road. Then, we have applied an algorithm to extract a connected subnetwork, which contains the roads in the ranking able to explain a fixed percentage of the observed mobility (see Additional file 1 for a brief description of the main steps of the algorithm). We are able to extract a subnetwork which explains the 64% of the observed mobility using 13% of the total road network length for the case of the Carnival dataset and 15% of the total length in the case of the *Festa del Redentore* dataset. The selected road subnetworks are plotted in Fig. 8 for both the datasets. As a matter of fact, many of the highlighted paths are also suggested by internet sites [38]. However, we remark some differences that can be related by the different nature of the considered events. During the Carnival of Venice the mobility seems to highlight three main directions connecting the railway station and the Piazzale Roma (top-left in the map), which are the main access points to the Venice historical centre, with the area around San Marco square, where many activities where planned during 26/02/2017. In the case of the *Festa del Redentore* the structure is more complex due to the appearance of several paths connecting the station and Piazzale Roma with the Dorsoduro district in front of the *Giudecca* island (see map in Additional file 1). This geometrical structure could have a double explanation: on one hand the *Festa del Redentore* introduces an attractive area near the Giudecca island, where the fireworks take place in the evening; on the other

Figure 7 Interpolation of the empirical elapsed time distributions by using the analytical distribution (4) the left picture refers to the Carnival dataset, whereas the right picture to the *Festa del Redentore* dataset. The continuous line is the distribution (4) with parameter $\alpha = 0.42$, $\tau = 1$ in the first case and $\alpha = 0.58$ and $\tau = 1$ in the second one
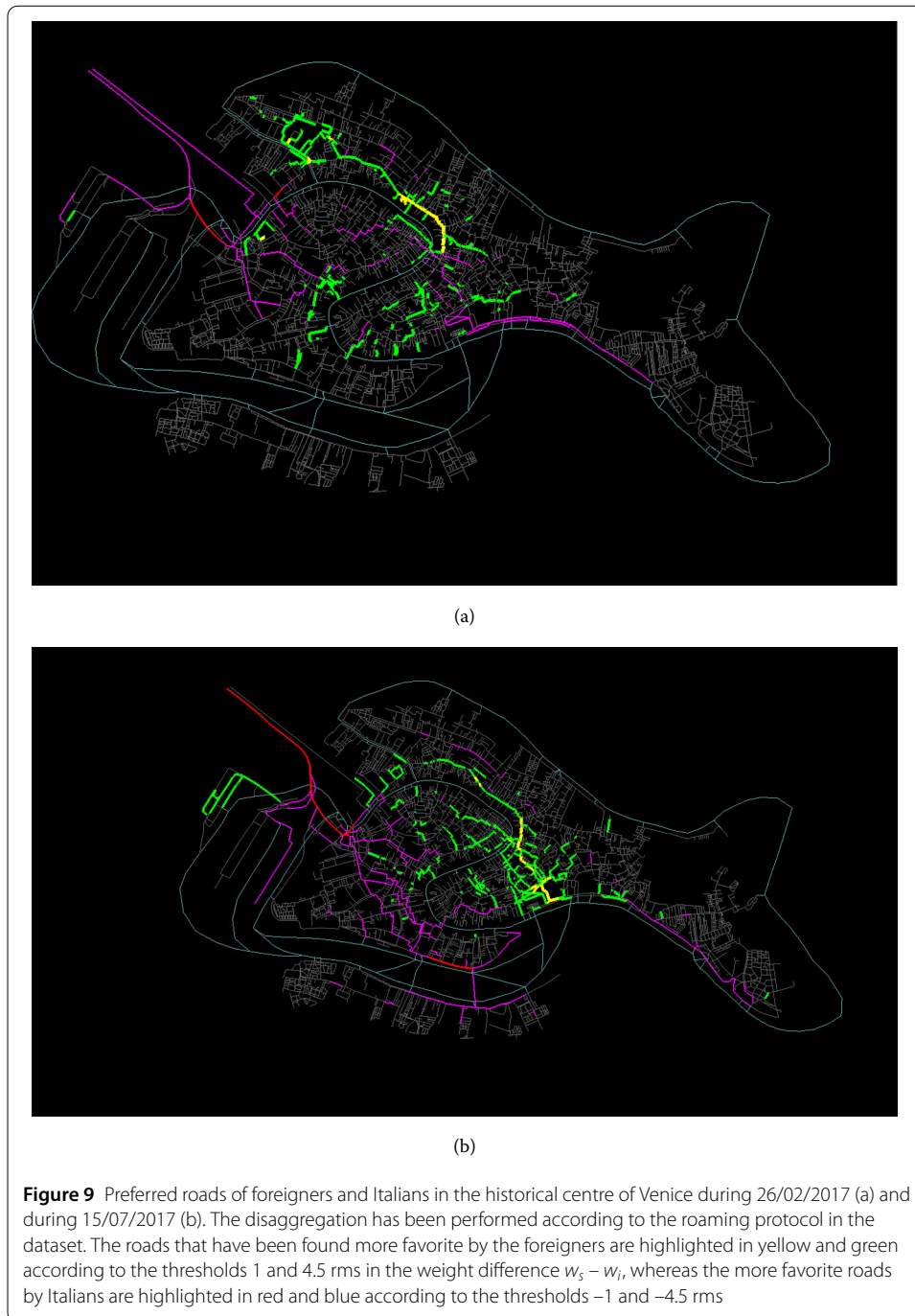
hand the *Festa del Redentore* is a festivity very much felt by the local population, that knows the Venice road network and performs alternative paths.

