Research Article

Vehicular Crowdsensing with High-Mileage Vehicles: Investigating Spatiotemporal Coverage Dynamics in Historical Cities with Complex Urban Road Networks

L. L. L. Starace, F. Rocco Di Torrèpadula, S. Di Martino, and N. Mazzocca

University of Naples Federico II, Department of Electrical Engineering and Information Technology (DIETI), Naples, Italy

Correspondence should be addressed to L. L. L. Starace; luigiliberolucio.starace@unina.it

Received 27 March 2023; Revised 15 April 2023; Accepted 24 April 2023; Published 12 May 2023

Academic Editor: Jaeyoung Lee

Background. Vehicular crowdsensing (VCS) can be a cost-effective solution to gather data in urban environments, leveraging the onboard sensors of modern vehicles moving around the city. Many experimental studies have proven that high-mileage vehicles, such as taxis, can be effectively used for VCS. However, these studies have been mostly carried out in cities with regular, grid-based, road networks. Conversely, little work has been conducted to assess the suitability of VCS in cities with more complex urban road networks, such as historical ones. Goal. As a step towards filling this gap, the present study investigates the feasibility of using different-sized fleets of taxis to crowdsense information in the urban areas of the historical cities of Porto (Portugal) and Rome (Italy), whose road networks evolved over the centuries and feature a complex topology. Data and Methodology. This work leverages massive real-world datasets of taxi trajectories collected over three contiguous weeks in the cities of Porto and Rome to estimate the spatiotemporal coverage achievable by different-sized fleets of taxis if they were used for VCS. Indeed, using these trajectories, several simulations were conducted, considering four sizes of taxi fleets, ranging from 50 to 400 vehicles, for both cities. The achievable spatiotemporal road network coverage metrics were computed at a fine-grained scale of single road segments. Results. Results show that the achievable coverage in both historical cities exhibits very similar trends, with as few as 50 vehicles being capable of visiting a relevant part of the road network at least once in the considered time frame. As expected, increasing the number of involved vehicles improves spatial and temporal coverage. Still, time gaps between subsequent visits can be possibly inadequate for some VCS use cases. As a consequence, recruiting more vehicles and/or devising specialized routing/incentivization mechanisms might be necessary to achieve more comprehensive coverage of the urban road network.

1. Introduction

Recently, new strategies have been developed to monitor large-scale phenomena, including leveraging explicit user feedback or sensors embedded in smartphones or wearable devices. Such an approach, referred to in the literature as mobile crowdsensing (MCS), has been shown to be a viable alternative to traditional approaches based on stationary sensing solutions [1]. Vehicular crowdsensing (VCS) is a particular case of MCS, based on exploiting the sensors installed in modern vehicles to collect contextual data useful for novel use cases. The sensed relevant data, such as the availability of a free parking slot, the current speed, or the presence of a heat island, are sent to a back-end server. Here, data from all the involved connected vehicles are aggregated and processed, to extract new contextual knowledge (e.g., the average speed on a road segment or the amount of rain in a given area) [2]. On top of VCS-collected information, several novel and compelling applications can be developed. Examples of these use cases include mobility recommender solutions [3], better surveillance of urban scenarios [4], more accurate mobility estimates [5], or air quality monitoring [6]. Recently, a study conducted by the McKinsey & Company consulting firm [7] reported that properly exploiting the knowledge that could be extracted from vehicle-collected data “could deliver $250 billion to $400 billion in annual incremental value for players across the ecosystem in 2030,” acknowledging the potential of VCS.
The achievable spatiotemporal distribution of the collected data, also referred to as sensing coverage, is one of the key performance indicators (KPIs) when assessing the feasibility of an MCS proposal [8]. In VCS scenarios, in particular, sensing coverage is mainly determined by two key factors: the number of probe vehicles participating in VCS and their spatiotemporal sensing distribution [9], which can enable different VCS-based use cases. For example, considering the relatively low rate of asphalt degradation, monitoring potholes in urban scenarios might require a probe vehicle to pass by a road once a day or even less frequently. Conversely, monitoring the availability of on-street parking requires way more frequent sensings [10]. As widely demonstrated by previous works, a swarm of passenger transportation or high-mileage vehicles, like those of delivery services, might achieve adequate spatiotemporal sensing coverage for a large number of these VCS-based use cases [10, 11]. For instance, the study conducted by Bock et al. in [12] evaluated the feasibility of using a fleet of 500 taxis to crowdsense real-time availability of on-street parking in some selected streets in the business district of San Francisco (USA). In that work, the authors leveraged three weeks of taxi positioning data and their analyses showed that (1) all the considered roads were traversed by the taxis in the investigated time frame, and (2) a relevant portion of the considered streets was visited by one of the taxis every few minutes during the day, confirming the feasibility of VCS for monitoring parking availability.

However, most of the studies investigating the spatiotemporal sensing distribution attainable in VCS by fleets of vehicles either focus on cities featuring a regular, grid-based road network topology (see Figure 1(b)) or abstract the underlying road network with a set of cells, as conducted by Masutani in [9]. This is a significant simplification of the problem and may not provide enough insight into VCS scenarios requiring specific road segment-specific sensing, such as on-street parking monitoring. To the best of our knowledge, no replication study has been conducted to thoroughly investigate the generalizability of these studies on different, more complex urban road networks such as the ones of historical European cities.

