Noname manuscript No.
(will be inserted by the editor)

Pedestrian behavior analysis on a road network using ICTs during great tourist events PRELIMINARY VERSION - NOT FOR DIFFUSION

The Venetia case study

UNIBO-TIM-A4SMART collaboration

Received: date / Accepted: date

Abstract Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

Keywords First keyword · Second keyword · More

1 Introduction

The main problem is to reveal how the tourist flows during big attractive events move on a road network of an historical city. The special features of Venice allow to test our approach on a large historical center where the mobility is mainly pedestrian or performed by the public ferry systems. The two chosen periods are both related to

2 Dataset

The dataset used in this study has been provided by TIM and contains georeferenced positions of tens of thousands devices (e.g. mobile phones, tablets, etc..), that performed an activity (e.g. a phone call or an internet access) during six days from 23/2/2017 up to 2/3/2017 when the Venetian Carnival was going on, and from 14/7/2017 up to 16/7/2017 on the occasion of the *Festa del Redentore*. The data set refers 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 . Moreover there exists a special flag that denotes the devices in a roaming mode, that can be associated to foreign people. Each valid record gives information on the GNSS data of the device, on the recording time and on the signal quality with two anonymized ID for the activity and for the

Address(es) of author(s) should be given

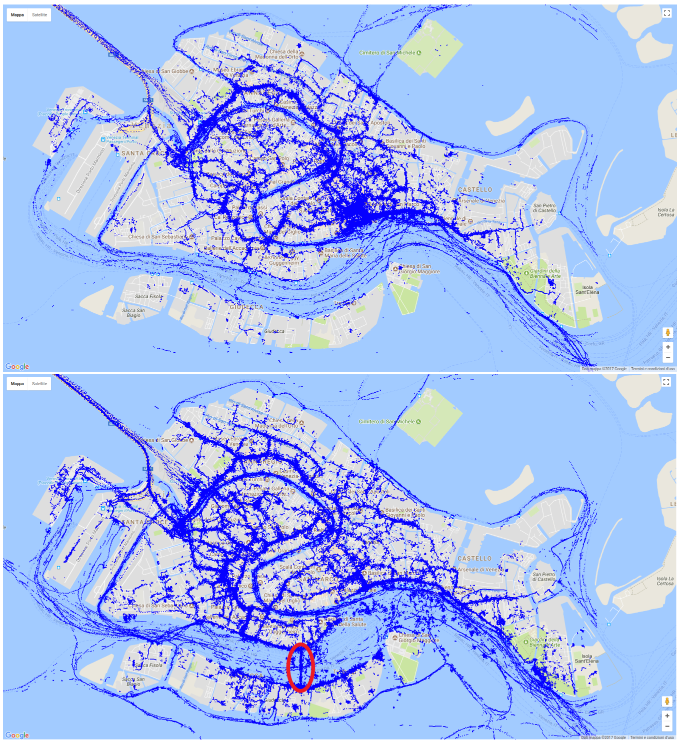

Fig. 1 Examples of the distribution of the GNSS data recorded in the Venice histrorical center: the top picture refers to the Carnival dataset and shows the data of 26/02/2017 from 12 to 14 o'clock. The bottom picture to the *Redentore* dataset and shows the data of 15/07/2017 from 19 to 21 o'clock. The red circle points out the the Redentore bridge location, which is a floating bridge built for the special occasion of the *Festa del Redentore*.

device. During each activity that lasts over time a sequence of GNSS data is recorded with a sampling rate from 2 sec., so that it is possible both to follow local trajectories and to detect point of interest at which people stop for a certain time. Both the Carnival and the *Festa del Redentore* datasets contains $\simeq 1.5 \times 10^6$ georeferenced records concerning the mobility in the historical center of Venice during each observation day. These data mobile phone ID allows to study the daily mobility of a sample of the device population of $\simeq 5000$ devices per day with the possibility of reconstructing the main paths used in the road network and to estimate the average activity rate of a device during the circadian rhythm. Since the presences in the historical center of Venice during the events were of the order of 10^5 individuals per day, expect an overall penetration of our sample of $\simeq 5\%$. The figure

