Geoinformatica (2022) 26:677–706
https://doi.org/10.1007/s10707-021-00460-z

An analysis of twitter as a relevant human mobility proxy

A comparative approach in spain during the COVID-19 pandemic

Fernando Terroso-Saenz[1] · **Andres Muñoz**[2] · **Francisco Arcas**[1] · **Manuel Curado**[1]

Received: 3 June 2021 / Revised: 16 November 2021 / Accepted: 21 December 2021 /
Published online: 15 February 2022

Abstract
During the last years, the analysis of spatio-temporal data extracted from Online Social Networks (OSNs) has become a prominent course of action within the human-mobility mining discipline. Due to the noisy and sparse nature of these data, an important effort has been done on validating these platforms as suitable mobility proxies. However, such a validation has been usually based on the computation of certain features from the raw spatio-temporal trajectories extracted from OSN documents. Hence, there is a scarcity of validation studies that evaluate whether geo-tagged OSN data are able to measure the *evolution* of the mobility in a region at multiple spatial scales. For that reason, this work proposes a comprehensive comparison of a nation-scale Twitter (TWT) dataset and an official mobility survey from the Spanish National Institute of Statistics. The target time period covers a three-month interval during which Spain was heavily affected by the COVID-19 pandemic. Both feeds have been compared in this context by considering different mobility-related features and spatial scales. The results show that TWT could capture only a limited number features of the latent mobility behaviour of Spain during the study period.

Keywords Human mobility · Spatio-temporal knowledge processing · Online social networks · COVID-19

1 Introduction

In this globalization era, human mobility across the globe has become much more complex, dynamic and heterogeneous than ever before. This calls for advanced techniques able to to understand where, when and how people move in fields like road-traffic management [33], city planning [38] or demographics [9].

The mobility mining discipline has evolved from the development of travel surveys aimed to capture human-movement distributions towards the massive adoption of information and communication technologies (ICTs), fostering the collection of mobility data in

✉ Fernando Terroso-Saenz
fterroso@ucam.edu

[1] High Polytechnic School, Catholic University of Murcia, Murcia, Spain

[2] Department of Computer Engineering, University of Cádiz, Puerto Real (Cádiz), Spain

a much larger and detailed scale. Hence, GPS spatio-temporal trajectories [28], mobile-phone location data [12] and smart-card transactions [39] are widely used as data sources to extract different types of human-mobility patterns. However, these sources often have limited access for the scientific community due to commercial and privacy issues [36].

At the same time, last decade has witnessed the emergence of Online Social Networks (OSNs) like Twitter or Instagram as prominent human-location feeds due to their capability to geo-tag users' posts. Unlike the aforementioned datasources, OSN-location data can be easily retrieved from Application Programming Interfaces (APIs). These data have been used in several types of mobility-related analyses at different spatial scales and scenarios [8, 23, 29, 34].

However, it is also true that OSN-location data suffer from serious data-sparsity problems in its spatial and temporal dimensions due to their volunteer nature [27]. For that reason, it is required to asses the actual feasibility of OSNs as reliable proxies for estimating the human-location information of a geographical region. In this line, most of the current proposals within the OSN-based human mobility mining discipline validate their results by just observing certain inner features of the extracted mobility patterns such as the radius of gyration or the probability of return [15, 20]. Nevertheless, the following limitations are observed in the existing studies:

- First, these approaches usually focus on analyzing the mobility during stationary time periods in which human mobility patterns exhibit a large regularity. However, they neglect the capability of OSN to actually detect human-mobility dynamic behaviours in a region (e.g., the increment or decrement of trips in real time). Nowadays, the human flows of a region might dramatically increase or decrease with respect a regular patterns due to multiple factors like lockdown policies in a pandemic scenario.
- Existing studies usually focus on a particular spatial scale for their analysis (e.g., at a city level). However, the mobility of a region might be modelled at different scales depending on the utility setting (e.g., human flows at street level vs. human mobility among cities). Thus, it is necessary to study whether the mobility patterns extracted from a particular OSN feed remains consistent among different spatial levels.
- Lastly, most of the current analysis rely on the raw spatio-temporal trajectories composed from the target OSN feed. Nonetheless, alternative mobility formats like Origin-Destination (OD) matrices or time series have not be fully considered.

Given these limitations, the goal of this work is to assess whether Twitter (TWT) could be considered a reliable datasource to capture abnormal variations of human displacements at different spatial granularities in a similar manner to a dataset based on mobile phone records. This makes the comparison much more challenging and allows us to evaluate the reliability of Twitter as a mobility proxy in such an abnormal setting. In that sense, there is already a large number of works that focus on evaluating the mobility captured by Twitter in regular conditions [19, 20, 24]. However, our work intends to enlarge such an evaluation to new conditions where such a regular and normal mobility did not occur.

To achieve this goal, we propose a comprehensive comparison between an official study about the human mobility in Spain developed by the Spanish National Institute of Statistics (INE)[1], which is based on a nation-wide mobile-phone location dataset, with a social media dataset crawled from TWT covering the same spatial and temporal dimensions. More in detail, both datasets cover a 3-month period from April to June, 2020. During these months,

[1]Instituto Nacional de Estadistica, https://www.ine.es/index.htm

the human mobility in Spain was seriously constrained by lockdown policies taken by local governments due to the COVID-19 pandemic. Additionally, this comparative analysis takes into account the features related to OD matrices, time and graph-based representations extracted from the mobility patterns. Thus, the term *abnormal variation of human displacements* mentioned above referred that the change in the number of trips through time were not mainly caused by the commuting routines of a population but due to the evolution of an absolutely exceptional pandemic.

The rest of the paper is structured as follows. Section 2 provides an overview about validation methods for social media data in the human-mobility scope. Section 3 describes the proposed validation study and Section 4 puts forward its results in detail. Finally, the main conclusions and the future work are highlighted in Section 5.

2 Related work

This section reviews how OSN data have been used and validated in domains and applications related to human mobility (see Table 1).

Some of the early works that explored the use of TWT as a proxy for human mobility focused on the study of demography. Thus, some works elsewhere explored how TWT was able to capture changes in the demographic characteristics (e.g., age, gender, and ethnic groups) of human mobility patterns from a city level [25] to a worldwide level [30].

TWT has been also used to better understand natural events related to human mobility. For example, the work in [6] focuses on forecasting dengue disease transmission by studying the public transport data, related Twitter posts and the incidence of such a disease in tropical areas. Similarly, the work in [1] aims to detect natural disaster situations by means of TWT posts without the need for geo-tagged data. The authors rely instead on the use of other features such as posting time, publication frequency and hashtag trends. However, no validation methods are used to test the reliability of this proposal.

More specifically related to this work, it is possible to find comprehensive studies that evaluate TWT as a reliable source for travel prediction at different spatial scales based on human mobility mining. At the city level, [20] calculated different mobility features like the radius of gyration or the distribution of the distance-based displacements from a TWT dataset comprising 10,000 unique users in New York. Then, authors analyzed the flows among the city boroughs obtained from TWT with the ones from different travel surveys and transportation councils. The results showed that the similarity among flows was quite high for Manhattan but less remarkable for the other boroughs. At the same level, [29] proposed the integration of TWT data with census tracts and LEHD Origin-Destination Employment Statistics (LODES) for New York city to estimate its commuting trip distribution. The results showed that the aggregation of TWT data to the model created from the more traditional data along with the use of the Random Forest technique yielded the best performance for developing dynamic models of trip distribution.

Regarding the study of mobility patterns at the inter-urban level, one of the first studies [24] analyzed the use of a check-in social network application in 370 Chinese cities, involving half a million users. The results showed that the inter-urban displacements in those cities followed an exponential distribution with a decay effect depending on the size and location of the cities. In a more recent study [16], the Weibo social media platform was used to explore the correlation between the mobility captured by this OSN and the inter-urban networks of population flow extracted from survey data in China. The results showed

Table 1 Comparison of mobility studies based on OSN. The number in brackets in the *Location* column indicates the number of entities under study. Acronyms: MPL: mobile-phone location data; PTD: Public Transport data

Ref.	Goal	Datasources	Granularity level	Location Setting	Validation method	COVID-19 related
[5]	Predict the spread of SARS-CoV-2 through the analysis of geolocated tweets	TWT	Inter-countries	Worldwide	none	yes
[4]	Prediction of the spatiotemporal spread of infectious diseases using social media data	TWT	Inter-city	China	none	yes
[6]	Early epidemic predictions	PTD	Intra-city	Fortaleza (Brazil)	OD and time series comparison	no
[29]	Estimation of commuting trip distribution	TWT & census tracts	Intra-city	New York	none	no
[1]	Detection of population's spatio-temporal and demographic features	TWT	Intra-city	Chicago	Census and activity centers data	no
[25]	Natural disaster detection	TWT	Inter-cities	N/A	none	no
[30]	Spatial distribution of the world-wide population	TWT	Inter-countries	Worldwide	Statistical methods	no
[20]	Human mobility study	TWT & travel surveys	Intra-city	New York	Statistical methods	no
[19]	Human mobility study	TWT & MPL	Nation-wide	Australia	Statistical methods	no
[17]	Human mobility study	TWT & MPL	Nation-wide	USA	Statistical methods	yes
[15]	Human mobility study	TWT & demographics	Worldwide	-	Statistical methods	no
[31]	Use of telephone antenna signals to establish human relationships between geographic areas	MLP	Inter-cities	Chile	Graph Clustering	no
Our approach	Human mobility study	TWT & MPL	Nation-wide	Spain	OD and time series comparison	yes

a relevant correlation between these two sources, again with some dependence with respect to factors such as distance or scarcity of connections among cities.

The study in [19] analysed a dataset at the national level comprising more than 4 million geo-tagged tweets in Australia. From the underlying trajectories, different human mobility features were extracted, like the radius of gyration, the probability of return or the displacement distribution. The results showed that there was a large heterogeneity in the movement captured by TWT as different statistical distributions were used to model each of these parameters. Moreover, it was observed certain similarities between the TWT dataset and previous studies based on call-data records regarding, for example, home-return patterns. This study also evaluated the location predictability using TWT by means of two entropy metrics. It was shown that TWT users could be grouped in two different mobility types, a highly predictable one and another exhibiting a much more diverse, and therefore less predictable behaviour.

Hawelka et al. [15] provided a suitability study of TWT at a worldwide scale. In this case, authors studied the correlation among certain mobility features (e.g. radius of gyration) and the demographics of certain countries. To validate some of their results, a comparison between the TWT inter-country flows and tourist statistics from the World Economic Forum was performed. Results showed a quite strong correlation with R^2 above 0.8.

Due to the COVID-19 pandemic, some works has recently used TWT location data to infer human mobility patterns so as to better understand and predict the epidemic evolution of the virus. In that sense, the proposal in [5] analysed geo-tagged TWT data from 2013 to 2015 to describe the geographical spreading of the virus SARS-CoV-2. In particular, its goal was to predict human mobility patterns in and outside of China so as to forecast the spatio-temporal spread of the COVID-19 virus. In the reported results, it was possible to observe a high correlation between country-level TWT user visits and reported COVID-19 cases. Similarly, in [4] authors compared the human mobility patterns estimated by analyzing geolocated tweets from 2013 to 2015, comprising trips from 161 Twitter users from Wuhan (China) travelling to other Chinese cities during the first few weeks of the COVID-19 outbreak. Besides, authors in [17] evaluated the differences among four human-mobility feeds extracted from different mobile-phone operators (e.g. Google, Apple) and TWT during the COVID-19 pandemic. The goal was to assess whether the four mobility sources were able to capture the changes in the human mobility patterns from March to June, 2020 in USA. The results showed that TWT exhibited a *reducing-and-recovering* pattern similar to the other sources with a Pearson Correlation Coefficient (PCC) ranging from 0.3 to 0.5.