As a first step towards filling this gap, Martino and Starace in [13] analysed the suitability of a swarm of 100 taxis for VCS in the city of Porto, a medium-sized city in Portugal. The road network of the considered city presents an organic, irregular topology, as a result of centuries of urban evolution. That work exploited an open dataset of more than 1.7 million trajectories collected from 441 taxis in Porto over a one-year time span and showed that 100 taxis could guarantee adequate coverage of the road network to support many VCS use cases. That work, however, considered only a single, medium-sized historic city (Porto) and did not investigate the impact of fleet size on the achievable spatiotemporal coverage. Moreover, the temporal coverage analysis was performed only on the entire three-week period, without considering fluctuations in different time slots during the day.

The present work significantly broadens that study in two ways. First, it also includes a new historic city, i.e., Rome, in Italy, which is way bigger than Porto and is characterized by a complex urban road network as well (see Figure 1(a)). Second, a new dimension is added to the investigation by also evaluating the impact of the number of vehicles involved in VCS activities on the resulting road network coverage. Indeed, in real-world scenarios, selecting an adequate number of participants in VCS is a crucial step. On the one hand, when few participants are selected, the achieved sensing coverage might be inadequate. On the other hand, selecting too many participants might result in wasting money on sensors and/or incentivization strategies [14, 15]. Hence, for both the cities of Porto and Rome, a further assessment was conducted, taking into account four alternative scenarios, each featuring a different number of taxis, ranging from 50 to the maximum number of involved vehicles. The experiments were conducted exploiting two datasets of real-world taxi trajectories, from Porto and Rome, where the latter contains traces from 315 taxis.

For all the considered scenarios, a number of spatiotemporal coverage metrics were computed, providing useful insight into decision makers of smart cities interested in understanding the potential of VCS-based solutions. For any practitioners interested in replicating our findings, a replication package containing the software and materials to reproduce the case studies presented in this work is publicly available [16].

2. Related Works

Modern vehicles are being equipped with an ever-growing number of environmental sensors, mostly meant to improve comfort and safety for passengers and drivers [17]. In the near future, thanks to the introduction of even more advanced driver assistance systems (ADASs), supporting autonomous driving levels above SAE L2 [18], the number of contextual sensors per vehicle might rise above 200 [19]. Other than those related to powertrains, common vehicular sensors include cameras, Lidars, GPS receivers, and dedicated sensors to monitor air temperature, pollution and humidity, rain intensity, seat occupancy, and so on.

The possibility to share this sensed information with a remote backend, generically falling in the so-called telematics domain [20], has been investigated for decades. Nevertheless, only recent advancements in communications [21] are making it technically viable to share contextual information sensed by fleets of vehicles among each other or with a remote back-end infrastructure [22, 23]. As connected vehicles will constitute one of the most pervasive sensor networks in urban areas, vehicular crowdsensing (VCS) has the potential to foster the development of collective intelligence, or contextual awareness, to unprecedented levels [17].

The user involvement level is one of the key characteristics of VCS, and two main approaches have emerged: (1) participatory sensing, in which users need to actively participate in sensing, explicitly deciding when to collect and share data or (2) opportunistic sensing, in which no explicit user involved is required, and a sensing software may automatically collect data in an opportunistic fashion [24, 25].
Overall, VCS can provide a practical trade-off between deployment costs and sensing coverage and has thus been largely investigated, both in academic and industrial settings. Indeed, in a number of scenarios, VCS may lead to significant benefits over traditional monitoring techniques, including reduced data acquisition costs or the possibility of collecting data that were previously unavailable [26, 27].

Since most private vehicles are stationary in a parking space for up to 95% of the time, as observed by Ruth in [28], VCS studies presented in the literature mostly focused on exploiting high-mileage vehicles, such as buses, garbage trucks, or taxis [29]. In particular, taxi trajectories have often been used to investigate several urban phenomena of interest and extract noteworthy insights into urban dynamics. For example, Castro et al. in [30] investigated the possibility of using taxis as probe vehicles in VCS with the goal of modelling and predicting traffic conditions, while Mao et al. in [31] analysed more than 35 million taxi traces collected from approximately 9k taxis in the city of Shanghai in China to understand commuting patterns of urban dwellers.

The study conducted by Mathur et al. in [32] was the first to suggest the use of taxis to crowdsense on-street parking availability, using a dataset of taxi traces from about 500 taxis in San Francisco (USA) [33]. The spatiotemporal distribution of vehicles within urban areas was investigated by Bock et al. in [12], analysing the average daily time gaps between consecutive passing of potential probe vehicles over some road segments in San Francisco, using again the taxi traces recorded in [33]. This analysis showed that some streets might have been visited by a probe vehicle with a frequency in terms of minutes, thus enabling very dynamic VCS-based use cases. In contrast, for some of adjacent minor streets, the probing frequency increases remarkably, going up to many hours between two consecutive sensings. Li et al. in [34] investigated air pollution due to traffic emissions in the city of Beijing in China by analysing the trajectories of more than 12K taxis. Historic data on taxi trajectories have also been used to improve the charging efficiency of electric vehicles charging networks [3, 35]. In [36], Zhong et al. proposed an approach that leveraged gyroscope and accelerometer data from driver smartphones to detect and monitor potholes and road pavement degradation during vehicle trips. It is worth noting that not only data from motor vehicles have been exploited for VCS. For example, the authors of [37] investigated the usage of trajectories collected from bike-sharing system (BSS) users in Chicago, USA, to discover key bike-sharing stations, intending to optimize BSS planning.