We have performed a filtering process on the available datasets to extract data that contain relevant information to study the mobility on the road network. We aggregate the GNSS data of each device-ID and we downsample the data starting from an initial position (pivot point) and computing the geodesic distance 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 $\simeq 60 \times 10^3$ per day in the Carnival dataset and to $\simeq 80 \times 10^3$ in the *Festa del Redentore* dataset. Each selected GNSS point is located in the nearest links of the road network within a distance of 60m included the ferryboat lines; the points that cannot be attributed to any link according to this criterion, are discharged. The positioning procedure further reduces the valid points by 20% and finally we have $\simeq 3 \times 10^5$ corrected georeferenced positions for the Carnival dataset and $\simeq 2.1^5$ points for the *Festa del Redentore* dataset. These positions allow to get dynamical information both on the most visited paths on the road network, and on the main points of interest where the rest time is significant. We have performed a direct check of the representativity of the considered sample at the spatial scale of a single road, by comparing the estimated flows using GNSS data with the measured flows (direct measures with people counting) on the Redentore bridge, that is a floating bridge on the *Canale della Giudecca* (see fig. 1 left). 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 p.m.. To estimate the pedestrian flow across the bridge we have counted the mobile devices that leave two signals at opposite sides of the bridge during the considered time slot, so that we distinguish between the bridge crossing into opposite directions. The results are reported in the figure 2 (left) where the estimated penetration of our sample is $\simeq$ 1.6% according to a best fit of the measures with an average error of 20% (excluding the flow measured after the reopening of the bridge). The reduced sample penetration with respect to the expected %5, is sue to the small spatial scale of the bridge that requires a coincidence of two signals from the same device at opposite sides of bridge in a short time interval. Then we expect that the variability of the activity rate reduces the sample penetration. We have computed the 10 minutes activity rate for the devices located in a area near the bridge from 7:00 p.m. of 15/7/2017; the results are reported in the fig. 2 (right). Finally, we remark that the estimated flows allows to reproduce with good accuracy the evolution of the empirical observation except for a single point after between midnight and one o'clock a.m. when the bridge was reopened after the firework. Indeed a big pedestrian flow was recorded between 12:30 p.m. and 1:00 a.m. that is not detected by the GNSS dataset. A possible explanation was that the device activity in the area was dropped down during and after the fireworks since most of people was mainly interested in attending the show and, afterwards, to cross the bridge towards the Venice center moving in a crowded environment. Indeed using the direct empirical observations we evaluate a net flow towards the Giudecca island of 8000 people from the opening of the bridge and a net flow of 14000 people after the bridge reopening (since probably some people reach the island by ferry). The GNSS dataset estimates correctly the incoming flow, but underestimate the outgoing flow of approximately 8000 people. This estimate could be consistent if the device activity at the bridge is reduced by

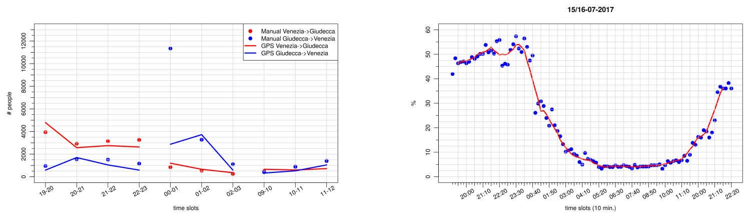

Fig. 2 Left picture: comparison of the hourly flows on the *Redentore* bridge estimated from the GNSS data base (continuous curves) and the empirical measures by a direct people counting (dots): the blue data refer to the pedestrian flow from *Giudecca* island toward Venice, whereas the red data refer to the pedestrian flows on the opposite direction. The scaling factor applied to the GNSS data corresponds to penetration of 1.6% of the sample. We recall that the bridge was closed between midnight and one o'clock. Right picture: empirical relative frequency to get a GNSS record in a time interval of 10 minutes, from a device of our sample present in the area of interest near the *Redentore* bridge; the red line is an mean average over one hour to smooth the fluctuation effect.

a factor 3 the time interval between 00:30 and 1:00 a.m.. We observe as the device activity rate increases before the firework show, probably due to the people excitation before the big event of the day, and quickly drop down by a factor 2 afterward. However the empirical evidence suggests that the selected sample recovers its representativity during the night after the first time slot. A possible explanation is that at the reopening of the bridge, the high level of crowding due to the pedestrian flow incoming to Venice discourages people to use the ICT while they are walking. As a consequence many people move away from the area without leaving any signal in the dataset. When a normal condition is recovered, the device activity of the remaining people is at the same level as before the fireworks, but we have a lower average value since we are still considering the whole device population. According to the previous discussion, a model for the pedestrian flows based on the ICT device activities could miss to detect localized critical situations.