As for the methods utilized to validate the aforementioned results, it is observed that most of them focus on a particular spatial granularity (e.g. city, nation or worldwide level). However, there is a scarcity of proposals that evaluate the mobility behaviour at different granularities from the same TWT dataset. This will allow to actually asses the coherence among patterns at different scales, which cannot be done when different TWT datasets are considered. In that sense, our work considers different spatial hierarchies to be compared as stated in Section 3.4.1.

Moreover, our approach also considers a time-based analysis of the human mobility patterns. Hence, the number of trips extracted from the platform is regarded as a time series and, based on such a representation, we utilize some similarity metrics as explained in Section 3.4.2. Although some studies like [17] consider the time-based pattern from TWT, they just measure the patterns with a particular baseline number of trips. On the contrary, we make use of a more varied palette of metrics to actually assess the time-dependent similarity.

Finally, an important part of our study focuses on modelling the mobility from the two target datasets, a mobile-phone location feed and the TWT one, by means of OD (origin-destinaton) matrices. This format has been widely used in the literature to represent human flows as it easily highlights the connections among entities within a geographical region [35]. Some works have already employed OD matrices as part of their methodology [20] but they comprise a very low number of entities. Thus, they do not fully take into consideration the *reliability* of complex OD matrices connecting a large number of entities extracted from an OSN and how such matrices might evolve through time.

3 Comparison of the mobile-phone and the OSN mobility studies

3.1 Mobile-phone location dataset

The mobile-phone location dataset has been downloaded from the INE web portal related to mobility patterns during the COVID-19 pandemic in Spain.[2] These data have been collected from more than the 80% of the mobile phones in Spain. The dataset covers the area for the entire country divided into 3,216 *mobility areas* (MA). These are population groups of between 5,000 and 50,000 inhabitants, representing a much more homogeneous area than the official municipalities to enable a more fine-grained analysis of the population's mobility. Thus, an MA in a depopulated zone is the sum of several small or very small municipalities (up to 5,000 inhabitants) whereas in large cities the MAs are the districts or even disaggregations of such districts. The INE has scaled the mobile-phone data collected for each MAs to the total population to estimate the real population on the move.

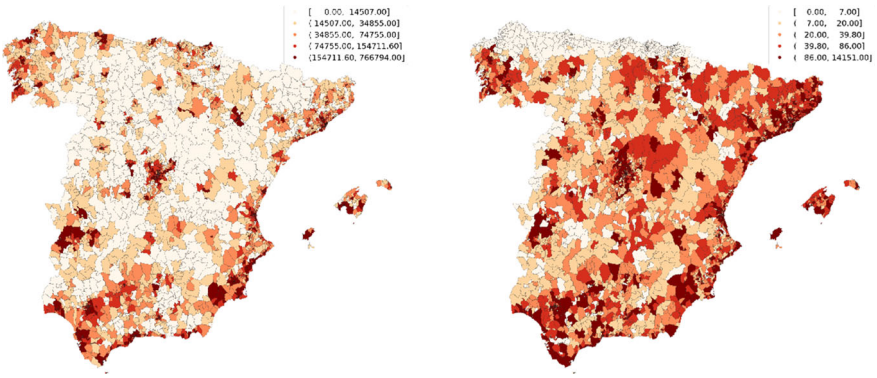
After processing the dataset, we obtained 1,082,387 entries with the following items for each entry: Origin mobility area (OMA), destination mobility area (DMA), number of outgoing trips from OMA to DMA and date. These entries comprise from April 1, 2020 to June 20, 2020 (a total of 81 days). Figure 1a shows the sum of all of outgoing trips in each MA during all the period of the study. Several areas of concentration of mobility can be observed in the center (Madrid), northeast (Barcelona), Mediterranean coast and some areas in the south (Andalusian region).

3.2 Twitter mobility dataset

The TWT dataset has been collected by means of the Streaming API of this social network. We collected geo-tagged tweets from whose point or polygon-based geometries spatially fitted into the boundaries of any of the MAs defined in the previous section from April 1, 2020 to June 20, 2020.

This initial set of documents was initially filtered by removing tweets accomplishing any of the two following criteria: 1) tweets from users who only posted in a single location during the whole time period or 2) posts from users who posted more than 50 tweets per day on average. The rationale behind these filters is to only retain tweets from users who moved around different locations and to discard potential bots. As a result, the final TWT dataset for the study comprised 8,210,773 cleaned tweets from 190,100 unique users. From those tweets 480,750 were geo-tagged with point coordinates and 7,730,023 included the coordinates of a polygon as location metadata.

[2] https://www.ine.es/covid/covid_movilidad.htm



(a) Total number of outgoing trips per MA in the mobile-phone dataset. (b) Total number of outgoing TWT trips per MA.

Fig. 1 Geographical distribution of trips in the mobile-phone and TWT datasets

The resulting set of tweets was used to extract the TWT-based OD matrix. To do so, we used a well-known trip-extraction procedure from tweets [10, 29]. In particular, a trip is regarded as sequence of two consecutive tweets $\langle tw^u_{ma_o} \rightarrow tw^u_{ma_d} \rangle$ from a user u posted at different MAs $\langle ma_o, ma_d \rangle$ ($ma_o \neq ma_d$) whose time difference is less than Δ_t, $tw^u_{ma_d}.timestamp - tw^u_{ma_o}.timestamp \leq \Delta_t$ ($\Delta_t > 0$). Hence, a TWT trip is regarded as a spatial displacement of u between the location indicated in the $tw^u_{ma_o}$ to the location reported by $tw^u_{ma_d}$.

In order to set the value of Δ_t, we studied the number of generated trips according to different values of the parameter as shown in Fig. 2. As a result of this study, the value of Δ_t was set to 24 hours. This is the cut-off value where the number of trips stabilizes. Moreover, the trips whose associate speed was over 500 km/h were also removed as they might represent unrealistic trips.

Eventually, 348,289 final trips were identified whose spatial distribution is shown in Fig. 1b. As it can be seen, the dark red MAs are more geographically spread than in the INE dataset (see Fig 1a). This reflects that the distribution of trips among the MAs is more balanced in the TWT dataset.

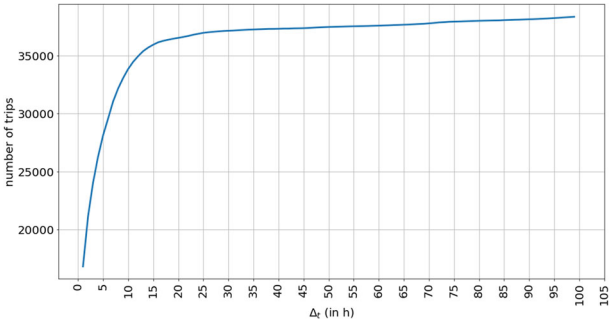


Fig. 2 Number of generated trips according to the Δ_t parameter

3.3 Comparison goals

In the light of the lack of proper comparisons among OSN and mobile-phone feeds representing human mobility and given the two datasets explained above in the context of the COVID-19 pandemic, the present work focuses on answering two different research questions (RQ):

RQ1 *Is it possible to infer human mobility patterns from TWT similar to the ones extracted from a mobile phone-based feed at different nation-wide spatial scales?*

To answer this question, several aspects like the level of connectivity among geographical areas (e.g. MAs) or the number of daily trips have been taken into account. Moreover, the existing human-mobility literature indicates that individuals actually move at different spatial scales, each one representing displacements at different granularities [2]. Therefore, in this study has been evaluated whether these scales might affect or not the capability of TWT to identify human mobility patterns in a reliable manner. To do so, three different spatial scales were defined as it is described in the next section of the manuscript.

RQ2 *Are the two mobility feeds (i.e. mobile phone and TWT datasources) able to capture in a similar way the change in the human flows due to the evolution of COVID-19 pandemic in a geographical area?*

For the second RQ we profit from the fact that the time period covered by this work fully included the lockdown restrictions in Spain to analyze the possible correlations among mobility and COVID-19 cases in both mobility feeds.

At this point, it is worth recalling that this study aims to assess whether the latent human-mobility behavior captured by the TWT dataset is similar to the mobility patterns reflected in the mobile-phone dataset. Consequently, we evaluate whether TWT could be regarded as a cost-effective datasource with respect to much more limited mobility studies in terms of availability. It should also be noted that the sampling size of the mobile-phone users in the INE study does not allow considering this survey as the actual ground-truth for human mobility patterns in Spain.

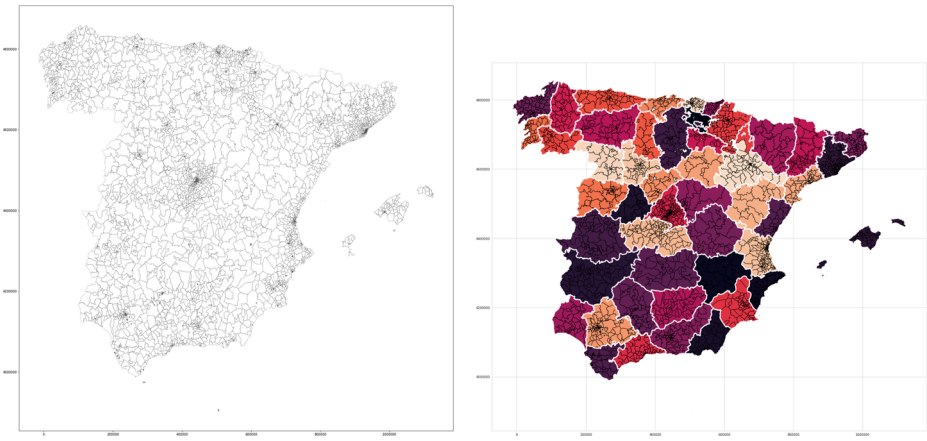
3.4 Comparison methodology

This section explains the methodology used to answer the aforementioned research questions.

3.4.1 Target spatial scales

Regarding RQ1, three different spatial levels of the Spanish geography were defined following a hierarchical approach, each level based on the previous one. In this manner we propose to measure the impact of the spatial *resolution* on the similarity of both datasets. The three spatial scales are as follows:

Level 1: *Mobility Areas (MAs).* The first granularity level just comprises the 3216 MAs defined in the INE study whose geographical distributions are shown in Fig. 3a. Furthermore, Fig. 3b shows the population distribution of the MAs. These areas cover a population of roughly 7,000 people on average which, in the case of Spain, is a quite small population segment. Hence, the mobility defined at this



(a) Spanish MAs.

(b) Spanish provinces. Each province is limited by a white line boundary.

Fig. 3 Geographical representation of the Spanish Mobility Areas (MAs) and provinces (PROV)

first level provides a quite fine-granularity overview of the human mobility in Spain.