More recently, Dokuz and Dokuz [38] proposed a novel approach to detect anomalies in daily traffic dynamics based on vehicular trajectories and leveraged a massive dataset of more than 60 million taxi trajectories collected in New York City to validate their proposal. In [39], Dokuz defined a new method, based on weighted spatiotemporal data mining, to estimate regional traffic conditions starting from massive datasets of vehicular trajectories. That work leveraged a dataset consisting of more than 80 million taxi trajectories, also collected in New York City.

Still, to the best of our knowledge, the number of studies investigating the feasibility of leveraging a fleet of high-mileage vehicles for VCS is limited, particularly in settings where road network topology is not grid-like, but rather complex and irregular. Indeed, most related studies are

Figure 1: Topology of the road networks of the cities of (a) Rome and (b) San Francisco.
focused on urban areas whose road network has a grid-based
topology, such as New York City, San Francisco, Shanghai,
and Chicago. In [13], Martino and Starace moved a first
in this direction, presenting a case study on the suitability
of a fleet of 100 taxis for VCS in the city of Porto in Portugal,
which is characterized by a very irregular road network
topology. In that work, the results showed that 100 taxis
could achieve significant road network coverage over one
month, but their visit frequency might not be enough to
adequately support some VCS use cases requiring high
sampling rates. That work showed that taxis can effectively
act as probes in VCS in cities with complex road networks,
but the considered setting was limited, and no insight was
given on the impact of the number of vehicles recruited for
VCS. This is a key factor for a decision maker to consider
when evaluating the feasibility of VCS-based use cases, as
recruiting few participants might lead to inadequate sensing
coverage, while recruiting too many of them might result in
wasting money.

3. Case Studies

To assess the potential of exploiting high-mileage vehicles as
probes for VCS in urban road networks with irregular to-

topology, like those of historical cities, two case studies were
conducted, leveraging two publicly available datasets of real-
world trajectories collected from taxis, respectively, in
the city of Porto in Portugal and that of Rome in Italy. In each
case study, to investigate the impact of the number of probe
vehicles on the achieved spatiotemporal coverage, first, the
entire fleet of taxis in the dataset was considered. Then,
additional repetitions of the experiments were conducted,
after considering only the trajectories of 50, 100, and 200
randomly subsampled taxis. Five repetitions of the case
study were performed for each sampling, to account for
oscillations due to the randomness in the taxi sampling
process. In the remainder of the paper, the average results
across these five repetitions are reported.

In the experiments, OpenStreetMap (OSM) [40] data
were used to represent the underlying road network. OSM
data are generally considered to be comparable in terms of
quality with authoritative datasets in urban areas [41]. The
trajectories in the datasets were matched to the OSM
representation of the road network. On top of this, several
spatiotemporal coverage statistics were computed, as dis-
cussed in the following, for each road segment class, as
defined by OSM (whose brief description is provided in
Table 1).

<table>
<thead>
<tr>
<th>Road segment class</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motorway</td>
<td>Major divided highways, typically with two or more lanes and restricted access</td>
</tr>
<tr>
<td>Trunk</td>
<td>The most important roads that are not motorways</td>
</tr>
<tr>
<td>Motorway link</td>
<td>Roads that serve as a connection to access motorways</td>
</tr>
<tr>
<td>Primary</td>
<td>Important thoroughfares, typically linking major cities</td>
</tr>
<tr>
<td>Primary link</td>
<td>Roads that serve as a connection to access primary roads</td>
</tr>
<tr>
<td>Secondary</td>
<td>Smaller roads, not qualifying as primary</td>
</tr>
<tr>
<td>Secondary link</td>
<td>Roads that serve as a connection to access secondary roads</td>
</tr>
<tr>
<td>Tertiary</td>
<td>Roads connecting local centres in larger cities</td>
</tr>
<tr>
<td>Tertiary link</td>
<td>Roads that serve as a connection to access tertiary roads</td>
</tr>
<tr>
<td>Residential</td>
<td>Roads connecting residential areas, often lined with houses</td>
</tr>
<tr>
<td>Living street</td>
<td>Residential streets that are often restricted to vehicular traffic</td>
</tr>
<tr>
<td>Service</td>
<td>Service roads that are used to access industrial estates, business parks, etc</td>
</tr>
<tr>
<td>Unclassified</td>
<td>Roads that have not been yet classified or not falling in any of the other categories</td>
</tr>
</tbody>
</table>

The pipelines to conduct the case studies were imple-
mented within the KNIME platform [42], leveraging
a custom extension, which is also freely available at the
GitHub repository https://github.com/luistar/knot. Software
and data to reproduce the case studies are made available to
the interested reader in the replication package [16]. In the
following, a detailed description of the adopted experimental
protocol is provided, in terms of employed data, procedure,
and considered spatiotemporal coverage metrics.