3 Mobility path reconstruction on the road network

The filtered GNSS positions are geolocalized on the Venice road network that has to be extended to include the ferryboat lines (see figure 3 left). The procedure considers separately the land mobility and the water mobility since the two mobility networks are physically separated and it is necessary to check carefully the transitions from one network to another. 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 when the path changes from land mobility to water mobility and on the estimated travel speed. In the first case we require at least two successive points inn the sequence are attributed on a ferry line to end a land path and to start a water path, otherwise we force the localization of the point attributed to a ferry line on the nearest road. In the second case we discharge the paths whose velocity is clearly not consistent

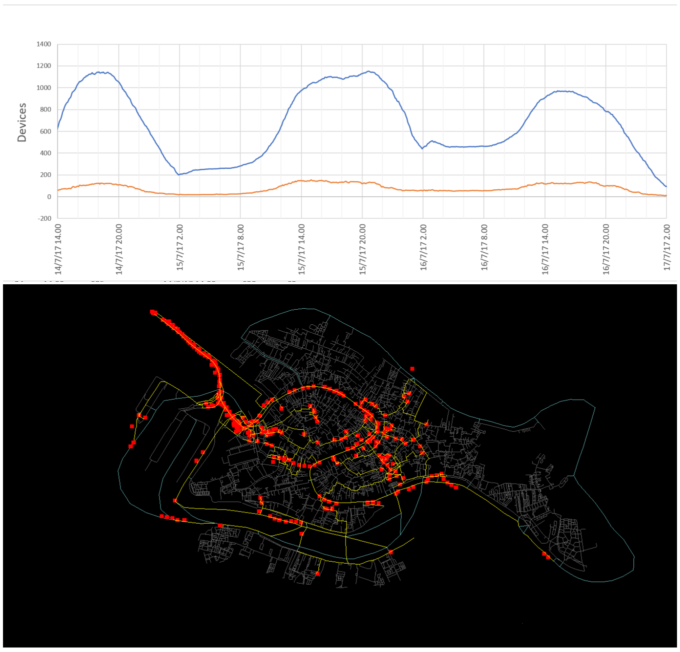

Fig. 3 Top picture: number of selected devices present the *Redentore* dataset collected during the three days: we observe the anomalous increase of the presence during the night of 15/7/2017. Bottom picture: some examples of mobility paths reconstructed (continuous lines) on the road network of the historical center on Venice using GNSS data (red dots).

with the typical pedestrian velocity (or ferryboat velocity). Finally we have neglected anomalous paths which crosses a very high number of roads (more than 200) or have a very low number of points (less than 3). In the first case we attribute these path to people performing a particular activity in Venice (like for example a postman), which not related to the tourist or citizen mobility, In the second case the associated paths are too short to study the mobility properties. In this way we reconstruct the daily mobility of $\simeq 4000$ different devices per day both for the *Festa del Redentore* dataset and the Carnival dataset. In fig. 3 (left) we show the measured number of moving devices detected in the historical center of Venice, whose mobility paths have been correctly reconstructed by the algorithms during the *Redentore* celebration: 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 fireworks. According estimated penetration of 1.6% of our sample the GNSS data estimate a peak of $\simeq 80,000$ people during the evening of 15th July and a number of passengers of order 10^4 on the ferryboat transportation system. These numbers are consistent with estimates provided by the local newspapers. In fig. 3 (right) we report some examples of the reconstructed mobility paths on the Venice road network joined with the ferry lines on the channels. The mobility paths provide dynamical information on how people realize their mobility demand on the Venice road network in Venice during the big tourist events. The elapsed time between two successive GNSS data is used to attribute a displacement velocity that of course is also affected by the rest times at the places of interest. A dynamical model to simulate the pedestrian flow dynamics on the Venice road network has to take into account the features of the microscopic dynamics that includes tracts covered at constant velocity and breaks due to the presence of points of interests or to necessity to recover from the walking effort. We have considered the some statistical properties of the mobility paths to check if their are consistent with other statistical laws suggested by the analysis of mobility dataset in urban contexts[]. In the figure 6 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 *Redentore* datasets. This difference can be explained both by the effect of weather (the Carnival takes place on winter whereas the *Festa del Redentore* is celebrated during summer time) and by the features of the two events. The Venice Carnival is an ensemble of events spread on the historical center even if San Marco square is always the main attractive location, whereas the *Festa del Redentore* program is concentrated in the area near *Canale della Giudecca* between the Giudecca island and *Riva degli Schiavoni*. Therefore one expects a mobility more influenced by an origin destination nature during the *Festa del Redentore* (the distance between the Station and the Giudecca island is $\simeq 2.3$ km), whereas the mobility during Carnival has a larger component with a random characters. We remark that in the path distribution in fig. 6 refers only to pedestrian mobility since we have excluded all the mobility paths with a tract on a ferry line. This criterion is satisfies by 2/3 devices in our sample, whereas the remaining 1/3 performs a mixed mobility between walking and using the ferry lines. This estimate could reflect the expensive cost of the ferry lines for tourists in Venice. We propose an exponential interpolation of the path length distribution for both the datasets (cfr. dashed lines in the figure 6); 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* used by pedestrians in Venice and the exponential decaying of the distribution tail is in agreement with a Maxwell-Boltzmann distribution. The different characteristic lengths of the exponential decaying, 3.0 km for the Carnival dataset and 3.6 km for *Festa del Redentore* dataset, confirms the propensity of individuals to perform greater mobility in the second case and they can be interpreted as average mobility energies. In both cases these distances are greater than the typical pedestrian mobility in a city, but they are reasonable average walking distance in a historical city like 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 of short daily mobility path lengths has also to be related to the possibility of using the public transportation system, that, however, is quite expensive for tourists. The exponential tail of path length distribution is consistent with the

