POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

Unraveling Changes of Pre- and Post-COVID Car Usage Through Mobility Network Data Analysis

Master of science in Mathematical Engineering

Author: Luca Meda

Advisor: Prof. Silvia C. Strada

Co-advisors: Prof. Sergio M. Savaresi

Academic year: 2022-2023

1. Introduction

The Covid-19 pandemic had a disruptive effect worldwide and different countries adopted specific containment measures with the common thread of limiting the citizens' possibility to travel for their daily purposes. The unique characteristics of this historical period lead to many research works trying to quantify the impact of the pandemic diffusion on people's mobility, while also trying to interpret its new trends in the post-pandemic years, [2].
A specific interest revolves around the study of private mobility in relation to the contrasting effects emerged during and after the pandemic period: on the one side, car is still in 2022 the preferred mode of transportation among Italian citizens with a 1.5% increase compared to 2019, [1], but on the other hand the diffusion of virtual communication mitigates possible increases in traffic despite the aforementioned increase in car usage, [3].
To address these topics, in this thesis we rely on real driving data collected over a period of 4 years, from 2019 to 2022, from on-board telematics devices installed on private vehicles in the provinces of Brescia, Milano and Pavia. The available timespan guarantees access to data covering pre- and post-pandemic periods, beside covering all the months in which mobility restrictions were enforced.
This thesis presents the development of a innovative methodology framework to address a longitudinal analysis of trips data based on Mobility Networks models, enabling, from one side, the study of collective movements of the vehicles composing the datasets, on the other hand the data-driven modelling framework enables the profiling of individual driving behaviour, capturing regularities and patterns among locations visited by single vehicles.

2. Data description

In this work we have access to data made available within a research collaboration between UnipolSai and PoliMove. Specifically, these data are collected by small telematics devices (Black Boxes), equipped with a GNSS receiver, which are firmly installed on insured cars and register high-frequency geo-coded data.

2.1. Events Dataset

The datasets consist of anonymized GNSS records of vehicles grouped according to their geographical source. The first contains records of

33,226 vehicles immatriculated in the province of Brescia collected from January 2019 to December 2020, and of 68,367 vehicles in the same province recorded from January 2021 to December 2022. The second refers to the provinces of Milano and Pavia, with 63,416 vehicles tracked from January to December 2019 and 167,502 from January to December 2022. Each record is stored approximately every two kilometers and it represents instantaneous information (event) about speed, position and the time and distance travelled from the previous recorded instant.

2.2. Trips-Stops Dataset

Sequences of events are aggregated to obtain the driving behaviour of each vehicle, namely the collection of complete trips and the relative stops between them. Each trip record provides information later exploited in the analysis, such as the trip distance, the duration, the average speed, the origin and destination coordinates and the duration of the following stop.

2.3. Datasets Matching Algorithm

Records belonging to the same vehicle are stored in different datasets anonymized with a different ID depending on the year they were recorded. Because of this setback, we devise an ad-hoc Matching Algorithm, with the clear intent of finding a connection criterion between the consecutive datasets in the same province, exploiting the data themselves.
This is achieved by picking out stops longer than six hours for each vehicle, and clustering these stops based on their proximity through the application of DBSCAN algorithm to identify important individual locations.
The algorithm proceeds to join together vehicles in different datasets presenting at least a common long-stops cluster, whose representative points have a distance lower than 50m. After the required cleaning step, we obtain 17384 vehicles in the province of Brescia with trips-stops data collected from January 2019 to December 2022, and 28204 vehicles in the provinces of Milano and Pavia with data available, collected only in the whole 2019 and 2022.

3. Mobility Network Models

Complex spatial network analysis gained popularity in modelling urban/regional interactions and understanding phenomena and regularities generated by human mobility, [7]. Graph representation can be indeed used to model mobility data in a geographical area of interest, while adopting powerful network science-enabled tools to study their characteristics and structural properties. In this thesis we develop two different categories of network-based models:

- **Global Mobility Networks**, which are designed to capture collective mobility insights, to study high-level characteristics of mobility interactions and distribution.
- **Individual Mobility Networks**, where a single vehicle's travelling behaviour is represented through a network graph structure, highlighting regularities and patterns in individual routines.