4.1 Foreigners versus Italians mobility

To study the possible effect on the mobility of a greater custom to visit Venice, we divide the devices in the datasets into Italian and foreign devices according to the roaming protocol. The technical details that allow this disaggregation are reported in the Supplementary Material. Of course we have no guarantee that all the Italians are more used to visit Venice than the foreigners, but this is a reasonable assumption on average, since many commuter

(a)

(b)

Figure 8 Picture (a): selected subnetworks (highlighted in blue) from the road network of the Venice historical centre (in the background), that explain 64% of the recorded mobility in the datasets. The top picture refers to the Carnival mobility during 26/02/2017 and corresponds to 13% of the total length of the Venice road network. The picture (b) refers to the *Fesat del Redentore* mobility during 15/07/2017 and corresponds to 15% of the total length of the Venice road network

visitors come from neighboring regions during the considered events. Then we have associated to each road two normalized weights $w_{fo,it}$ proportional to the number of mobility paths of Italians and foreigners on the road itself (the detected Italians are approximately 10 times the foreigners). In this way we select the roads that are respectively preferred by the Italians and by the foreigners considering the distribution of the difference $w_{fo} - w_{it}$ and introducing thresholds at ± 1 rms and ± 4.5 rms. In Fig. 9 we plot the results for the two datasets. We remark that not all the highlighted roads are present in the subnetworks in Fig. 8 since it was not possible to connect them using the high ranked roads in our list. It is noteworthy to observe that the majority of highlighted roads show a well defined preference by one of the two populations (i.e. their difference $|w_{fo} - w_{it}|$ is greater than 4.5 rms). During the Carnival the foreigners follow a path passing through *Strada Nuova*
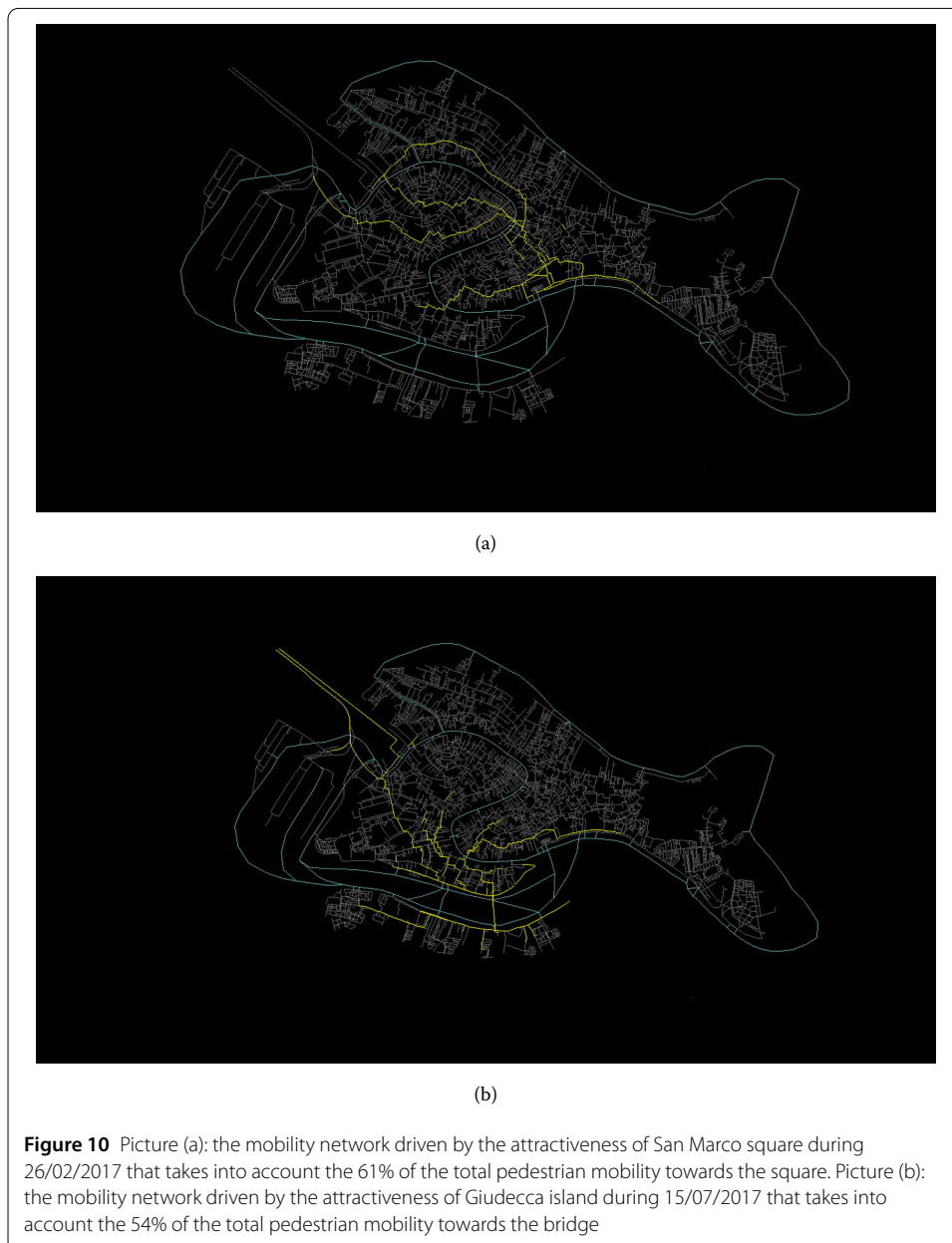
Figure 9 Preferred roads of foreigners and Italians in the historical centre of Venice during 26/02/2017 (a) and during 15/07/2017 (b). The disaggregation has been performed according to the roaming protocol in the dataset. The roads that have been found more favorite by the foreigners are highlighted in yellow and green according to the thresholds 1 and 4.5 rms in the weight difference $w_s - w_i$, whereas the more favorite roads by Italians are highlighted in red and blue according to the thresholds −1 and −4.5 rms