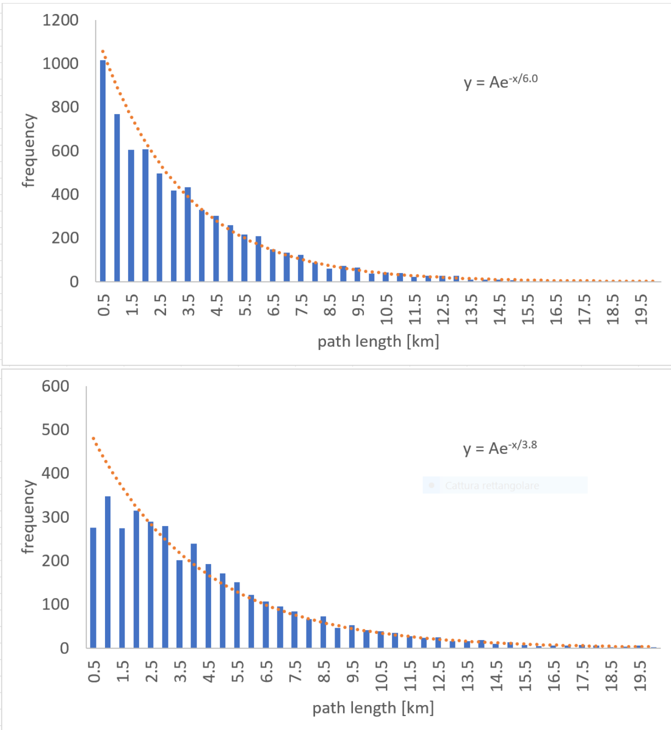

Fig. 4 Distribution of the mobility path lengths reconstructed during the Carnival (top picture) and the *Festa del Redentore* (bottom picture) in the Venice historical center. The dashed line is an exponential interpolation of the distribution tail whose formula is reported in the pictures.

concept of *travel time budget* proposed in other studies of urban mobility[]. To understand the statistical features of the observed mobility we have also considered the elapsed time distribution associated to the path lengths computed as the elapsed time between the first and the last recorded position of a device (see fig. 7). The elapsed time is the sum of the travel time and the rest time during the day and the distribution tail can be affected by the device activity during the night not directly related to the mobility. Assuming that if a device has spent more than 8 h in Venice then it has a high probability to spend also the night in Venice, we have $\simeq 1/3$ of the devices in the dataset that satisfy this criterion. The exponential interpolation is less justified in this case due to the effect of the rest times with respect to the mobility times and we have considered the relation between the mobility path lengths s and the elapsed time. The mobility path length is defined

$$s = \int_0^t v(u)du$$

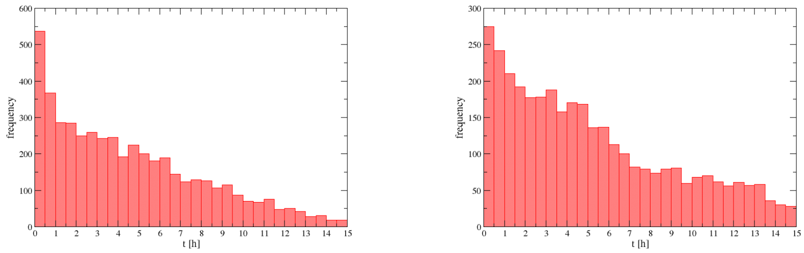

Fig. 5 Distribution of the elapsed time associated to the daily mobility paths reconstructed during the Carnival (left picture) and the *Festa del Redentore* (right picture) in the Venice historical center.