Level 2: *Administrative provinces (PROVs).* In administrative terms, Spain is divided into 50 different provinces (Fig. 3b) each one with an average population of about 904,000 people (±120,000). Since each MA geographically fits into a single province, then a province can be regarded as an aggregation of MAs providing a corse-grained representation of the mobility flows of both datasets.

Level 3: *Clusters of provinces (CPROVs).* The administrative provinces were aggregated into clusters by means of the DBSCAN algorithm [11]. We opted for this algorithm because of its non-parametric characteristics for clustering applications based on point density. Unlike other clustering algorithms, DBSCAN does not require to indicate a number of clusters but a maximum distance (ε) between two samples to be considered neighbors, in addition to the number of points (*minPoints*) in a cluster to be considered a group of its own. This algorithm was feed with a new dataset were each PROV was described by 29 variables in four socio-economic aspects in the year 2019 (Table 2). Furthermore, the geographical center and the page-rank [26] and betweeness [37] centralities of each PROV considering the underlying connectivity graph from TWT and INE were also included.

From this process, a cluster of 5 INE-based and 6 TWT-based CPROVs were obtained. On average, each TWT cluster had population of 917,000 people (± 437,000) and each INE cluster covered a population of 950,000 people(± 609,000). To generate each cluster set, only the centrality values associated with each fed were used. The visual results of this process can be seen in Fig. 4a and b where the circles represent each PROV, the circle size is the betweenness centrality and a set of PROVs with the same colour stands for a CPROV.

This third level of abstraction helped us obtain a new point of view focused on the socio-economic reality of the individuals of each region. It shows that this reality is conditioned by the region where they live in and it does not necessarily have to be shared with other neighbouring provinces but may be similar to other non-neighbouring regions or belonging

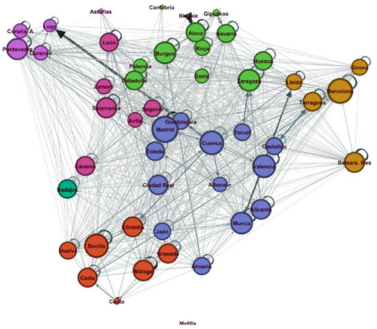
Table 2 The socio-economic variables used in the clustering algorithm for each Spanish province in 2019. CV stands for *Coefficient of Variation*

Category	Variable	Description	Mean	CV
Employment	Agriculture	% of people employed in agriculture	6.34	0.79
	Variability	Employment variability in agriculture	0.58	0.78
	Industrial	% of people employed in industry	13.82	0.43
	Variability	Employment variability in industry	0.60	0.46
	Building	% of people employed in building	6.19	0.23
	Variability	Employment variability in building	0.42	0.53
	Services	% of people employed in services	66.59	0.09
	Variability	Employment variability in services	1.05	0.56
	Unemployment	% of people unemployed	7.06	0.48
	Variability	Unemployment variability	0.70	0.66
Vehicles	Trucks	# of trucks sold in the year	49,009	1.02
	Vans	# of vans sold in the year	47,451	1.34
	Buses	# of buses sold in the year	1,259	1.40
	Cars	# of cars sold in the year	472,271	1.32
	Motorcycles	# of motorcycles sold in the year	69,369	1.52
	Tractors	# of tractors sold in the year	4,474	1.11
	Trailers	# of trailers sold in the year	9,381	0.99
	Others	# of other vehicles sold in the year	8,990	0.88
Population	Population	# of inhabitants	904,961	1.32
	%Population	% of inhabitants	0.02	1.31
	Density/Km^2	Population density	348.93	3.27
	Area	Area of a province	9,726.52	0.52
	%Area	% of the area over the country	0.02	0.52
	GDP	Gross domestic product in Euro	22,325,022.50	1.64
	GDP per capita	Gross domestic product per capita in Euro	23,226.00	0.20
Companies	Transport	# of transport companies	23,378	1.28
	Construction	# of construction businesses	8,122	1.34
	Industry	# of Industrial companies	3,828	1.23
	Services	# of Services businesses	29,348	1.77
Relationships	Journeys	Mean of journeys between provinces per day	3019,95	2.65

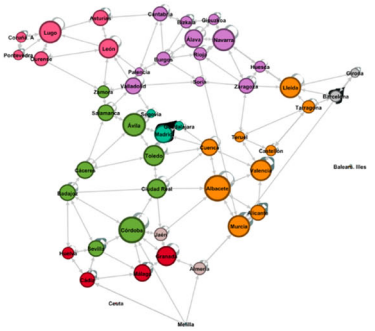
to different Spanish Autonomous Communities. Indeed, as observed in Fig. 4, there are some major differences in the obtained CPROVs. Whilst TWT identified a large CPROV covering the Southeast of the country (the blue groups of provinces in Fig. 4a), the INE dataset also identified a large CPROV but in this case at the West side of the nation (the green CPROV in Fig. 4b). In terms of similar results, note that the North coast of Spain was split into two different clusters in both datasets.

3.4.2 Human mobility views

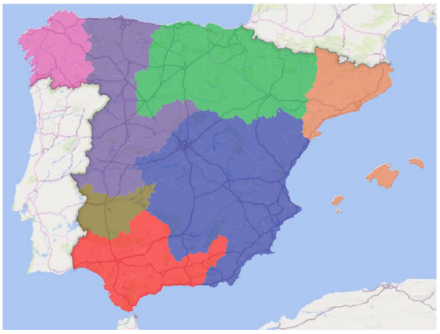
The comparison between the two feeds has been performed based on three different representations of the latent human mobility contained in them, as explained next:



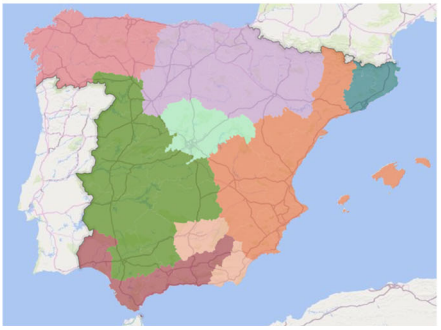
(a) TWT CPROVs.



(b) INE CPROVs.



(c) Final map of TWT CPROVs.



(d) Final map of INE CPROVs.

Fig. 4 Spatial representation of the CPROVs generated with DBSCAN. In figures a and b each single Spanish province is represented as a circle. Each CPROV is a group of circles with the same colour

Time series A set of time series, one for each mobility dataset and spatial hierarchy level, was defined as $\langle \mathcal{TS}_f^s \rangle \ \forall f \in \langle TWT, INE \rangle, \forall s \in \langle MA, PROV, CPROV \rangle$. In this case, each $\mathcal{TS}_f^s = \langle nt_f^{s,1}, nt_f^{s,2}, .., nt_f^{s,81} \rangle$ where $nt_f^{s,d}$ indicated the total number of trips at the d-th date according to the f feed in the s spatial scale.

OD matrix. A set of global OD matrices $\mathcal{OD}_f^{s,g} \ \forall f \in \langle TWT, INE \rangle, \ \forall s \in \langle MA, PROV, CPROV \rangle$ was composed based on all the trips during the whole period of study where each cell $o_{f,s}^{jk} \in \mathcal{OD}_f^{s,g}$ indicated the total number of estimated trips from area i to area j. It is worth mentioning that the aforementioned OD matrices were normalized due to the difference in the number of trips between the datasets (348,289 TWT trips vs. 1,082,387 mobile-phone trips). Consequently, these normalized matrices comprise the rate of trips from one area to the others. In that sense, these trip rates can be also useful in many application scenarios. For instance, they could reveal the importance of a particular city in terms of visitors/tourists in another city with respect to other urban areas.

Mobility graphs. Finally, we used a graph-based representation of the trips. In particular, it was defined a graph $\mathcal{G}_f^s = \langle \mathcal{N}, \mathcal{E} \rangle \ \forall f \in \langle TWT, INE \rangle, \ \forall s \in \langle MA, PROV, CPROV \rangle$. While the set of nodes in $\mathcal{N}$ are the particular MAs, PROVs or CPROVs

under consideration, the edges in $\mathcal{E}$ reports the sheer number of trips between each pair of nodes.

4 Comparison results

This section describes the main results obtained following the methodology described in previous section, splitting the results according to the three mobility views described in Section 3.4.2. Then, the results are discussed to extract the main lesson learnt in this study.

4.1 Comparison based on time series

The comparison based on the time-series views of the datasets allowed to test whether both feeds similarly captured the evolution of the sheer number of human trips in Spain during the COVID-19 pandemic. To do so, three different criteria have been used as described next.

4.1.1 Pearson Correlation

In order to compare both time series views, the first step was to visually analyze their associated time series $\mathcal{TS}_{TWT}^{MA}$ and $\mathcal{TS}_{INE}^{MA}$ as depicted in Fig. 5. Both datasets exhibited a quite similar behaviour through the whole period under study, with a continuous increment in the number of daily trips.

This initial similarity was further analyzed by decomposing each time series in their seasonal, trend and residual components in the three spatial scales as shown in Fig. 6. In this case, the trend component of all the time series confirmed the steady increment in the number of trips during the whole period under study.

Furthermore, we calculated the Pearson Correlation (PC) between the two time series for the three spatial levels. As observed in Table 3, there is a very high linear correlation among the time series generated from the trips at MA and PROV levels (PC values of 0.85 and 0.71, respectively). However, the correlation at the CPROV level is much lower (0.32).

In order to find an explanation to this PC drop at the CPROV level, its time series were decomposed as shown in Fig. 6. It is observed that the trend component of $\mathcal{TS}_{TWT}^{CPROV}$ (Fig. 6e) does not follow a increasing trend as clear as the one in $\mathcal{TS}_{INE}^{CPROV}$ (Fig. 6f). This