3.1. Geo-localized Origin-Destination matrices

To study private mobility in this work, we rely on Origin-Destination (OD) matrices.
In formal terms, an OD matrix F is a n x n matrix where n is the number of possible Origin and Destination zones, while F_{ij} is the number of vehicles travelling between i and j in the time considered. The necessary condition in this work's setting is the domain knowledge definition of the spatial tessellation of a geographical area, from which the n is derived and every origin and destination can be localized.

3.2. Global Mobility Networks

The construction of Global Network models relies on the building elements of a network graph, meaning the tuple G = (N,E,W) where N are nodes, E are edges and W weights of edges.

- **Nodes** in the network correspond to each area of the spatial partition on which the OD matrices are based, so n nodes are selected corresponding to every origin/destination in the respective matrix.
- **Edges** are the undirected links connecting nodes in the network, where an edge (i,j) connects node i to node j.
- **Weights** are associated to every edge (i,j) with a formula derived from the OD matrix definition, with ΔT the specific time domain, as each network is built from the OD matrix computed from trips data hav-

ing their departure date on day t:

$$w_{ij}(t) = F_{ij}(t) + F_{ji}(t) = w_{ji}(t) \qquad t \in \Delta T$$

- **Lengths** of edges are a network-based metric of distance between nodes, computed as the reciprocal of weights.

We also introduce a specification of this network model denominated **Commuting Global Mobility Networks**. These networks are introduced to put a deeper focus on those time ranges during weekdays where car movements are supposed to be mainly dominated by work necessities. These networks are characterized by *directed edges*, and their weights present an updated formula exploiting hourly OD matrices, differentiating between morning (6 AM,9AM) and evening (4 PM,7 PM) time ranges.

3.3. Individual Mobility Networks

As defined in [6], an *Individual Mobility Network (IMN) describes the individual mobility of a vehicle through a graph representation of its trips and visited locations, grasping relevant properties of individual mobility while removing unnecessary details.*

In formal terms, an IMN is a directed weighted network graph. In this setting however nodes are defined with a data driven approach designed specifically for each vehicle; the result is no longer a uniform partition of the geographical domain, but instead nodes are identified as visited locations, where a location is a proxy for a subjective place of interest in which the vehicle remains parked.

In each individual network locations are obtained clustering together the vehicle's stops coordinates based on their spatial proximity, considering a specific time slot (day, month, year). To this purpose we apply again DBSCAN algorithm. Once nodes are defined, an edge e is simply defined as:

$$e = (l_i, l_j) \ \ \forall \ l_i \in L, \ \ i \neq j$$

where L is the set of visited locations by the vehicle. Denoting by m the cardinality of the set of trips performed by the vehicle in the time slot considered, with t_k each trip and with (c_s^k, c_e^k) the start and end coordinate of the trip t_k, finally we define weights associated to each edge in the network as:

$$w_{ij} = \sum_{k=1}^{m} t_k, \quad t_k = 1 \ \ if \ c_s^k \in l_i \wedge c_e^k \in l_j$$

For every node in an IMN we also find an unsupervised semantic annotation of locations: home and work nodes.

4. Mobility analysis during the Covid-19 pandemic

In this section we exploit the study of vehicles in the province of Brescia, in 2020 and 2021, to examine the state of private mobility during the two years in which mobility was affected by state and regional-level restrictions.

The first macro effect involves the huge drop in the absolute number of trips which reshaped car-travelling in a more local sense. In Figure 1, we show the 7 days moving average for the quantity $\Delta f(t)$, defined as:

$$\Delta f(t) = \left(\frac{F(t)}{F_0(t)} - 1 \right) * 100$$

where F(t) is the total number of trips on day t ($t \in [2020, 2021]$) and $F_0(t)$ is the average number of trips in the pre-pandemic reference year 2019 in the same month and weekday of t. We divide the curves in 4 different trips distance ranges to highlight how long-distance trips are heavily affected by restrictions.

The pandemic effects on private mobility are reflected by the Global Mobility Network perspective with an increased sparseness observed in the network and the formation of local clusters, to form a lattice structure.