to reach San Marco square and the Rialto bridge, whereas Italians prefer to go through the central part of the Venice historical centre. Moreover we have two clear attraction areas for the foreign people at the *Old Getto* (up left in the picture) and near *Palazzo Grassi* (in the center of the picture). These preferences are also observed during the *Festa del Redentore* with the exception of the *Old Getto* that was not pointed out by the algorithm. But the attractiveness of San Marco square is increased for the foreigners with respect to the Italians that prefer to reach the area in front of *Giudecca* island. This is consistent with the

structure of the mobility subnetwork in this area that seem to be used manly by Italians (Fig. 8 (bottom)).

4.2 Attractiveness of the main areas of interest

Finally we analyze the mobility driven by the areas of greatest attractiveness like San Marco square during the Carnival and the Giudecca island during the *Festa del Redentore*. We select the mobility paths passing through San Marco square (or the Redentore bridge) and we reconstruct the mobility network defined by incoming paths. The results are plotted in Fig. 10: for the Carnival dataset we select $\simeq$1200 mobility paths corresponding to the 42% of the total mobility, whereas for the *Festa del Redentore* dataset we select $\simeq$700 mobility paths corresponding to 19% of the total mobility. The highlighted road networks explain

(a)

(b)

Figure 10 Picture (a): the mobility network driven by the attractiveness of San Marco square during 26/02/2017 that takes into account the 61% of the total pedestrian mobility towards the square. Picture (b): the mobility network driven by the attractiveness of Giudecca island during 15/07/2017 that takes into account the 54% of the total pedestrian mobility towards the bridge

the 61% (resp. 54%) of the total pedestrian mobility towards the San Marco square (resp. towards the Giudecca island) in the datasets. In the first case the analysis points out three main mobility pedestrian paths starting from the main entry points (the railways station and the Piazzale Roma parking area) that joins near the Rialto bridge. From the Rialto bridge the observed mobility presents a more diffusive character and it does not clearly define a path. Then we have an incoming path from the *Riva degli Schiavoni* due to the ferryboat line contribution and a well defined path between San Marco and the Accademia Bridge (see map in Additional file 1).

In the second case we observe a single pedestrian path from the main entry points towards the Giudecca island, whereas we have various incoming paths along the canal banks, indicating that people arrived by ferryboat. Noteworthy, there is not a clear connection between the San Marco square and the Giudecca island suggesting that the most of the people interested in the *Festa del Redentore* in the evening have not visited San Marco before.