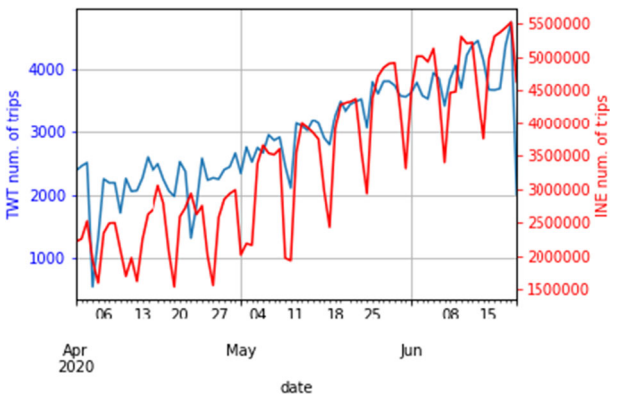
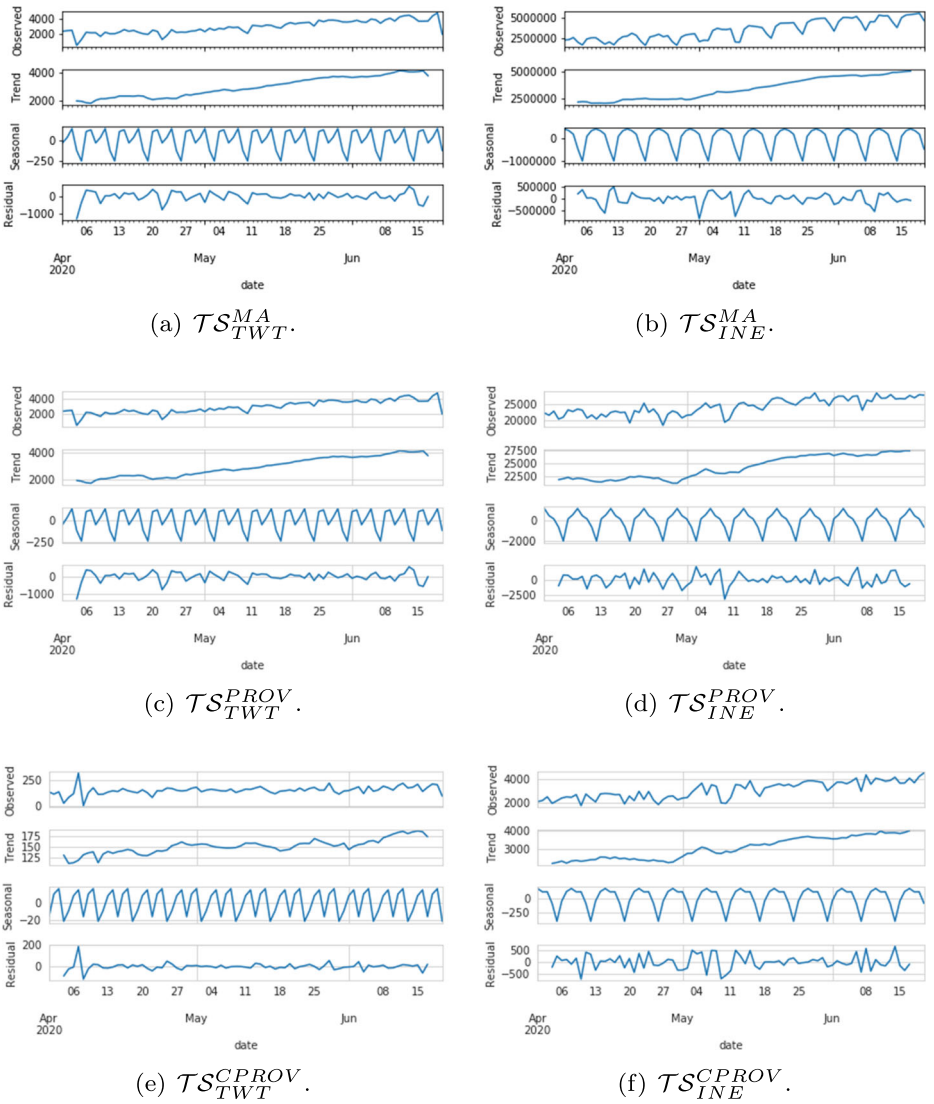


Fig. 5 Total number of MA trips per date under study



(a) $\mathcal{TS}_{TWT}^{MA}$. (b) $\mathcal{TS}_{INE}^{MA}$.

(c) $\mathcal{TS}_{TWT}^{PROV}$. (d) $\mathcal{TS}_{INE}^{PROV}$.

(e) $\mathcal{TS}_{TWT}^{CPROV}$. (f) $\mathcal{TS}_{INE}^{CPROV}$.

Fig. 6 Decomposition of the time series at the three target spatial scales

Table 3 PC among each pair of time series ($\mathcal{TS}_{INE}, \mathcal{TS}_{TWT}$) at the three spatial levels

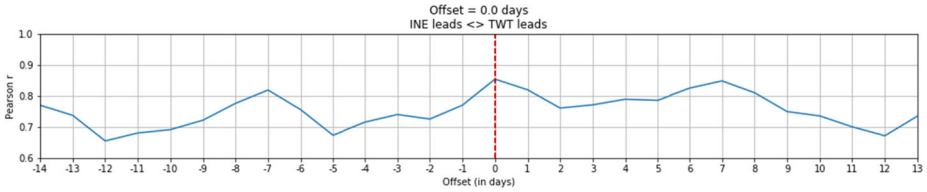
Spatial scale	PC
MA	0.8543 (P< 10^{-23})
PROV	0.7147 (P< 10^{-17})
CPROV	0.3260 (P< 0.0031)

suggests that the TWT study reflected a quite stationary pattern in terms of displacements among regions with different socio-economic profiles. However, the INE dataset did not captured this *silo effect* but a steady increment of trips similar to the other two spatial levels. Hence, the INE dataset seems to provide a more consistent behaviour across spatial levels.

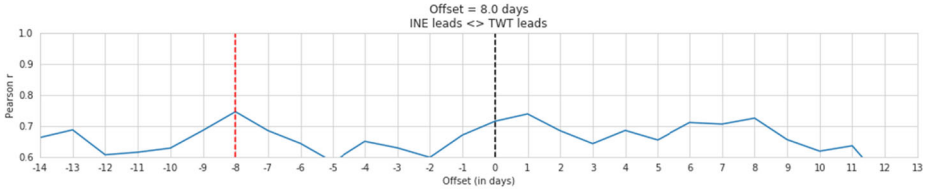
4.1.2 Time-lagged Cross Correlation

Next, the study focused on the level of *synchronization* of the time series. The goal in this case was to quantify if there was any type of leader-follower relationship between the TWT and INE time series. To do so, the time lagged cross correlation (TLCC) metric was applied [7]. In brief, this metric assesses the PC between two time series shifted relatively in time. The results of this analysis are depicted in Fig. 7.

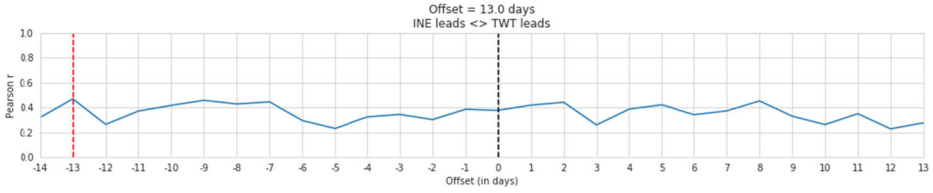
At MA level, the highest TLCC is achieved when the time offset is set to 0, that is, when none of the two time series is shifted (Fig. 7a). This indicates that both series are synchronized in time and thus their associated number of daily trips fluctuate at the same time. However, the TLCC peak is achieved with much larger offsets for the other two levels. Thus, the highest time-lagged PC is achieved with 7-day and 13-day shifts at PROV and CPROV levels, respectively (Fig. 7b and c). This indicates that TWT is a *slower* source of mobility than the INE dataset in capturing changes in the movement behavior of a population when



(a) MA level.



(b) PROV level.



(c) CPROV level.

Fig. 7 Time-lagged correlation between time series at the target three spatial scales

the spatial granularity is greater. In applicability terms, these delays seem too large to use TWT as a real-time feed to estimate the evolution of the mobility at spatial scales where the target spatial areas cover several thousands of people as PROVs and CPROVs do.

A possible reason for these gaps is that TWT users were more *reluctant* to initiate long-distance and non-regular trips (as the ones that mainly occurred at PROV and CPROV levels) during the COVID-19 pandemic than the Spanish population segment captured by the INE study. In that sense, we could assume that the INE study captures the movement of people involved in certain transport services that needed to operate again during the pandemic period as soon as possible giving rise to a much more accelerated increment of long-distance trips. The fact that both datasets are synchronized at the MA level also reflects that such TWT users changed their behaviour in terms of the number of short and regular trips at a similar pace than the INE target population thorough the period of study.

4.1.3 Granger Causality Test

In order to deep into the time-based similarity between TWT and the INE study, we also made use of the Granger Causality Test (GCT) [32]. Basically, the GCT allows determining if one time series (x) is helpful for predicting another (y). To do so, it tests the null hypothesis that the coefficients of past values of x in the regression equation to predict y are zero. Hence, GCT compares an *unrestricted model*, in which the time series y is explained by the lags of y and the lags of an additional series of observations from x (both lags up to a same fixed order), with a *restricted model*, in which y is only explained by the lags of y. Table 4 shows the p-value of this test for different number of lags (i.e., past days) for the time series $\mathcal{TS}^{s,g}_{INE}$ and $\mathcal{TS}^{s,g}_{TWT}$ when each one adopts the x or y role.

From this table, it can be observed that $\mathcal{TS}^{s,g}_{TWT}$ is more useful to explain or predict the behaviour of $\mathcal{TS}^{s,g}_{INE}$ than the other way round. As a matter of fact, the values in bold in Table 4 show that the null hypothesis can be rejected with a significance level of 0.05 when $\mathcal{TS}^{s,g}_{TWT}$ adopts the role of the exogenous variable x at MA and PROV scales for time lags between 1 and 7 days. However, TWT does provide this explanatory capability at CPROV level. A reason for this pattern is found in the TLCC results stated in Section 4.1.2. Following this analysis, we discovered that the time gaps of TWT with respect the INE study were 0, 7 and 13 days for each of the spatial scales MA, PROV and CPROV, respectively. Consequently, the gap between both feeds is small enough in the first two levels given the

Table 4 P-values of the GCT for different number of lags related to time series $\mathcal{TS}^{s,g}_{TWT}$ and $\mathcal{TS}^{s,g}_{TWT}$. The values in bold are below the significance level 0.05, and therefore the null hypothesis can be rejected

Level	x	y	Number of lags (in days)						
			1	2	3	4	5	6	7
MA	INE	TWT	**0.0345**	0.1473	0.2219	0.299	0.0512	**0.0015**	**0.0001**
	TWT	INE	**0.0171**	**0.0035**	**0.0001**	**0.0000**	**0.0000**	**0.0014**	**0.0000**
PROV	INE	TWT	**0.0001**	**0.0063**	**0.0360**	0.1077	0.3406	0.1499	0.4816
	TWT	INE	**0.0001**	**0.0010**	**0.0008**	**0.0001**	**0.0035**	**0.0011**	**0.0255**
CPROV	INE	TWT	**0.0001**	**0.0014**	**0.004**	**0.0335**	0.1350	0.2253	0.2627
	TWT	INE	0.1535	0.1178	0.1689	0.1059	0.0768	0.0740	0.0516

lags considered in the GCT evaluation. However, it is too large at CPROV level for TWT to have some influence on the INE time series in a predictive model.

Table 4 also shows that $\mathcal{TS}_{INE}^{s,g}$ has some predictive influence on $\mathcal{TS}_{TWT}^{s,g}$ in the three scales for lags equal or below 4 days. This is consistent with the *reluctant* trips behaviour detected in TWT users (Section 4.1.2). Given such a leader-follower relationship, it seems reasonable that the INE time series can anticipate the movement behaviour of TWT users some days ahead.

4.1.4 Comparison with COVID-19 cases

At this point, the next step was to *contextualize* the uncovered steady increment of daily number of trips in both datasets. To do so, it was used the daily number of official COVID-19 cases in Spain reported by the Spanish Ministry of Science[3]. Thus, Fig. 8 depicts the time series of each spatial level along with the number of COVID-19 cases whereas Table 5 comprises the PC of each time series with such a number of cases.

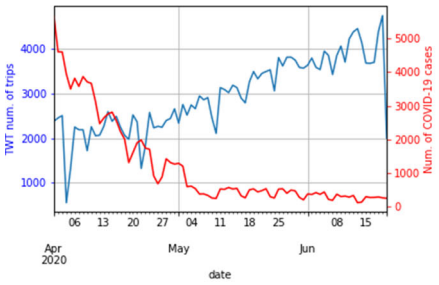
As the negative PC scores in Table 5 suggest, the mobility increment captured by the time series evolved in parallel with the progressive decrease of COVID-19 cases. As a result, the intuitive idea that the mobility behaviour of Spain, in terms of number of daily trips, was directly affected by the evolution of the COVID-19 pandemic was properly captured not only by the INE dataset but also by the TWT one in a similar manner. Indeed, it is observed PC scores around -0.6 in all the time series except for $\mathcal{TS}_{TWT}^{CPROV}$, which achieved a much lower PC (-0.35). This is consistent with the stationary behaviour of this time series discussed in Section 4.1.1.