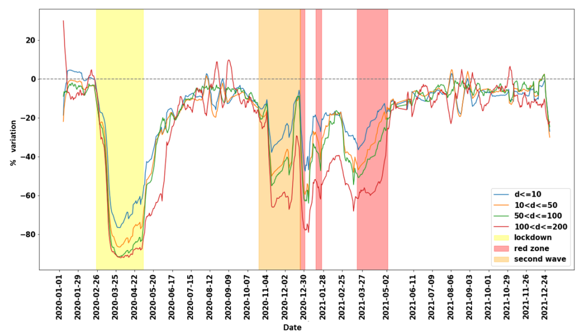


Figure 1: Relative percentage variation of total number of daily trips divided by distance ranges, d, with respect to the same month and weekday in the pre-Covid year 2019

In Figure 2 we can indeed observe the time evolution of the two parameters determining this structural property, namely the ratio of average shortest path length L(T) and average Clustering coefficient $C^w(T)$ in 2020 and 2021 compared to the respective average values in March 2019. We recall that path length is the distance between nodes according to the network-based definition of length, while clustering coefficient measures the local cohesiveness of an area.

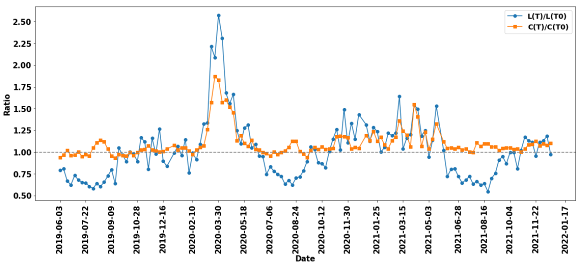
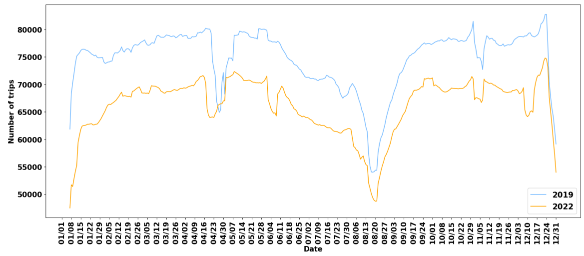


Figure 2: Average Weekly values of L(T) and C^w(T) for the Global Mobility Network relative to the reference pre-Covid March 2019 (T0)

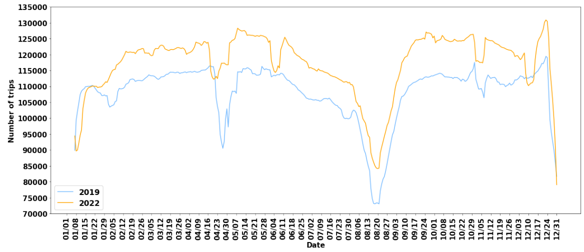
A common effect to all the network-based metrics is the substantial return towards a pre-pandemic condition as soon as pandemic limitations are lifted.

5. Pre- and post-Covid-19 mobility comparison

We also provide a comparison, in terms of overall and individual mobility indicators, between the pre-pandemic year 2019 and the post-pandemic year 2022, in Brescia and Milano-Pavia.
Very interesting differences emerged from the two datasets. In Brescia, the daily total number of trips falls below the values recorded in 2019, with an average decrease in 2022 equal to 9.5% (std = 3.4), as seen in Figure 3. We also report a decrease, stable throughout 2022, of the daily number of active vehicles (i.e. vehicles recording at least a trip in a given day), with an average drop of 5.2% (std = 2.3). In Milano-Pavia we instead give evidence to an increase in the total number of trips registered daily, with an average growth of 8.5% in 2022 vs 2019 (std = 4.8), alongside an higher number of active vehicles daily through the whole 2022, on average 4.5% (std = 3.1).



(a) Comparison of 7-days moving averages of total number of trips in Brescia dataset (2019 vs 2022)



(b) Comparison of 7-days moving averages of total number of trips in Milano-Pavia dataset (2019 vs 2022)

Figure 3: 7-days moving averages of total number of trips (2019 vs 2022)

The data reported above brought as a consequence also a change in the daily frequency of trips for vehicles on workdays.
No evidence instead emerges on changes in the distribution of trips across the 24 hours, and no statistically significant differences in the distributions on all vehicles of the median trip duration and median trip length on workdays.