5 Conclusion

The possibility of recording accurate anonymous georeferenced positions of mobile ICT devices whenever they perform an activity, provides dynamic information on the people mobility on a whole road network. Even if the requirements to reconstruct reliable daily mobility paths strongly reduce the penetration of the considered samples, we succeed to study some statistical and dynamic properties of pedestrian mobility in Venice. We explicitly analyze the pedestrian mobility during two large tourist events, but our methodologies apply to any dynamic GPS dataset containing information on individual mobility on a road network. The historical centre of Venice is an ideal experimental field to study the features of pedestrian mobility and the choice of two big tourist events (the Carnival of Venice 2017 and the *Festa del Redentore*) as case studies allows on one hand to increase the representativeness of the sample and on the other hand to provide quantitative information to the stakeholders that are in charge of the management of tourist flows. Our result are consistent with the existence of a 'mobility energy' and they point out the relevance of a 'fatigue effect' that reduces the average speed of a mobility path as the mobility time increases. Moreover the distribution of the mobility paths on the Venice road network, allows both to reconstruct connected subnetworks able to explain the majority of the observed mobility and to give information on how people use the road network to reach the main areas of interest. These results can be also relevant for the realization of a monitoring system of the pedestrian flows in Venice, suggesting where to install the people counting devices and how the local measures can be correlated to the mobility state of the road network.

The possibility of disaggregating Italians from foreigners by the roaming protocol, shows some different behaviors that should be further analyzed to understand if they can be related to the different knowledge of Venice road network. The different features of the two events (the Venice Carnival takes place during two weeks in winter, whereas the *Festa del Redentore* is a religious holiday in summer) are reflected by different dynamic properties of the observed mobility. Our results show the possibility of using the quality of the GPS data on small sample of mobile devices to build useful tools to study the individual mobility at the spatial scale of the road and to tune dynamic models [39] of pedestrian flows that perform a nowcasting and forecasting of the mobility state of the whole road network to avoid critical states. We expect that in the next future the quality and the quantity of GPS

datasets provided by the ICT will continuously increase and that their study contribute to the debate on the development of the Smart City paradigm.

Additional material

> **Additional file 1:** Supplementary material (PDF 1.8 MB)

Abbreviations
ICT, Information Communication Technologies; GPS, Global Positioning System.

Availability of data and materials
Due to the Italian law on privacy the original data are not of public domain. The dataset is property of TIM and its availability requires a non disclosure agreement with TIM (contact e-mail: davide.micheli@telecomitalia.it). The data in a aggregated form are available on request.

Competing interests
The authors declares that they have no competing interests.

Authors' contributions
SR, AF, CM, FB, SS performed the data analysis and elaborated the georeferencing algorithms to reconstruct individual paths; NC, RL, GV performed the mobility network analysis and participated to the data collection on the *Ponte del Redentore*; MD, GM, AV collected and prepared the GPS datasets, AB conceived and supervised the work. All authors read and approved the final manuscript.

Author details
[1]Physics and Astronomy Department, University of Bologna, Bologna, Italy. [2]TIM S.p.A., Roma, Italy.

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

References
1. Vespignani A (2012) A modelling dynamical processes in complex socio-technical systems. Nat Phys 8:32–39
2. Batty M, Axhausen KW, Giannotti F, Pozdnoukhov A, Bazzani A, Wachowicz M, Ouzounis G (2012) Smart cities of the future. Eur Phys J Spec Top 214(1):481–518
3. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. Nature 439:462–465
4. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility pattern. Nature 453:779–782
5. Song C, Koren T, Wang P, Barabasi AL (2010) Modelling the scaling properties of human mobility. Nat Phys 6(10):818–823
6. Gallotti R, Bazzani A, Rambaldi S (2012) Towards a statistical analysis of human mobility. Int J Mod Phys C 23:1250061
7. Yan XY, Han XP, Wang BH, Zhou T (2013) Diversity of individual mobility patterns and emergence of aggregated scaling laws. Sci Rep 3:2678
8. Gallotti R, Bazzani A, Degli Esposti M, Rambaldi R (2013) Entropic measures of individual mobility patterns. J Stat Mech Theory Exp 2013:P10022
9. Zhao K, Musolesi M, Hui P, Rao W, Tarkoma S (2015) Explaining the power-law distribution of human mobility through transportation modality decomposition. Sci Rep 5:9136
10. Gallotti R, Bazzani A, Rambaldi S, Barthelemy M (2016) A stochastic model of randomly accelerated walkers for human mobility. Nat Commun 7:12600
11. Song C, Qu Z, Blumm N, Barabasi AL (2010) Limits of predictability in human mobility. Science 327(5968):1018–1021
12. Lin M, Hsu WJ, Lee ZQ (2012) Predictability of individuals' mobility with high-resolution positioning data. In: Proceedings of the 2012 ACM conference on ubiquitous computing. ACM, New York, pp 381–390
13. Cuttone A, Lehmann S, Gonzalez MC (2018) Understanding predictability and exploration in human mobility. EPJ Data Sci 7:2
14. Batty M (2016) Big data and the city. Built Environ 42(3):321–337
15. Shelton T, Matthew Zook M, Wiig A (2015) The 'actually existing smart city'. Camb J Reg Econ Soc 8:13
16. Kitchin R (2014) The real-time city? Big data and smart urbanism. GeoJournal 79:1
17. Batty M, Desyllas J, Duxbury E (2003) Safety in numbers? Modelling crowds and designing control for the Notting Hill Carnival. Urban Stud 4(8):1573–1590