4.1.5 Answers to the RQs based on the time series analysis

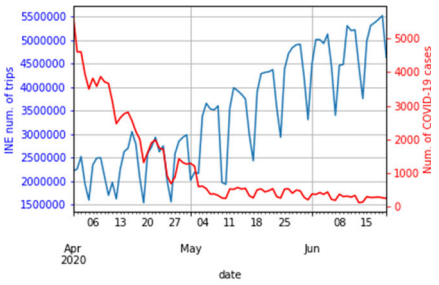
Based on the analysis of the time series, we now can provide some answers to the two RQs. Regarding RQ1, we have observed that the time series extracted from TWT followed a steady increment in the number of trips that was similar to the one observed in the INE study at low spatial resolutions. This similarity is not only related to the amount of trips (as discovered in Section 4.1.1) but also to the *pace* of this rate (Section 4.1.2). However, this similarity meaningfully decreases when we move to upper spatial scales specially at CPROV level. The GCT has also confirmed this disparity between the MA level and the upper ones (Section 4.1.3).

Concerning the RQ2, the TWT timeseries exhibited a negative correlation with respect the COVID-19 cases evolution, and this correlation was similar to the one obtained by the INE study at MA and PROV levels (Section 4.1.4). This indicates that TWT and INE captured in a similar manner the changes in the mobility patterns of Spain, in terms of sheer number of trips, in the first two spatial scales. These changes unsurprisingly indicated that the aforementioned increment in the number of trips was proportional to the decreasing evolution of the COVID-19 pandemic. However, the patterns of TWT and INE meaningfully diverged at CPROV level. In particular, the stationary pattern observed the TWT time series was not consistent with the pattern of increment of trips extracted from the INE study at this level.

[3] https://cnecovid.isciii.es/covid19/

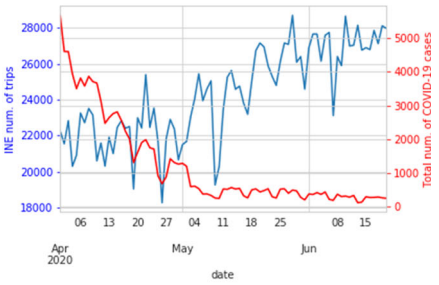


(a) $\mathcal{TS}_{TWT}^{MA}$ and daily COVID-19 cases.



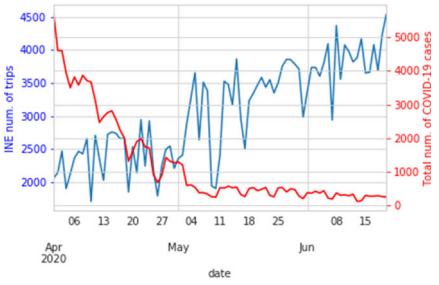
(b) $\mathcal{TS}_{INE}^{MA}$ and daily COVID-19 cases.

(c) $\mathcal{TS}_{TWT}^{PROV}$ and daily COVID-19 cases.



(d) $\mathcal{TS}_{INE}^{PROV}$ and daily COVID-19 cases.

(e) $\mathcal{TS}_{TWT}^{CPROV}$ and daily COVID-19 cases.



(f) $\mathcal{TS}_{INE}^{CPROV}$ and daily COVID-19 cases.

Fig. 8 Evolution of the number of trips at the three spatial levels and the number of daily confirmed COVID-19 cases in Spain

Table 5 PC among time series and the number of COVID cases

Spatial level	Time series	PC with COVID-19 cases
MA	$\mathcal{TS}_{TWT}^{MA}$	-0.6584 (P< 10^{-10})
	$\mathcal{TS}_{INE}^{MA}$	-0.6265 (P< 10^{-9})
PROV	$\mathcal{TS}_{TWT}^{PROV}$	-0.6630 (P< 10^{-10})
	$\mathcal{TS}_{INE}^{PROV}$	-0.5541 (P< 10^{-7})
CPROV	$\mathcal{TS}_{TWT}^{CPROV}$	-0.3532 (P< 0.0013)
	$\mathcal{TS}_{INE}^{CPROV}$	-0.6231 (P< 10^{-11})

4.2 Comparison based on OD matrices

After studying the similarity of both datasets in terms of the magnitude of trips through time, we focused on analyzing the similarities of those trips in terms of their actual origins and destinations. To this end, we made use of the OD matrices described in Section 3.4.2.

4.2.1 Number of origins and destinations

First, we evaluated whether each pair of global matrices $\langle \mathcal{OD}^{s,g}_{INE}, \mathcal{OD}^{s,g}_{TWT} \rangle$ comprised a similar level of connectivity. To do so, we focused on the number of counterparts that a particular MA, PROV or CPROV is connected to. Thus, Fig. 9 depicts the distribution of origins and destinations of each spatial scale based on these global matrices. It is worth noting that only meaningful connections comprising at least 5% of the outgoing trips of an MA, PROV or CPROV were considered in this part the study (i.e., cells in both global matrices with values below 0.05 were discarded).

Concerning the MA level (Fig. 9a), both studies returned quite similar results in terms of MA connectivity. More in detail, MAs usually have 2 primary origins in the case of TWT and 3 in the case of INE (see the blue box-plots at Fig. 9a). Regarding the primary destinations, both studies concluded that the majority of trips from an MA finished, on average, at five different MAs (see the orage box-plots at Fig. 9a).

Concerning the PROV level (Fig. 9b), TWT and INE had ranges of origins and destinations with a large overlap degree. Hence, whilst a PROV absorbs, on average, human trips from 2 different provinces according to TWT, in the case of INE this number is 3 (see blue box-plots in Fig. 9b). As for the destinations (orange box-plots in Fig. 9b), the population of a PROV usually moves to 3 different provinces according to both mobility sources.

Regarding the connectivity at CPROV scale, we must take into consideration that the TWT and INE trips are distributed at different number of CPROVs (Section 3.4.1). However, it is observed that the TWT CPROVs had a much higher connectivity than their INE counterparts, as reflected in box-plots of Fig. 9c. In particular, the blue boxes representing the number of meaningful connections of the TWT CPROVs cover a much larger range of values that the INE (orange) ones. As a result, there are isolated CPROVs in the TWT clustering setting that do not receive trips from other CPROVs (i.e., number of origins 0) whereas there are others that receive trips from all the other 5 CPROVs (number of origins

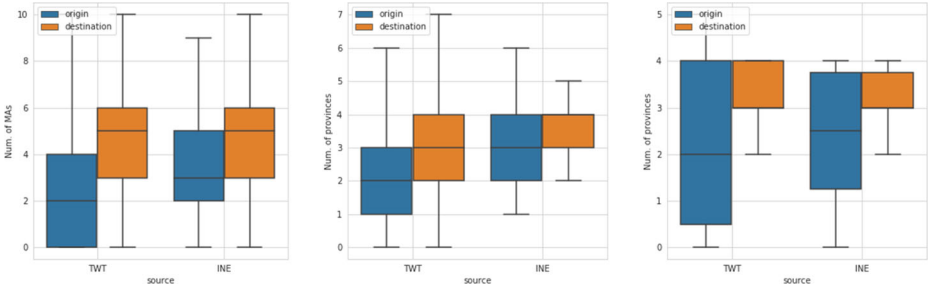


Fig. 9 Distribution of the number of origins and destinations per MA, PROV and CPROV

5). On the contrary, the number of connections in the INE setting is much more balanced with origins and destinations ranging from 2 to 4. This finding is consistent with the reluctant behavior of TWT users to move across CPROVs discovered in the time series analysis (Section 4.1.2). This behaviour not only caused a slower increment of the number of trips but also a much lower connectivity among CPROVs.

4.2.2 Ranking of origin and destinations based on distance

Another aspect that we studied from the OD matrices was the distance between areas based on their trip-rate rank, as shown in Fig. 10.In that sense, it is obvious that increasing the geographical granularity would also increase the spatial distances of the trips in absolute terms. Hence, the goal of this comparative is to evaluate whether this change occurs following the same pattern in the two mobility feeds.

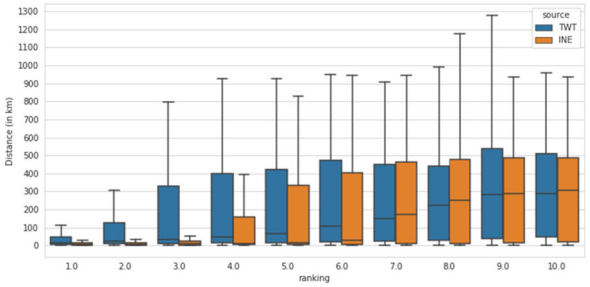
This plot reflects that the most relevant destination of an MA is, on average, at a distance of 5 km according to the INE study (see the leftmost orange box-plot of Fig. 10a) whereas in the TWT study this distance is 11 km (see the leftmost blue box-plot in the same figure).

At PROV level (Fig. 10b), we observed a similar pattern as for the MA level. Nonetheless, it should be pointed out that the distances for the most relevant destination of a province ranged from 105 km to more than 400 km for TWT whereas in the case of INE this range was much more narrow (see the two leftmost orange and blue box-plots of Fig. 10b).

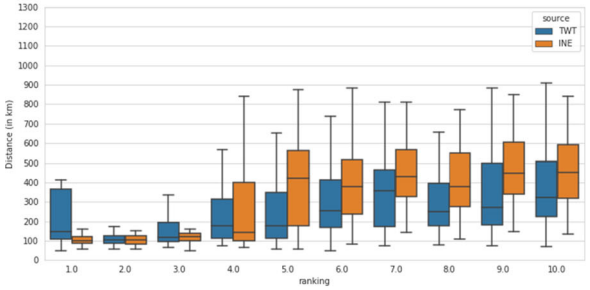
Given these results, it should be noted that a well-known human-mobility pattern study elsewhere [22] states that the relevant destinations of a person are usually generated by his/her frequent and daily trips (e.g. home-work itineraries), which are mainly short distance ones. According to Fig. 10a and b, both INE and TWT reflected this mobility behaviour as the average distances for the first five destinations in their rankings were below 50 km at MA level and 150 km at PROV level. Next, this distance increased as long as the relevance of the destination decreased in both feeds. As a matter of fact, the 10th most important destination of an MA in terms of volume of trips was located, on average, at 312 km based on the INE study and 283 km in case of TWT (see the two rightmost box-plots in Fig. 10a). A similar pattern is observed at PROV scale This was also consistent with the power law distribution of human trips comprehensively discussed in the literature [3].

However, it should be noted that Fig. 10a and b also show that the height of the blue box-plots corresponding to TWT were much larger than the orange ones (INE), specially in the first four positions in the ranking. This indicates that the ranges of distances of these positions were much larger in the case of TWT than in the INE study. Thus, some MAs and PROVs in TWT had very meaningful connections to quite distant counterparts. For example, certain MAs had their second most relevant destination more than 100km away according to TWT. Consequently, TWT seemed to be a more noisy feed in terms of MA and PROV connectivity than the INE study. This second study exhibited a quite homogeneous behavior in terms of range of distances.