5.1. Individual Networks based measures

The IMN model requires the introduction of a suitable set of properties, adapting for example to a number of nodes which is vehicle and time-dependent in each network. In a recent study, [5], the authors highlight six network metrics which are specific in individual mobility analysis and that have been proved to be stable with respect to the time partition. These six metrics are also designed to address the main questions we want to tackle in our analysis, that can be summarized in three main categories: **vehicle behaviour complexity**, **home and workplace importance** and description of **spatial characteristics**. The individual network metrics suitable to answer these questions are:

- **Degree distribution coefficient** and **weight distribution coefficient**, which measure the spread of vehicle's trips on its nodes and edges.

- **Median trip distance and 9th decile trip distance**, which measure how far the vehicle travels.
- **Average Journey Length** and **Hub Size**, which measure how many locations are visited habitually.

We test the differences of these quantities' mean between 2019 and 2022. In Brescia no significant evidence is computed, signalling a complete return to a pre-pandemic driving behaviour scenario. In Milano we register a significant decrease in the mean of average journey length and hub size and an increase in weight distribution coefficient. The effect of these combined coefficients highlights a slight reduction in the complexity of driving routine, with a lower number of locations visited with regularity and trips concentrated on a reduced number of nodes and edges.

5.2. Network based vehicle behaviour clustering

Moreover, we exploit the IMN model to cluster individual driving behaviour.
The algorithm requires at first the computation of the six network-based features and their normalization to z-scores for each vehicle. Then K-Means clustering algorithm is applied, with a procedure based on Mann-Withney U test designed to guarantee a better quality of the results, with the network features relative to 2019 for vehicles in both Brescia and Milano-Pavia. 8 clusters are found (2 of those are discarded because less than 1% of vehicles were assigned to them) and we introduced specific clusters' denominations to highlight their meaning: **Long Distance Driver, Short Distance Town-Based Driver, Short Distance Home-Based Driver, Commuter, Generic Driver and Medium Distance Driver**. In Figure 4 we show the scatterplot of the IMN features across the identified clusters.
After training a Random Forest Classifier on the cluster assignments from the 2019 data and applying it to the 2022 features for each vehicle, we observe that in Brescia the relative number of vehicles belonging to each group is substantially unaltered. In Milano-Pavia, instead, in 2022 a notable growth (14%) emerges in the number of vehicles that are classified as Short Distance Home-Based, the most regular and short-ranging driving profile.

6. Commuting with private car

Global Network Entropy

A methodological approach to measure mobility features related to work habits in an area is given by the application of entropy concepts to Mobility Networks, [4], specifically Commuting Global Mobility Networks, which were defined previously. The definition of global entropy reflects the structural homogeneity of a network and is derived from information theory as a measure of uncertainty inside probability distributions.

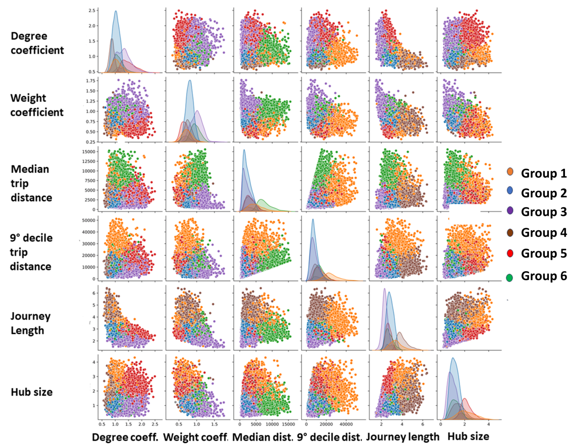


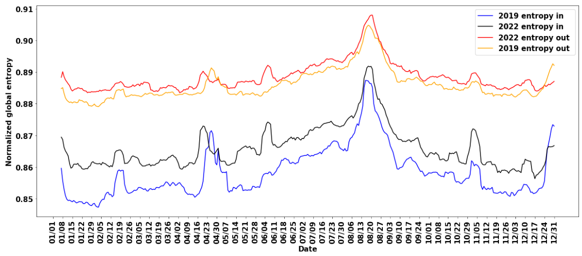
Figure 4: Scatterplot of the distribution of each vehicle's IMN features across identified vehicles' groups