18. Omodei E, Bazzani A, Rambaldi S, Michieletto P, Giorgini B (2014) The physics of the city: pedestrians dynamics and crowding panic equation in Venezia. Qual Quant 48(1):347–373
19. Moussaïd M, Perozo N, Garnier S, Helbing D, Theraulaz G (2010) The walking behaviour of pedestrian social groups and its impact on crowd dynamics. PLoS ONE 5(4):e10047
20. https://www.tim.it/
21. Candia J, Gonzalez MC, Wang P, Schoenharl T, Madey G, Barabasi AL (2008) Uncovering individual and collective human dynamics from mobile phone records. J Phys A, Math Theor 41(22):224015
22. Becker R, Caceres R, Hanson K, Isaacman S, Loh JM, Martonosi M, Rowland J, Urbanek S, Varshavsky A, Volinsky C (2013) Human mobility characterization from cellular network data. Commun ACM 56(1):74–82
23. Csáji BC, Browet A, Traag VA, Delvenne JC, Huens E, Van Dooren P, Smoreda Z, Blondel VD (2013) Exploring the mobility of mobile phone user. Phys A, Stat Mech Appl 392(6):1459–1473
24. Xu Y, Shaw SL, Zhao Z, Yin L, Lu F, Chen J, Fang Z, Li Q (2016) Another tale of two cities: understanding human activity space using actively tracked cellphone location data. Ann Am Assoc Geogr 106(2):489–502
25. Ratti C, Frenchman D, Pulselli RM, Williams S (2006) Mobile landscapes: using location data from cell phones for urban analysis. Environ Plan B, Plan Des 33(5):727–748
26. Calabrese F, Di Lorenzo GD, Liu L, Ratti C (2011) Estimating origin-destination flows using mobile phone location data. IEEE Pervasive Comput 10(4):0036
27. Xu Y, Shaw SL, Zhao Z, Yin L, Fang Z, Li Q (2015) Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. Transportation 42(4):625–646
28. Bazzani A, Giorgini B, Rambaldi S, Gallotti R, Giovannini L (2010) Statistical laws in urban mobility from microscopic GPS data in the area of Florence. J Stat Mech Theory Exp 2010:P05001
29. Toole JL, Colak S, Sturt B, Alexander LP, Evsukoff A, Gonzalez MC (2015) The path most traveled: travel demand estimation using big data resources. Transp Res, Part C, Emerg Technol 58(Part B):162–177
30. Mokhtariana PL, Chenb C (2004) TTB or not TTB, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets. Transp Res, Part A, Policy Pract 38(9–10):643–675
31. Gallotti R, Bazzani A, Rambaldi S (2015) Understanding the variability of daily travel-time expenditures using GPS trajectory data. EPJ Data Sci 4:18
32. https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration
33. http://www.veneziatoday.it/eventi/carnevale-venezia-2017-numeri-record.html, https://www.ilgazzettino.it/nordest/venezia/redentore_venezia_2017_foto-2563968.html
34. https://www.openstreetmap.org/#map=14/45.4365/12.3546
35. http://www.corila.it/
36. Böcker L, Martin Dijst M, Prillwitz J Impact of everyday weather on individual daily travel behaviours in perspective: a literature review. Transp Rev 33(1):71 (2013)
37. Geroliminis N, Daganzo CF (2008) Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transp Res, Part B, Methodol 42(9):759
38. http://www.camminandoavenezia.com/itinerari/
39. Barbosa-Filho H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M Human mobility: models and applications. https://arxiv.org/abs/1710.00004