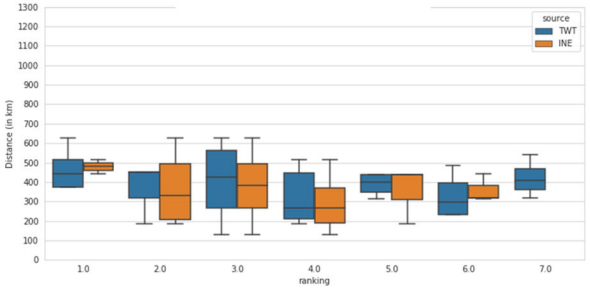
Lastly, an interesting finding at CPROV level (Fig. 10c) was that the increment of the distance to a destination based on its decreasing importance is not preserved as in the other two lower scales. From Fig. 10c it can be seen that the distances among CPROVs remained flat regardless of the destination ranking with values between 200 km and 550 km. In this case, we must clarify the that distances were defined on the basis of the spatial centroids of the CPROVs. Moreover, the ranges of the two feeds had a large level of overlap in all the rank positions. Indeed, the 4th most important CPROV was located, on average, at roughly



(a) MA scale



(b) PROV scale



(c) CPROV scale. Remark that the 7th position only includes the TWT boxplot as the INE setting only gave rise to 6 CPROVs

Fig. 10 Distribution of distance to the destination area based on its trip-rate rank

280 km in both datasets. This shows that increasing the spatial scale allowed to homogenise the diversity of trips between TWT and the mobile-phone feed to a large extent.

4.2.3 RMSE between matrices

Another metric that was used to measure the similarity between the global OD matrices, $\mathcal{OD}_f^{s,g}$, was the Root Mean Squared Error (RMSE). In this case, it is worth mentioning that the i-th row in each OD matrix comprises the normalized rate of trips from the i-th area (at MA, PROV or CPROV level) to the rest of areas. Hence, RMSE values were computed

for each pair of similar columns in the two related matrices. Next, the final RMSE was calculated as the mean of these pairwise values. For instance, the RMSE between the two global OD matrices at MA resolution was calculated using the formula

$$RMSE(\mathcal{O}_{TWT}^{MA,g}, \mathcal{O}_{INE}^{MA,g}) = \frac{1}{3216} \times \sum_{i=i}^{3216} RMSE(\mathcal{O}_{TWT}^{MA,g,i,*}, \mathcal{O}_{INE}^{MA,g,i,*})$$

where $\mathcal{O}_{TWT}^{MA,g,i,*}$ represents the i-th row of $\mathcal{O}_{TWT}^{MA,g}$ and $\mathcal{O}_{INE}^{MA,g,i,*}$ represents the i-th row of $\mathcal{O}_{INE}^{MA,g}$ and 3216 is the total number of MAs. Note that this metric has been computed over the normalized matrices, and therefore values range from 0 to 1.

However, Fig. 9 shows that the connectivity among MAs and PROVs is rather poor and, consequently, the OD matrices are very sparse. As a result, each row is mostly filled with 0s. Consequently, we computed the RMSE by considering two different combinations of the OD rows as follows:

- Firstly, we computed the *Raw RMSE* by considering the original rows of the OD matrices.
- Secondly, we compute the *Restricted RMSE* by only considering the elements of the rows extracted from the INE and TWT OD matrices with values above 0. Therefore, the positions with value 0 in both rows are removed before computing the RMSE. In this manner, we reduce the impact of these no-connections in the error computation.

Table 6 shows the values for each RMSE. The column *Raw RMSE* shows that the raw form of the matrix obtained quite low values. For example, the error at MA level was limited to 0.0002. Nevertheless, we should take into account that these low values are due to the sparse nature of the OD matrices.

Regarding the *Restricted RMSE*, the values are still low (below 0.07 in all the cases) and no large differences are observed across spatial scales. As a matter of fact, the error was 0.0656 at MA scale and 0.0640 at CPROV level. What this reveals is that the rate of trips among regions was very similar in both mobility feeds. In order to explain this 6% deviation between TWT and INE, we made use of the radius of gyration metric as stated in the next section.

4.2.4 Radius of Gyration

The radius of gyration is a well-known metric within the human-mobility analysis discipline [14]. This metric allows measuring the distance travelled by an individual or a population during a particular time period. Based on this metric, we have defined the radius of gyration of an area a according to a feed f, $r_g^f(a)$, by means of the following formula

$$r_g^f(a) = \sqrt{\frac{1}{n_a} \sum_{i=1}^{n_a} dist(l_a, l_i)^2}$$

Table 6 Values of the different types of computed RMSE at the three spatial scales. The standard deviation is shown in brackets. The lower value per RMSE type is shown in bold

Level	Raw RMSE	Restricted RMSE
MA	**0.0002 (±0.0001)**	0.0656 (±0.0295)
PROV	0.0076 (±0.0078)	**0.0613 (±0.0436)**
CPROV	0.0124 (±0.0123)	0.0640 (±0.0282)

where $dist(l_i, l_a)$ is the distance in kilometers between the area a and the i-th area, and n_a is the total number of areas that a is meaningfully connected to showing a trip rate above 5% according to the mobility feed f.

Figure 11 shows the value of this metric in each of the three spatial scales. It is observed a similar Gaussian distribution of values in both feeds at PROV and CPROV scales (Figs. 11b and c).

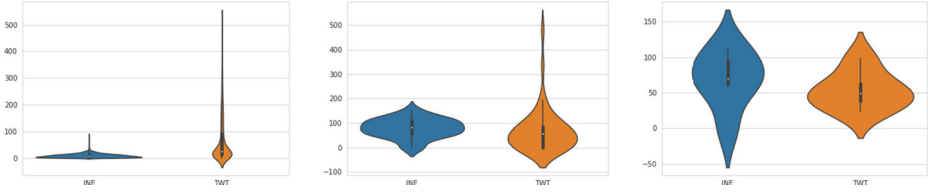
At MA scale, TWT had a very different Gaussian distribution with a very long tail up to 500 km (Fig. 11a). On the contrary, the INE dataset had a very narrow distribution of this level showing that most of the trips at this level covered very short distances. As previously discussed in Section 4.2.2, this confirms that the TWT dataset gave raise to some abnormal connections between distant regions that are not present in the INE study. This might be due to the fact that the mobility patterns of certain areas are mainly defined by the behavior of a limited number of TWT users that generate such random and abnormal connections. This would also explain the fact that the connection profiles of the INE study and TWT tends to have more similar distributions at PROV and CPROV scales. In that sense, the inherent aggregation of trips that occurs at such levels reduces the impact of the abnormal behaviour of individual users.

4.2.5 Answers to the RQs based on the OD matrices analysis

The study based on OD matrices shed some light on RQ1. Indeed, the latent mobility found in the OD matrices from TWT and INE reflected similar connectivity models among regions. This similarity is related to the distribution of trips with respect to distance (Section 4.2.2) and the rate of trips from each area to the others (Section 4.2.3). However, TWT generated some connections between distant areas at low-resolution spatial scales that are not observed in the INE study (Section 4.2.4). In that sense, the highest similarity in connectivity terms between both feeds occurs at PROV and CPROV levels where those *noisy* connections do not occur.

4.3 Comparison based on Graphs

The third comparison focuses on the graph representation of the mobility flows of both feeds, $\langle \mathcal{G}^s_{INE}, \mathcal{G}^s_{TWT} \rangle$. By considering the whole graph, this last analysis allowed us to study the connections among regions more comprehensively than the OD matrices. To do so, we rely on two well-known centrality measurements from the graph analysis domain.



(a) MA scale (b) PROV scale (c) CPROV scale

Fig. 11 Violin plots showing the rotated kernel density of the radius of gyration of both feeds in the three spatial scales

4.3.1 Comparison based on the PageRank centrality

The Pagerank method (PR) ranks the importance of a web page through the hyperlink structure of a web system [21]. This value of a web page (p_i ($i = 1, \cdots, n$)) can be calculated as follows:

$$PR(p_i) = \sum_{P_j \in D(p_i)} \frac{PR(p_j)}{N(p_j)}, \tag{1}$$

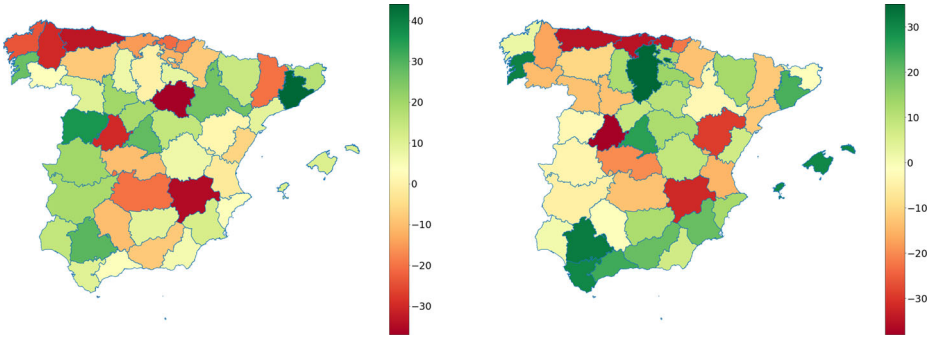
where $PR(p_j)$ are all pages pointing to p_i and $N(p_j)$ is the number of link pages starting from p_j. This index is widely used in networks to measure the importance of a node in a network.

By means of this measure, we estimated the importance of an MA in the network. Thus, we calculated the PR centrality in a directed network of 3215 nodes (MAs), where the number of trips between two nodes represents the weight of the link. We compared the PRs obtained by each MA based on the TWT and INE graph obtaining a Person Correlation (PC) of 0.95. This was consistent with the RMSE obtained from the OD matrices (Section 4.2.3).

Concerning PROVs, we observe a clear drop of the PC score at this level (0.17). To further analyze this result, we composed two rankings of provinces: one of them ranked the provinces decreasingly based on their PR scores from $\mathcal{G}_{INE}^{PROV}$ and the other one did the same based on $\mathcal{G}_{TWT}^{PROV}$. Next, we calculated the position difference of each province in the TWT rank with respect to the INE one. As a result, we composed the map shown in Fig. 12a.

It is observed that the yellowish color of most of the eastern provinces in Spain indicates a quite high similarity of their PR scores in the INE and TWT graphs. However, the green colors in the western regions clearly indicate that those provinces were much more relevant in the INE study than in the case of TWT. Finally, the northern provinces were clearly more relevant in $\mathcal{G}_{TWT}^{PROV}$ than in the INE graph.

This finding provides an overview of the connectivity differences among regions based on both feeds that is wider than the one extracted from the OD analysis. Although the plain origin-destination tuples at PROV level were quite similar in both feeds (Section 4.2.3), when considering the connections at a global scale we observe that such global connectivity



(a) PR centrality differences (b) BT centrality differences.

Fig. 12 Differences among Spanish provinces based on its PR and BT centrality measurements from the INE and TWT graphs. Red provinces are those whose position in the TWT ranking was higher than in the INE one. On the contrary, the green ones are provinces whose position in the INE ranking was higher than in the TWT one

is not homogeneous but varies across different parts of the considered area in terms of north-south or east-west dimensions. This reflects that the bias of TWT may not be only limited in terms of socio-economic aspects or population density of its users [18] but also in terms of the actual spatial coverage of the extracted mobility patterns.