3.1. Datasets. The empirical evaluations are based on two
publicly available datasets of real taxi trajectories collected in
two different projects. The first dataset [43] has been col-
lected from 441 taxis operating in the city of Porto and
contains 1,710,671 trajectories spanning over one year, from
July 2013 to June 2014. Each trajectory is characterized as
a sequence of GPS positions and associated with a starting
timestamp. The second one [44], on the other hand, was
collected from 315 taxis in the city of Rome, over a one-
month period starting from February 2014, and consisted of
more than 20 million distinct GPS positions, corresponding
to approximately 70k vehicular trajectories. Note that, since
the two datasets span over significantly different time
frames, as better detailed in the Empirical Procedure de-
scription, for both the datasets, the three contiguous weeks
containing the greatest number of taxi trajectories were
selected. It is worth noting that both the datasets include
only a fraction of the total number of taxis in these cities. For
example, in Rome, there are about 3,000 taxi licences. The
GPS points contained in the datasets are depicted in Fig-
ure 2, where each GPS position is represented as a
black point.

Note that, even though these datasets are not very recent,
there have been only minimal changes in the road networks
of the considered cities since they were recorded. In many
European countries, there have been no disruptions in taxi
mobility dynamics, since transportation network companies
TNCs like Uber or Lyft are banned. Nevertheless, the findings presented in this work could also be applied in scenarios where TNC vehicles also act as probes in presence of appealing rewarding mechanisms. Therefore, it is reasonable to assume that the obtained results are still representative of the spatiotemporal coverage achievable by fleets of high-mileage vehicles in historical cities.

3.2. Empirical Procedure. The employed empirical procedure is formalized in Algorithm 1 and described as follows: The input of the process is a dataset containing taxi trajectories.

As a first step, preliminary filtering of taxi trajectories was performed based on spatial and temporal criteria (lines 2–8). In particular, only the taxi trajectories matching the following criteria were retained: (1) they are entirely contained within the urban area of the considered city, as defined in OpenStreetMap, and (2) they are recorded in a considered three-week timespan. The investigation was restricted to urban areas because most potential VCS use cases involve urban environments [9], and the taxis mostly operated in the urban areas of the considered cities.

As for the temporal filtering step, its rationale is to make the current results comparable to those reported in other studies, such as those conducted by Bock et al. in [12], and allow for future replications on additional datasets, which typically contain trajectories collected over briefer time spans, such as the one from San Francisco presented in [33]. Indeed, replicating existing studies to assess whether the results of a previous investigation can be reproduced in new contexts with different data is critical for building a cumulative and wider body of research knowledge [45].

More in detail, the three contiguous weeks containing the most trajectories were selected for the Porto dataset, namely, the ones between 2014/05/02 and 2014/05/22. For the Rome dataset, the first three weeks in the dataset were selected, namely, the ones from 2014/02/01 to 2014/02/21. After this spatiotemporal filtering, approximately 100k trajectories were retained for the Porto dataset and more than 65K for the Rome one. The difference in the figures can be explained by the fact that the Porto dataset contains data from 440 taxis, while the Rome one includes data from 315 vehicles.

As for the logical representation of the road network on which to conduct the coverage analysis, the empirical evaluation leveraged freely available data from the OpenStreetMap project (see Line 9, Algorithm 1), which is generally considered qualitatively comparable to authoritative datasets in urban areas [40, 41]. Table 2 reports some statistics on the considered OSM datasets for investigated urban areas. In particular, for both Porto and Rome and for the main road types considered in the OSM standard (https://wiki.openstreetmap.org/wiki/Key:highway), the absolute number of road segments and the corresponding percentage of the total are reported. Note that the road types “service” and “unclassified” were excluded from the investigation, since, as reported in the OSM standard, these types of segments, which generally correspond to urban parks or industrial estates, might not be accessible to general vehicular traffic. Similarly, “living street” segments were also excluded, as they are generally not accessible to public traffic.
and are mostly used by pedestrians and cyclists. As for the “Motorway,” “Trunk,” “Primary,” “Secondary,” and “Tertiary” (also with corresponding links) segments, they were retained for the analysis. Indeed, even though these segments represent a small fraction of the overall road network, they still account for dozens of thousands of segments, thus being in our opinion worth being investigated to understand VCS dynamics.

Subsequently, the map-matching procedure is performed, which is a preliminary step necessary to compute accurate road network coverage metrics. Indeed, it is worth noting that the GPS positions in the considered datasets are inherently affected by positioning errors [10], as highlighted in Figure 3.

Thus, the goal of map matching (see Line 10, Algorithm 1) is to align raw vehicular trajectories, which consist of series of possibly inaccurate GPS points, with the considered OSM road network. The employed map-matching procedure is detailed in Algorithm 2 and leverages OSRM (Open Source Routing Machine) [46], a well-known state-of-the-art routing solution, widely used in other empirical studies such as the study by Singh et al. presented in [47].

In particular, for each trajectory in the datasets, a query is performed to an instance of OSRM, requesting a route traversing all the GPS positions in the trajectory in the same order (see Line 5, Algorithm 2). If such a route exists, the instance of OSRM returns a sequence of OSM road segments that were traversed with visit timestamps, which is the map