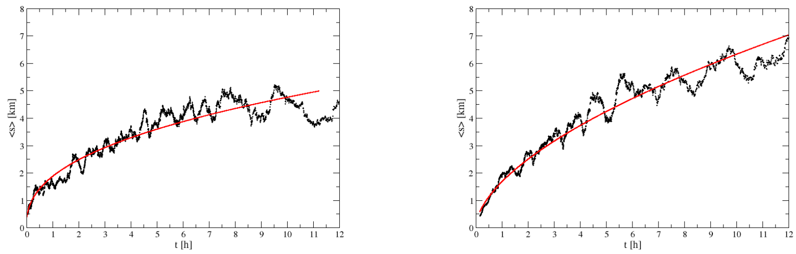

Fig. 6 Relation between the average path lengths s and the elapsed times: the left picture refers to the Carnival dataset and the right picture to *Festa del Redentore* dataset. The plots are the obtained performing a running average of length 100 on the (t, s) data. The continuous line is the result of an power law interpolation (cfr. eq. (1)) with exponents $\alpha = .41$ in the first case and $\alpha = .58$ in the second one, whereas the proportionality coefficient is $\simeq 1.7$ in both cases.

where $v(u)$ is the absolute value of velocity at time u and we have considered a power law relation

$$< s >= ct^{\alpha} \tag{1}$$

where we perform an average over the mobility paths with similar elapsed times. In the figure 7 we show the result of the interpolation of empirical data with the law (1) In normal crowding conditions, the pedestrian dynamics is performed at a constant velocity v_0, with a possible stochastic variation among individuals, and a linear relation $s = v_0 t_m$ is expected. The statistical law $< s > \propto t^{\alpha}$ with $\alpha < 1$ implies that the rest time $t - t_m$ increases t simulating a fatigue effect of individuals during pedestrian mobility. The interpolation of the empirical data gives an exponent $\alpha = .41$ in the case of the Carnival dataset and $\alpha = .58$ in the case of *Redentore* dataset. This different suggests a less effective mobility during Carnival than during *Festa del Redentore*, probably due to the weather conditions in winter, but also to crowding effects that were more relevant during Carnival. To link the empirical observations with a microscopic dynamics, we propose a relation between the mobility time t_m and the elapsed time t to interpret the empirical law of the form

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

where τ is a fatigue scale time for pedestrian mobility and α is an efficiency degree during the mobility (α could be affected by the frequent stops near the shops or by the crowding along the roads). The relation (2) implies that if $t < \tau$ the elapsed time practically coincides with the elapsed time, whereas the mobility time fraction reduces when $t \gg \tau$ as fast as α tends to 0. For a typical visit of 6 h in the Venice historical center, the formula (2) implies that the mobility time fraction is $t^alpha\tau^{1-\alpha} \simeq 2.5\ h$ for a fatigue time scale $\tau \simeq 1$ h (see below). A simple calculation gives

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

so that one recovers eq. (1)

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

where $\bar{s}$ denites an average on the path lengths conditioned to a given elapsed time t. We remark that the relation (2) is singular when $\alpha \to 0$ (i.e. there is not any mobility); then the validity of eq. (2) for long elapsed time is questionable since they are related to the device activity at home or in the hotels where the mobility is zero. The numerical interpolation provides

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

so that assuming $\bar{v}_0 \simeq .5$ m/sec. as a typical average pedetrian velocity during a tourist visit, one estimates the fatigue time scale $\tau \simeq 1$ h. This approach provides an analytical formula for the elapsed time distribution once the distribution of $\bar{s}$ is known. As a matter of fact, we have a great individual variability in the recorded mobility (see supplementary material) so that the distribution of $\bar{s}$ is no more exponential and the approximation with a constant distribution is reasonable at this stage. Then one derives a elapsed time distribution

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

We remark that this distribution is not summable and we expect a validity for a limited a elapsed time interval. In the figure 7 we show the comparison between the empirical elapsed time distribution and the analytic distribution (4). The parameters used in the interpolations are derived directly from fig. 6 with $\tau = 1$. We remark as the analytical law provides a quite good interpolation of the elapsed time distributions with $t \in [0 : 6]$ h whereas the tail queue is still of exponential nature.