Furthermore, we also computed the PC between COVID-19 incidence and these centrality metrics at PROV level. We obtained a rather high correlation value in the case of INE (0.91) and a slightly lower one for TWT (0.37). As a result, the INE dataset seems to better support the intuitive idea that the most relevant a region was in terms of human mobility, the more exposed to the COVID-19 incidence. In the case of TWT, this relation was weaker.

4.3.2 Comparison based on the Betweenness centrality

We define the classical betweenness centrality (BT) as an undirected graph $G = (V, E, W)$, where V is the set of nodes, E the set of edges and W the weight of each node.

Considering σ_{ij} as the number of shortest paths from a node i to node j and σ_{itj} as the number of shortest paths from node i to node j through a transition node t, then the fhe formal definition of betweenness centrality measure of a node t is [13],

$$BT_t = \sum_{i \neq t} \sum_{j \neq i,t} \frac{\sigma_{itj}}{\sigma_{ij}}. \tag{2}$$

This metric is relevant in our problem because it allows uncovering whether a particular area is likely to be the final destination or it is just an intermediate stop in longer displacements. In that sense, the PC at MA level between the BT centralities obtained from $\mathcal{G}^s_{INE}$ and $\mathcal{G}^s_{TWT}$ was very high, 0.99.

At PROV scale, this PC value remarkably dropped to 0.14. Therefore, we followed the same rank-comparison approach than the one used in the PR metric to discover its cause. The final map comprising the BT ranking differences is shown in Fig. 12b. It is observed that the yellowish provinces indicating a high similarity in the BT rank based on TWT and INE are mostly located in the western part of Spain. However, most of the green provinces (indicating that the position of the province in the INE rank was higher than in the TWT one) are mostly located in the south and eastern coast of Spain. This reflects that the INE study capture a high connectivity of these regions with the rest of provinces of Spain (they potentially acted as intermediate stops of displacements across other regions) that was not observed in the TWT mobility set.

On the contrary, most of the northern provinces are coloured in red indicating that, in this case, the OSN capture a much richer connectivity among these regions. Besides, it is worth mentioning the two provinces clearly marked in red in the eastern side of Spain in Fig. 12b. These two provinces actually act as *transportation hubs* connecting the Mediterranean coast of the country (where the tourist sector is very relevant) with the central provinces as they are crossed by multiple transportation infrastructures (e.g. highways and high-speed train railways). In that sense, TWT captured this role of both provinces in a more clear manner than the INE dataset by assigning them a much higher BT rank.

Finally, as in previous section, we calculated the PC between the COVID-19 incidence and these centrality rankings at PROV scale. The results were similar to the ones obtained based on PR in the case of INE, 0.83. Again, the high correlation in the INE study made sense because a PROV with high BT receives travellers from many parts of Spain and, thus, it increases the probability of a large propagation of the virus. In the case of TWT we still observed a rather weak PC, 0.50.

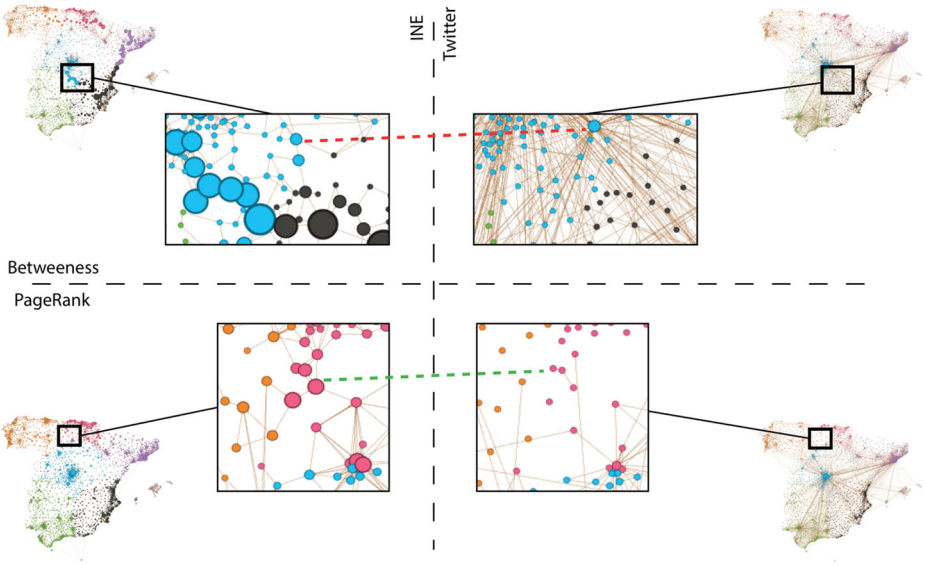


Fig. 13 Comparison of two centrality metrics (betweeness in top, PageRank in bottom) in TWT (left) and INE (right) for the MAs associated to the municipalities of Tarancon (top) and Reinosa (bottom). Each circle represents an MA and its size indicates the value of the particular centrality metric

4.3.3 Use case

For the sake of completeness, we introduce two use cases to explain some particular differences between the TWT and INE mobility flows in the graph dimension and how such mobility datasets can help to give insight into the virus spreading risk of a region during the COVID-19 pandemic (see Fig. 13).

To begin with, we detected a non-touristic small town that became a relevant source of virus spreading. This is the case of the MA associated to the municipality of Tarancón in the administrative province of Cuenca (see the two top figures in Fig. 13). Its population is only about 15,000 people, but this MA had the highest COVID-19 incidence in its province[4]. According to the TWT-based connections with other areas, it can be seen that this MA has a high value of BT centrality (see Fig. 13 top-right).

This is because this town is a well-known *refreshment stop* for travellers moving across 3 important and distant cities (Madrid, Barcelona and Valencia). In that sense, eating-out is a quite dangerous activity in terms of virus propagation. At the same time, it is a social activity that could facilitate the publication of geo-tagged *tweets* in that region. As a side effect, this allowed to uncover many TWT trips whose origin or destination was this MA. This made it very relevant in terms of connectivity. On the contrary, it was not possible to highlight this area based on INE data (see Fig. 13 top-left). This is because this dataset includes all trips, regardless of the causes of the stop (e.g. refueling, etc.).

Secondly, we studied how an isolated and small MA was seriously affected by COVID-19 cases due to its location. This is the case of an MA in the border of three administrative provinces in the north of Spain (see the two bottom plots in Fig. 13). This MA includes

[4]https://sanidad.castillalamancha.es/evolucion-de-coronavirus-covid-19-en-castilla-la-mancha

the town of Reinosa whose population is under 10,000 people, but it suffered from a higher COVID-19 incidence than close bigger cities.[5] We calculated the BT and PR centralities based on the MA links with other areas. By using the TWT dataset, it was not possible to capture any anomaly as the centrality values were similar to the ones of its neighbour MAs (see Fig. 13 bottom-right). However, according to the INE dataset, the Pagerank centrality highlighted this MA (see Fig. 13 bottom-left).

The rationale of this finding is the location of the MA. As previously pointed out, this MA is close to the border of 3 PROVs, thus it was influenced by the inter-province mobility. However, these three northern provinces were *over-represented* in the PR list based on TWT with respect to the INE study (Fig. 12a). Due to this global increment of the PR at province level, the target MA did not have an abnormally high PR as it had in the INE study.

4.3.4 Answers to the RQs based on the graph analysis

This graph-based analysis has provided a bigger picture of the similarities between the target mobility feeds to answer RQ1. It has been noted that the similarities in the trip connections among regions reported by TWT and INE varied meaningfully in different parts of Spain at PROV level.

Figure 12 reveals that the human flows captured by TWT and INE are more similar in the eastern side when we focus on the capability of the regions to *attract* trips from other regions at global scale (PR metric). However, both feeds provide a similar mobility model in the western side of the country when we focus on the regions' capability to act as *transportation hubs*. This reveals some important differences at nation-wide mobility scales that were not captured by the OD study.

As for RQ2, we have seen that the centrality metrics obtained from the INE graph were more correlated with the COVID-19 cases than the TWT study at PROV scale. One explanation to this difference is that some of the most-highly populated Spanish provinces, like Madrid or Barcelona, were under-represented in the metrics extracted from TWT causing this mismatch. However, these regions had a very high incidence of COVID-19 cases during the time period under consideration.

4.4 Lessons Learnt

After the analysis performed in the previous sections, it is possible to comprehensively answer the two RQs stated as initial hypothesis.

RQ1. Is it possible to infer human mobility patterns from TWT similar to the ones extracted from a mobile phone-based feed at different nation-wide spatial scales?

TWT has proved to be a reliable source to capture the mobility trends of Spain. Nevertheless, this reliability varies depending on the mobility feature, the spatial resolution or event the part of the country under consideration.

In the time dimension, TWT allows extracting a mobility behavior quite similar to the INE study in terms of daily number of trips at MA and PROV scales as showed by the high PC and TLCC measurements (Figs. 5 and 7). However, increasing the granularity of the geographical tessellation makes this similarity to meaningfully decrease. At CPROV scale,

[5] https://experience.arcgis.com/experience/9fc123d100e540dda44529d5aff5fd67

the INE feed also reported a continuous increment of number of inter-cluster trips, but it was not possible to infer such a behavior from TWT. On the contrary, the number of TWT trips remained more or less stationary (around 250 trips per day) at this upper level.

Focusing on the capability of TWT to mirror the trip connections among pairs of areas of the INE study, TWT actually captured the power law distribution of the distances based on destination importance of the INE dataset in the first two hierarchical levels, as Fig. 10 depicted. Furthermore, the RMSE values suggested that TWT actually captured the trip rates among target areas in a quite similar manner than the INE feed with values below 0.07 in all the spatial levels (Table 6).

Regarding the global trip connectivity provided by TWT, the graph-based comparison revealed that its similarity with the INE feed was different based on the geographical region under consideration. Depending on whether we consider the mobility relevance of a region to act as a final or as an intermediate destination of human trips, the similarity between the INE and TWT feeds changed in the eastern and western side of Spain.

To conclude, TWT seems to capture well the evolution of the short-distance displacements included at fine-grained spatial scales along with their origin-destination distributions. These short displacements are generated by the work and leisure trips that people do on regular basis. However, when the spatial scale is enlarged, and thus only the trips covering larger and more unusual distances are taken into account, TWT becomes a slightly more unstable and incoherent datasource in terms of volume of trips. However, the connectivity among regions (the origins and destination of these trips) reflected by TWT at these coarse-grained spatial scales is similar to the one reflected in INE.

All in all, given the high similarity of TWT and the INE study in the evolution of the number of daily trips, the direct origins and destinations of these displacements and the global relevance the destinations (PR score) in the eastern coast of Spain, it seems feasible to consider this OSN a mobility source as reliable as one based on mobile-phone records in such geographical area of the country.

There are possible explanations to all these findings. One is that TWT users may represent a population sample that, in mobility terms, is less prone to travel long distances that the one represented by the INE feed. Another possible explanation of this pattern is that the volunteer nature of TWT data made users more reluctant to post geo-tagged tweets far from their usual areas of movement during the pandemic scenario covered by the study. This eventually made the inference of long trips from TWT rather scarce. Also, this TWT population is not only biased in socio-economic terms but also in the spatial coverage of the mobility patterns extracted from the OSN.