### Algorithm 1: Experimental procedure.

#### Table 2: Road segments in the considered maps.

<table>
<thead>
<tr>
<th>Road segment class</th>
<th>Porto</th>
<th>Percentage of total (%)</th>
<th>Rome</th>
<th>Percentage of total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of segments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motorway</td>
<td>2,461</td>
<td>1.7</td>
<td>2,575</td>
<td>0.7</td>
</tr>
<tr>
<td>Trunk</td>
<td>1,036</td>
<td>0.7</td>
<td>6,276</td>
<td>1.8</td>
</tr>
<tr>
<td>Motorway link</td>
<td>4,445</td>
<td>3.1</td>
<td>6,964</td>
<td>2.0</td>
</tr>
<tr>
<td>Primary</td>
<td>7,936</td>
<td>5.5</td>
<td>13,329</td>
<td>3.8</td>
</tr>
<tr>
<td>Primary link</td>
<td>607</td>
<td>0.4</td>
<td>2,042</td>
<td>0.6</td>
</tr>
<tr>
<td>Secondary</td>
<td>10,692</td>
<td>7.4</td>
<td>26,143</td>
<td>7.4</td>
</tr>
<tr>
<td>Secondary link</td>
<td>701</td>
<td>0.5</td>
<td>1,318</td>
<td>0.4</td>
</tr>
<tr>
<td>Tertiary</td>
<td>14,271</td>
<td>9.8</td>
<td>49,035</td>
<td>14.0</td>
</tr>
<tr>
<td>Tertiary link</td>
<td>322</td>
<td>0.2</td>
<td>1,971</td>
<td>0.6</td>
</tr>
<tr>
<td>Residential</td>
<td>66,657</td>
<td>45.9</td>
<td>141,672</td>
<td>40.3</td>
</tr>
<tr>
<td>Living street</td>
<td>4,075</td>
<td>2.8</td>
<td>448</td>
<td>0.1</td>
</tr>
<tr>
<td>Service</td>
<td>26,335</td>
<td>18.1</td>
<td>84,440</td>
<td>24.0</td>
</tr>
<tr>
<td>Unclassified</td>
<td>5,735</td>
<td>3.9</td>
<td>15,022</td>
<td>4.3</td>
</tr>
<tr>
<td>Total</td>
<td>145,273</td>
<td>100</td>
<td>351,235</td>
<td>100</td>
</tr>
</tbody>
</table>
matching of the current trajectory. If, on the other hand, no such route exists (i.e., mapMatchedTrajectory is nil at Line 7, Algorithm 2), the trajectory is discarded as unfeasible. An approach to map matching similar to the one adopted in the present study was recently presented by Saki and Hagen in [48]. In that study, the authors reported that the solution was able to correctly map match approximately 95% of the input trajectories. Similar success rates (≈ 96%) were also observed in the map-matching process for both the considered dataset.

After map matching, to investigate the impact of the number of involved probe vehicles on the achieved spatiotemporal coverage, different analyses are performed, varying the number of considered taxis (see Lines 11–17, Algorithm 1). In particular, the spatiotemporal coverage achieved by 50, 100, 200, and the maximum number of taxis in the dataset is investigated. For each considered number of taxis \( n \), first, a random sample of \( n \) taxis is produced, retaining only the map-matched trajectories belonging to the sampled vehicles (Lines 13–14, Algorithm 1). To account for the randomness introduced by the taxi sampling process, five repetitions were performed for each of the considered numbers of taxis. For each repetition, spatial and temporal coverage metrics (see Line 15, Algorithm 1) based on the selected trajectories were computed.

To investigate the spatial coverage achievable by the fleet of taxis, the number of times each road segment was crossed by one of the taxis in the considered time span was computed. As for temporal coverage, the average time gap between consecutive visits on each road segment was computed, as conducted also by Mathur et al. in [32]. In greater detail, assuming that a segment is traversed by \( n \) vehicles at the times \( t_1, \ldots, t_n \), respectively, with \( t_i \leq t_{i+1} \) for all \( i \in [1, \ldots, n-1] \), the average time gap between consecutive visits can be computed as

\[
\frac{\sum_{i=1}^{n-1} (t_{i+1} - t_i)}{n-1}.
\]

Starting from these coverage metrics, also, additional spatiotemporal metrics were computed. In particular, for spatial coverage, the percentage of road segments that were visited at least once in the considered three-week timespan was computed, and a similar analysis was performed for the main OSM road types found in the central area of Porto and Rome, which are reported in Table 2. Similar aggregation was also computed for temporal coverage metrics,

![Figure 3: Detail of the considered datasets, highlighting the inherent inaccuracy of GPS positions. Each recorded GPS position is represented by a black point.](image-url)
<table>
<thead>
<tr>
<th></th>
<th>Porto</th>
<th>Rome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50 taxis (%)</td>
<td>100 taxis (%)</td>
</tr>
<tr>
<td>Covered road segments</td>
<td>64</td>
<td>72</td>
</tr>
</tbody>
</table>

Table 3: Overall road network coverage percentage achieved by different-sized fleets of taxis.
aggregating by OSM road types and reporting the median for each of the categories.

Finally, for each of the considered number of taxis, the average of the metrics obtained in each repetition was computed.

### 4. Results and Spatiotemporal Coverage Discussion

In this section, the results of the conducted case studies are presented and discussed, highlighting similarities and differences with respect to the findings of similar studies, but conducted in cities with a grid-like road network topology. The description first focuses on the spatial coverage and then reports on the temporal coverage results.

#### 4.1. Spatial Coverage

The overall spatial coverage results of the empirical investigations, in terms of the percentage of road segments that are visited at least once by one of the taxis in the evaluated three-week period, are reported in Table 3.

Moreover, to highlight coverage trends depending on the type of roads, in Table 4, the details of spatial coverage results for each road segment class are reported.