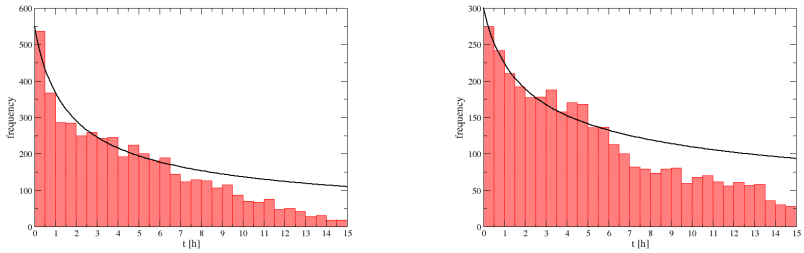

Fig. 7 Interpolation of the empirical elapsed time distributions by using the analytical distribution (4) the left picture refers to the Carnival dataset, whereas the right picture to *Redentore* dataset. The continuous line is the distribution (4) with parameter $\alpha = .42$, $\tau = 1$ in the first case and $\alpha = .58$ and $\tau = 1$ in the second one.

4 Pedestrian mobility network

The reconstruction of the mobility paths allows to study how people perform their mobility on the road network. We consider the problem to determine the most used subnetwork of the Venice road network. Indeed assuming a limited knowledge of tourists of the Venice road network, the presence of signals on the road network, the use of smartphone apps to determine the best paths and the presence of a herding effect in the pedestrian behavior could considerable limit the use of the road network in the historical center of Venice. To define a mobility subnetwork, we rank the roads according the their contribution to the total mobility recorded in the datasets: the road weight is the product of the road length times the number of path that pass through the road. Then varying the threshold for the weight road, we select a connected subnetwork that is able explain the 64% using 13% of the total road network length in the case of Carnival dataset and 15% of the total length in the case of the *Redentore* dataset. The selected subnetwork are plotted in the figure 8

5 Discussion

The reduce mobility network can be used as a backbone to model that simulate the mobility in the historical center and limits the information required to perform governance of the mobility based on a model for simulating the pedestrian flows on a road network.

Acknowledgements We are indebted with CORILA and CISET for their help to organize the data acquisition and the several helpful discussions.

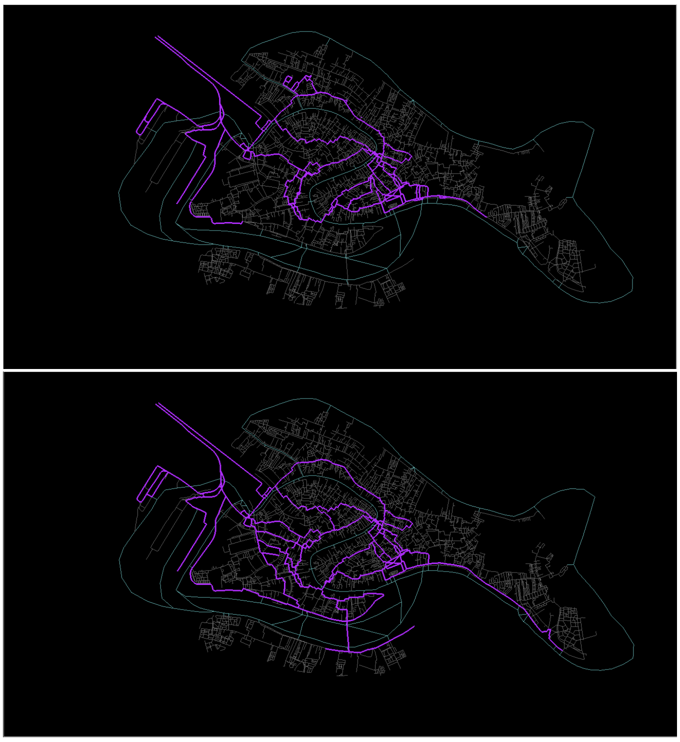

Fig. 8 Selected subnetworks (highlighted in blue) of the total road network of the Venice historical center (in the background) that explain 64% of the recorded mobility in the datasets.The top picture refers to the Carnival mobility during 26/02/2017 and corresponds to 13% of the total length of the Venice road network. The bottom picture refers to the *Redentore* mobility during 15/07/2017 and corresponds to 15% of the total length of the Venice road network.