RQ2. Are the two mobility feeds (i.e. mobile phone and TWT datasources) able to capture in a similar way the change in the human flows due to the evolution of COVID-19 cases in a geographical area?

Both mobility feeds have proved that the mobility of Spain during the lockdown period followed an incremental number of daily trips as long as the number of COVID-19 cases decreased. This has been reflected with clear negative PC scores in the three spatial scales (see Table 5). Therefore, the intuitive idea that the mobility in Spain, and thus the economic activity, was activated at the same time the COVID-19 pandemic subsided was confirmed by both the TWT and the INE study. However, the graph centrality measurements showed that the INE study was much more reliable to correlate the COVID-19 cases with the actual relevance of a region in mobility terms.

5 Conclusion and Future Works

Human mobility studies are gaining focus on computer-based research related to urban systems as an alternative to improve our understanding about how people move within a city, a region or a country.

In this paper we aimed to demonstrate the use of Online Social Networks (OSNs) as a compelling and reliable source of human mobility data at different granularity levels. In particular, we used Twitter to infer human mobility patterns at three different national-wide levels in Spain during the COVID-19 lockdown situation.

To demonstrate the utility of this OSN in this scenario we compared the results of the Twitter analysis with a dataset provided by a Spanish official mobility study based on mobile phone data during the same period. As a result, Twitter has been found a reliable mobility indicator to capture short distance trips at a fine-grained spatial level, whereas inter-province trips involving long distances have been found difficult to detect in comparison with the mobile phone-based study. Moreover, the use of Twitter has also enabled the detection of an increment of the mobility patterns during the COVID-19 lockdown scenario as the number of cases decreased and the mobility restriction policy became more flexible, in the same level as it was inferred from the official mobility study.

Regarding the applicability of the results, these findings could be used to develop two different types of applications based on TWT mobility data. First, the number of trips estimated from TWT could be used by authorities and stakeholders to proactively adjust the ticketing prices of certain public means of transport considering the estimated demand from short-distant trips. Next, the information on the origins and destinations of long-distance trips could foster the development of smart tourism applications able to analyze the origin of most of the visitors of a particular geographical area, and for example help in the design of more effective tourism marketing campaigns.

The most imminent future work in this line is the exploration of other OSNs such as Foursquare and Yelp as well as other open data sources such as economic data activity to better capture the mobility patterns during the period of this study and the subsequent partial lockdowns taking places nowadays.

Funding Financial support for this research has been provided under grant PID2020-112827GB-I00 funded by MCIN/AEI/10.13039/501100011033, the Fundación Séneca del Centro de Coordinación de la Investigación de la Región de Murcia under Project 20813/PI/18 and by the Spanish Ministry of Science, Innovation and Universities under Grant RTC2019-007159-5.

Code Availability The source code of the project is available at https://github.com/fterroso/comparison_TWT_INE.

Declarations

Conflict of Interests Authors declare no conflict of interests.

References

1. Ahmouda A, Hochmair HH, Cvetojevic S (2019) Using twitter to analyze the effect of hurricanes on human mobility patterns. Urban Sci 3(3):87
2. Alessandretti L, Aslak U, Lehmann S (2020) The scales of human mobility. Nature 587(7834):402–407

3. Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Human mobility: Models and applications. Physics Reports 734:1–74. https://doi.org/10.1016/j.physrep.2018.01.001, http://www.sciencedirect.com/science/article/pii/S037015731830022X. Human mobility: Models and applications
4. Bisanzio D, Kraemer MU, Bogoch II, Brewer T, Brownstein JS, Reithinger R (2020) Use of twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of covid-19 at global scale. Geospatial Health 15(1)
5. Bisanzio D, Kraemer MU, Brewer T, Brownstein JS, Reithinger R (2020) Geolocated twitter social media data to describe the geographic spread of sars-cov-2. Journal of Travel Medicine 27(5):taaa120
6. Bomfim R, Pei S, Shaman J, Yamana T, Makse HA, Andrade Jr JS, Lima Neto AS, Furtado V (2020) Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. Journal of the Royal Society Interface 17(171):20200691
7. Chatfield C, Xing H (2019) The analysis of time series: an introduction with R. CRC press
8. Chen Z, Gong Z, Yang S, Ma Q, Kan C (2020) Impact of extreme weather events on urban human flow: A perspective from location-based service data. Computers, Environment and Urban Systems 83:101520. https://doi.org/10.1016/j.compenvurbsys.2020.101520, http://www.sciencedirect.com/science/article/pii/S0198971520302532
9. Crols T, Malleson N (2019) Quantifying the ambient population using hourly population footfall data and an agent-based model of daily mobility. GeoInformatica 23(2):201–220
10. Cuenca-Jara J, Terroso-Sáenz F, Valdés-Vela M, Skarmeta AF (2020) Classification of spatiotemporal trajectories from volunteer geographic information through fuzzy rules. Appl Soft Comput 86:105916. https://doi.org/10.1016/j.asoc.2019.105916, http://www.sciencedirect.com/science/article/pii/S1568494619306970
11. Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise, pp 226–231
12. Fan Z, Pei T, Ma T, Du Y, Song C, Liu Z, Zhou C (2018) Estimation of urban crowd flux based on mobile phone location data: A case study of beijing, china. Computers, Environment and Urban Systems 69:114–123. https://doi.org/10.1016/j.compenvurbsys.2018.01.005, http://www.sciencedirect.com/science/article/pii/S0198971517302636
13. Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry, pp 35–41
14. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. Nature 453(7196):779–782
15. Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C (2014) Geo-located twitter as proxy for global mobility patterns. Cartogr Geogr Inf Sci 41(3):260–271
16. Hu X, Wang C, Wu J, Stanley HE (2020) Understanding interurban networks from a multiplexity perspective. Cities 102625:99
17. Huang X, Li Z, Jiang Y, Ye X, Deng C, Zhang J, Li X (2021) The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the us during the covid-19 pandemic. International Journal of Digital Earth 14(4):424–442
18. Jiang Y, Li Z, Ye X (2019) Understanding demographic and socioeconomic biases of geotagged twitter users at the county level. Cartography and Geographic Information Science 46(3):228–242
19. Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D (2015) Understanding human mobility from twitter. Plos One 10(7):e0131469
20. Kurkcu A, Ozbay K, Morgul E (2016) Evaluating the usability of geo-located twitter as a tool for human activity and mobility patterns: a case study for nyc. In: Transportation research board's 95th annual meeting, pp 1–20
21. Langville A, Meyer C (2006) Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Oxford. https://doi.org/10.2307/j.ctt7t8z9
22. Lenormand M, Bassolas A, Ramasco JJ (2016) Systematic comparison of trip distribution laws and models. J Transp Geogr 51:158–169
23. Li M, Westerholt R, Fan H, Zipf A (2018) Assessing spatiotemporal predictability of lbsn: a case study of three foursquare datasets. GeoInformatica 22(3):541–561
24. Liu Y, Sui Z, Kang C, Gao Y (2014) Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. PloS One 9(1):e86026
25. Luo F, Cao G, Mulligan K, Li X (2016) Explore spatiotemporal and demographic characteristics of human mobility via twitter: a case study of chicago. Appl Geogr 70:11–25
26. Lv L, Zhang K, Zhang T, Bardou D, Zhang J, Cai Y (2019) Pagerank centrality for temporal networks. Phys Lett A 383(12):1215–1222
27. Martí P, Serrano-Estrada L, Nolasco-Cirugeda A (2019) Social media data: Challenges, opportunities and limitations in urban studies. Computers, Environment and Urban Systems 74:161–174.

https://doi.org/10.1016/j.compenvurbsys.2018.11.001, http://www.sciencedirect.com/science/article/pii/S0198971518302333
28. Orakzai F, Pedersen TB, Calders T (2021) Distributed mining of convoys in large scale datasets. GeoInformatica 25(2):353–396
29. Pourebrahim N, Sultana S, Niakanlahiji A, Thill JC (2019) Trip distribution modeling with twitter data. Computers, Environment and Urban Systems 77:101354. https://doi.org/10.1016/j.compenvurbsys.2019.101354, http://www.sciencedirect.com/science/article/pii/S019897151930119X
30. Provenzano D, Hawelka B, Baggio R (2018) The mobility network of european tourists: a longitudinal study and a comparison with geo-located twitter data. Tourism Review
31. Sotomayor-Gómez B, Samaniego H (2020) City limits in the age of smartphones and urban scaling. Computers, Environment and Urban Systems 79:101423. https://doi.org/10.1016/j.compenvurbsys.2019.101423, http://www.sciencedirect.com/science/article/pii/S0198971519301814
32. Spirtes P, Glymour CN, Scheines R, Heckerman D (2000) Causation, prediction, and search. MIT press
33. Tempelmeier N, Dietze S, Demidova E (2020) Crosstown traffic-supervised prediction of impact of planned special events on urban traffic. GeoInformatica 24(2):339–370
34. Terroso-Saenz F, Muñoz A, Arcas F (2020) Land-use dynamic discovery based on heterogeneous mobility sources. Int J Intell Syst 36(1):478–525
35. Tolouei R, Psarras S, Prince R (2017) Origin-destination trip matrix development: Conventional methods versus mobile phone data. Transportation Research Procedia 26:39–52
36. von Mörner M (2017) Application of call detail records - chances and obstacles. Transportation Research Procedia 25:2233–2241. https://doi.org/10.1016/j.trpro.2017.05.429, http://www.sciencedirect.com/science/article/pii/S2352146517307366. World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016
37. White DR, Borgatti SP (1994) Betweenness centrality measures for directed graphs. Social Networks 16(4):335–346
38. Xu Y, Chen D, Zhang X, Tu W, Chen Y, Shen Y, Ratti C (2019) Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. Computers, Environment and Urban Systems 75:184–203. https://doi.org/10.1016/j.compenvurbsys.2019.02.002, http://www.sciencedirect.com/science/article/pii/S0198971518305192
39. Zhang Y, Sari Aslam N, Lai J, Cheng T (2020) You are how you travel: A multi-task learning framework for geodemographic inference using transit smart card data. Computers, Environment and Urban Systems 83:101517. https://doi.org/10.1016/j.compenvurbsys.2020.101517, http://www.sciencedirect.com/science/article/pii/S0198971520302507

Fernando Terroso-Sáenz obtained his B.S. and PhD in Computer Science from the University of Murcia in 2009 and 2013. Since 2017 he is associate professor at Catholic University of Murcia (UCAM). He has published than 30 articles in international journals congresses. His research areas include smart mobility, human-generated data analysis and mobile sensing.

Andrés Muñoz received the Ph.D. degree in computer science from University of Murcia in 2011. He is currently an Assistant Lecturer at University of Cadiz. His main research interests include applications in smart environments, semantic Web technologies and data fusion.

Francisco Arcas is Ph.D. degree in Computer Science from University of Murcia in 2008. He is currently Senior Lecturer at Catholic University of Murcia. Interested in knowledge representation in general and natural language processing in particular.

Manuel Curado obtained his PhD in Computer Science from the University of Alicante in 2018. Since 2020 he is associate professor at Catholic University of Murcia (UCAM). His main research interests include pattern recognition and computer vision.