These numbers highlight that as few as 50 taxis can achieve remarkable spatial coverage in both Porto and Rome, with a significant portion of the urban road network (64% in Porto and 47% in Rome) being visited at least once over the considered three-week period. The coverage varies significantly across different road types. When considering major road types such as motorways, primary, secondary, and trunks, about 80–90% of the respective road segments are sensed at least once, even with as few as 50 vehicles. When considering minor road types, such as residential and tertiary, the coverage is generally much lower, with approximately only half of the residential road segments being visited in Porto and only 27% of them being visited in Rome. Such variability in coverage rates between major and minor road segments has also been observed in other studies conducted on real vehicular trajectories from cities exhibiting a regular, grid-like structure [14, 49] and can be explained by the fact that vehicles do not distribute uniformly over the road network but rather tend to concentrate over the main city thoroughfares to reach their destination efficiently. Thus, it can be expected that those major streets (which are generally fewer than the minor ones) will indeed be visited more frequently by vehicles.

Moreover, as expected, increasing the number of involved vehicles leads to improvements in spatial coverage. As highlighted by Figure 4, however, such an improvement is not linear with respect to the number of vehicles, but rather the coverage percentage increases sublinearly.

More in detail, when considering the maximum number of taxis in the datasets (i.e., 440 taxis in Porto and 315 in Rome), the main road types are almost entirely visited (more than 95% of coverage) at least once in the three-week period, with improvements with respect to the 50 taxi scenarios ranging from 5% to 10%. As for minor roads, they are the ones that benefit the most, in terms of coverage, from the increase in taxi fleet size. For example, 78% (resp., 54%) of residential road segments are visited at least once in Porto (resp., Rome) when considering the maximum number of taxis, with improvements with respect to the 50 taxi scenarios going up to 30%.

The results also show that the spatial coverage achievable in Rome is in general lower than that achievable in Porto. This is due to the fact that the Rome road network is much bigger, containing more than twice the road segments of the Porto one, and hence is more difficult to cover.

These spatial coverage results appear to be generally comparable with those reported in other studies investigating the feasibility of using high-mileage vehicles for VCS in cities with a grid-like road network topology, such as the one presented by Di Martino and Starace in [14]. In particular, that work investigated the coverage achieved by a fleet of 500 taxis over a three-week period in the city of San Francisco, which features a regular road network (see Figure 1). Even though no investigation on the influence of taxi fleet size on the achieved spatial coverage is conducted in that work, the results achieved using the maximum number of taxis can be roughly compared with ours. This comparison shows that, for the main road types, a fleet of >300 taxis can achieve almost complete coverage (>95%) in three weeks, regardless of the road network topology. As for minor road segments, experiments in historical cities, with an irregular road network, show generally lower coverage rates. For example, only 54% of residential road segments are covered by 315 taxis in Rome, whereas 78% of residential road segments are covered in San Francisco.

<table>
<thead>
<tr>
<th>Road segment class</th>
<th>50 taxis (%)</th>
<th>100 taxis (%)</th>
<th>200 taxis (%)</th>
<th>440 taxis (%)</th>
<th>50 taxis (%)</th>
<th>100 taxis (%)</th>
<th>200 taxis (%)</th>
<th>315 taxis (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motorway</td>
<td>94.7</td>
<td>95.7</td>
<td>99.5</td>
<td>99.7</td>
<td>88.7</td>
<td>96.5</td>
<td>97.3</td>
<td>98.4</td>
</tr>
<tr>
<td>Trunk</td>
<td>93.9</td>
<td>98.1</td>
<td>98.7</td>
<td>99.1</td>
<td>83.4</td>
<td>89.9</td>
<td>95.0</td>
<td>94.7</td>
</tr>
<tr>
<td>Motorway link</td>
<td>82.5</td>
<td>87.2</td>
<td>91.4</td>
<td>95.7</td>
<td>44.2</td>
<td>67.1</td>
<td>81.7</td>
<td>88.7</td>
</tr>
<tr>
<td>Primary</td>
<td>89.8</td>
<td>94.4</td>
<td>96.2</td>
<td>97.5</td>
<td>90.5</td>
<td>95.3</td>
<td>97.7</td>
<td>97.8</td>
</tr>
<tr>
<td>Primary link</td>
<td>92.0</td>
<td>95.6</td>
<td>96.8</td>
<td>98.3</td>
<td>77.9</td>
<td>89.0</td>
<td>91.9</td>
<td>94.8</td>
</tr>
<tr>
<td>Secondary</td>
<td>89.2</td>
<td>92.8</td>
<td>95.5</td>
<td>97.2</td>
<td>81.8</td>
<td>85.9</td>
<td>92.5</td>
<td>94.4</td>
</tr>
<tr>
<td>Secondary link</td>
<td>73.6</td>
<td>75.5</td>
<td>77.6</td>
<td>79.3</td>
<td>75.3</td>
<td>82.0</td>
<td>93.3</td>
<td>94.9</td>
</tr>
<tr>
<td>Tertiary</td>
<td>80.5</td>
<td>86.2</td>
<td>91.2</td>
<td>94.4</td>
<td>65.7</td>
<td>75.2</td>
<td>85.8</td>
<td>91.1</td>
</tr>
<tr>
<td>Tertiary link</td>
<td>76.8</td>
<td>87.3</td>
<td>91.8</td>
<td>93.7</td>
<td>53.4</td>
<td>65.0</td>
<td>82.1</td>
<td>81.0</td>
</tr>
<tr>
<td>Residential</td>
<td>50.6</td>
<td>60.2</td>
<td>69.9</td>
<td>78.1</td>
<td>27.0</td>
<td>35.2</td>
<td>46.7</td>
<td>54.3</td>
</tr>
</tbody>
</table>

Table 4: Road network coverage percentage achieved by different-sized fleets of taxis in the considered scenarios.
4.2. Temporal Coverage. The results of the temporal coverage analyses are reported in Table 5 and Figure 5.

As observed with spatial coverage, temporal coverage also exhibits sensible differences among different road types, with major road segments being visited generally more frequently than those belonging to minor road types. Moreover, as expected, an increase in the number of vehicles involved in the VCS leads to noticeable reductions in the time gaps between subsequent visits. More in detail, when considering only 50 taxis in Porto in the three-week period, the median time gaps range from about 12 hours for motorway segments to 43 hours for residential ones. When increasing the number of taxis to 440, the median time gaps for motorway segments go down to less than 2 hours, while for residential segments, they are reduced to about 28 hours. Similarly, in Rome, 50 taxis achieve a 26-hour median time gap on segments belonging to primary roads and a 51-hour median time gap for residential segments. When increasing the number of involved taxis to 315, the median time gaps for primary segments are reduced to slightly more than 7 hours, while the median time gaps between subsequent visits on residential segments decrease to about 38 hours. Again, as highlighted by the trends in Figure 5, the median time gaps decrease sublinearly with respect to the number of involved vehicles. Moreover, the improvement (i.e., decrease) in time gaps due to increasing taxi fleet size is generally greater for the main road types than it is for minor roads. This is explained by the fact that fewer main roads exist and are largely covered even by smaller fleets of taxis (see Table 4). Thus, increasing fleet size leads to smaller improvements in spatial coverage for these kinds of segments but to larger improvements in temporal coverage, as a greater number of vehicles traverse the same main road segments. On the other hand, as previously noted in 10, when increasing the taxi fleet size, the improvement in spatial coverage is greater for minor road segments. This means that a greater part of the additional vehicles visits minor road segments that were never visited when considering smaller fleet sizes, leading to smaller improvements in temporal coverage.

Moreover, the temporal coverage results are worse in Rome, with significantly higher median time gaps between subsequent visits than in Porto. This is mainly due to the fact that Rome features a much bigger road network than Porto (see Table 2), and hence, it is not possible for a comparable number of vehicles to visit its road segments as frequently).

These temporal coverage results computed over the entire three-week period can be roughly compared with the ones presented by Di Martino and Starace in [14], which used the same metric to assess the temporal coverage achievable by a fleet of taxis in San Francisco. The comparison suggests that swarms of probe vehicles in cities with a grid-like road network can generally achieve noticeably better temporal coverage, regardless of the road segment type. For example, the median time gaps for primary segments in San Francisco, as reported in [14], are only 40 minutes, while the largest considered fleets in Porto (resp., Rome) achieve a median time gap of 5 hours (resp., 7.2 hours). Similarly, the median time gaps for residential road segments reported in [14] in San Francisco are 24 hours, while for Porto and Rome, the largest considered
To gain a better insight into the temporal coverage dynamics and make the current analysis more comparable with other previous investigations conducted in cities featuring a grid-like road network, like [10, 12], the average time gaps between subsequent visits on a road segment were also computed on a daily basis, i.e., by considering only visits happening in the same day. The results of these additional analyses are reported in Table 6.

Furthermore, to investigate daily traffic dynamics and their impact on coverage, daily average time gaps were also computed in four selected time slots, namely, the ones from midnight to 7.59, from 8.00 to 13.59, from 14.00 to 19.59, and from 20.00 to 23.59, which are referred to as follows: 00–08, 08–14, 14–20, and 20–24, respectively. These fine-grained results for the cities of Porto and Rome are reported in Tables 7 and 8.

These results can be indicatively compared against those reported in [12], which investigated the feasibility of using a fleet of 486 taxis to crowdsense on-street parking availability in San Francisco, using average time gaps computed on a daily basis as a key metric. Even though the analyses presented in [12] focused only on a small part of the urban road network of San Francisco, a comparison with the results presented in that paper highlights substantial differences in the temporal coverage achievable by fleets of taxis between cities featuring a regular road network topology and cities with an irregular one. Indeed, as reported by Bock et al. [12], the average daily time gaps achieved by the fleet of taxis in San Francisco on primary road segments amount to just 11 minutes. In Porto (resp., Rome), the results showed that
the average daily time gaps for primary segments achieved by the largest considered fleets of 440 (resp., 315) taxis amounted to 1.1 hours (resp., 2.3 hours). For minor road segments such as residential ones, the differences are even more significant, with average daily time gaps being smaller than 1 hour in San Francisco while amounting to 11.1 and 18.2 hours in Porto and Rome, respectively.

As for the time-slot analyses, the results show that taxis in Porto and Rome are mainly active during day hours (from 8.00 to 20.00), leading to better temporal coverage during those hours, similar to what was observed in San Francisco by Bock et al. in [12]. Nevertheless, even when considering only peak traffic hours, the temporal coverage achieved by the considered fleets of taxis in Rome and Porto is still significantly worse than the one achieved by a comparable fleet of taxis in San Francisco, as reported by Bock et al. [12].

The results of the conducted temporal coverage analyses highlight that, in both the considered urban scenarios, 50 or 100 taxis might be adequate to support only VCS use cases needing lower sensing frequencies, since for a significant

<table>
<thead>
<tr>
<th>Table 6: Average time gaps (in hours) in a single day per road segment type.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Road segment class</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Motorway</td>
</tr>
<tr>
<td>Trunk</td>
</tr>
<tr>
<td>Motorway link</td>
</tr>
<tr>
<td>Primary</td>
</tr>
<tr>
<td>Primary link</td>
</tr>
<tr>
<td>Secondary</td>
</tr>
<tr>
<td>Secondary link</td>
</tr>
<tr>
<td>Tertiary</td>
</tr>
<tr>
<td>Tertiary link</td>
</tr>
<tr>
<td>Residential</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 7: Average time gaps (in hours) in selected time slots per road segment type in Porto.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Road segment class</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Motorway</td>
</tr>
<tr>
<td>Trunk</td>
</tr>
<tr>
<td>Motorway link</td>
</tr>
<tr>
<td>Primary</td>
</tr>
<tr>
<td>Primary link</td>
</tr>
<tr>
<td>Secondary</td>
</tr>
<tr>
<td>Secondary link</td>
</tr>
<tr>
<td>Tertiary</td>
</tr>
<tr>
<td>Residential</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 8: Average time gaps (in hours) in selected time slots per road segment type in Rome.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Road segment class</strong></td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>Motorway</td>
</tr>
<tr>
<td>Trunk</td>
</tr>
<tr>
<td>Motorway link</td>
</tr>
<tr>
<td>Primary</td>
</tr>
<tr>
<td>Primary link</td>
</tr>
<tr>
<td>Secondary</td>
</tr>
<tr>
<td>Secondary link</td>
</tr>
<tr>
<td>Tertiary</td>
</tr>
<tr>
<td>Tertiary link</td>
</tr>
<tr>
<td>Residential</td>
</tr>
</tbody>
</table>

12 Journal of Advanced Transportation
portion of the road network, the interval between consecutive probe vehicle traversals is in the range of many hours or even days. This is anyhow still enough for services such as pothole [50] or air quality monitoring [51].

With larger fleet sizes, the situation becomes more interesting. In particular, the full fleet of 440 taxis is enough to monitor most of the urban road segments hourly, which could enable many more VCS-based services, such as heat island monitoring [52]. The situation in Rome is worse, mostly because the fleet of considered vehicles is smaller, while the total road network is far more extended. Consequently, even with the full fleet, the best case is to have road segments monitored every couple of hours. Still, it is worth noting that the number of vehicles considered for these experiments is a fraction of the total number of available taxis. For Rome, the considered dataset contained data from about one-tenth of the total fleet of taxis. Thus, in presence of advantageous incentivization mechanisms, it is easy to envision a scenario in which the spatiotemporal coverage of these potential probe vehicles can be significantly improved.

5. Conclusions

A great deal of research has been conducted to investigate the feasibility of leveraging high-mileage vehicle fleets in vehicular crowdsensing (VCS) scenarios, to collect contextual data in urban areas in an economically effective way. Still, most of these works either focused on urban road networks featuring a regular, grid-based topology or excessively simplified the underlying road network by abstracting it with a set of coarse-grained areas.

This work investigated a new setting, by presenting two case studies aimed at investigating the feasibility of leveraging fleets of high-mileage vehicles, such as taxis, on topologically different urban road networks from two historical European cities, and at assessing the impact of the number of vehicles recruited in VCS on the achieved coverage. In particular, the study analysed the potential spatiotemporal sensing coverage achievable by different-sized fleets of taxis and computed over real trajectories from different-sized fleets of taxis in the cities of Porto (Portugal) and Rome (Italy). The results showed that as few as 50 taxis can be adequate to achieve the basic spatial coverage of the road network in the urban areas of both Porto and Rome, enabling a number of possible use cases. When increasing the number of vehicles, spatial coverage improves, even though sublinearly, up to reaching almost full coverage of the road network in both cities in the considered three-week timespan. As for temporal coverage, the results showed that the sensing frequency when leveraging just 50 taxis is probably inadequate to support many VCS scenarios, requiring more frequent sampling. When increasing the number of vehicles taking part in crowdsensing, however, the time gaps between subsequent visits improve remarkably. Still, the temporal coverage achievable by taxis in historical cities remains significantly lower than that achievable in cities featuring a more regular road network and appears to be generally insufficient to support VCS use cases requiring very frequent probing. Hence, there is room for the introduction of incentivization mechanisms [15] or smarter routing algorithms [49], which could also help achieve better sensing coverage.

Future works could conduct further case studies involving also Asiatic cities such as Beijing, for which a well-known taxi trajectory dataset exists as well [53], to provide useful insights into the suitability of using taxis as probes in VCS scenarios. Moreover, investigating the spatiotemporal coverage that could be achieved by fleets of taxis in additional time frames, such as, for example, one day, three days, one week, and two weeks, could also provide useful and valuable insights.

Data Availability

The software and materials to reproduce the case studies presented in this paper are publicly available in the Zenodo repository at the DOI https://doi.org/10.5281/zenodo.7408877.

Disclosure

This work was performed as part of the employment of the authors at the University of Naples Federico II, Naples, Italy.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