Considering $p_{ij} = \frac{w_{ij}}{\sum_{i,j} w_{ij}}$ the probability of observing in the network an out-flow from node i directed to node j (we remind that w_{ij} is the weight on the edge from i to j), $\sum_j p_{ij}$ the probability of having an out-flow from node i and $\sum_i p_{ij}$ the probability of in-flow into node j, global node-level In-flow and Out-flow Entropies are defined as:

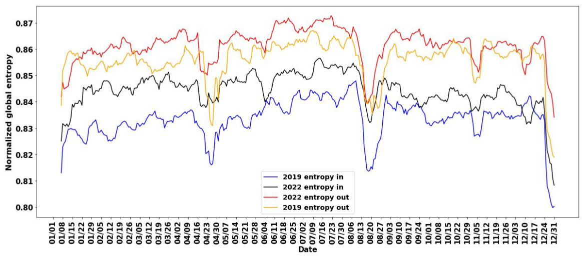
$$H_{GN}^{in} = -\sum_j \left(\sum_i p_{ij}\right) log \left(\sum_i p_{ij}\right)$$

$$H_{GN}^{out} = -\sum_i \left(\sum_j p_{ij}\right) log \left(\sum_j p_{ij}\right)$$

The situation in which we observe a monocentric system with one or few employment hubs leads



(a) $H^{in}_{GN}/log(n)$ and $H^{out}_{GN}/log(n)$ computed on Morning Commuting Global Networks of Brescia over a whole year (2019 and 2022)



(b) $H^{in}_{GN}/log(n)$ and $H^{out}_{GN}/log(n)$ computed on Morning Commuting Global Networks of Milano-Pavia over a whole year (2019 and 2022)

Figure 5: Global in- and out-entropies in Morning Commuting Networks compared between 2019 and 2022 in the provinces of Brescia and Milano-Pavia.

to a value of entropy $H^{out}_{GN} > H^{in}_{GN}$, while the difference between these two values indicates the degree of monocentricity of the structure. The opposite, namely $H^{in}_{GN} > H^{out}_{GN}$, would be the indication of a polycentric structure where destinations are spread out on various nodes while origins are concentrated in a smaller subset.
Analyzing global entropies in the morning commuting range in both Brescia and Milano-Pavia provinces, the first thing we notice in Figure 5 is that in both years evidently $H^{out}_{GN} > H^{in}_{GN}$, signalling that we are dealing with two employment hub-based mobility systems. We can observe that global in-entropy has increased in the post-pandemic year and it remains permanently higher over the year. This growth of in-entropy is the reflection of a tendency in the direction of a more uniform distribution of in-flows in the network, tendency that is confirmed by the fact that the gap between $H^{in}_{GN}/log(n)$ and $H^{out}_{GN}/log(n)$ is getting smaller in 2022.

6.1. Workplace activity

We apply the IMN model and the semantic representation of home and work nodes to compute for every vehicle the number of days in 2019 and 2022 in which a trip to work has been registered, to analyze if a sensible increase in remote-working has an effect after the end of the pandemic on private mobility. For Brescia the mean of the number of days in which vehicles have recorded at least a trip directed to a workplace decreases from 139.5 (std=99.6) to 127.8 (std=98.6) and the median from 160 to 138, while for Milano and Pavia the mean move from 128.6 (std=97) to 127.9 (std=92.4) and the median from 142 to 136. We realize that in 2022 a greater proportion of individuals reduces the usage of their vehicles to reach a destination classified as a work-node over the whole year.

7. Conclusions

In this work we develop a theoretical framework based on spatial and temporal networks, namely Mobility Network Models, to analyze both collective mobility scenarios and to provide informative insights on individual driving patterns. We design a complete data analysis pipeline to process telematics raw measurements into Origin-Destination matrices necessary to build the respective network models. The proposed methodology has then been applied to two case studies, namely the provinces of Brescia and Milano-Pavia to perform analysis on the changes in private mobility through the Covid-19 pandemic and after the end of the emergency situation. The results relative to 2020 and 2021 can be useful to investigate empirically the effectiveness of the rules enforced during the pandemic phases and can be used as a blueprint for more data-informed policies to adopt in a similar future scenario.
The pre- and post-Covid situations are investigated by exploiting data relative to years 2019 and 2022 to capture possible long-lasting changes in private mobility after the end of the pandemic restrictions. While we show different responses in 2022 in terms of general traffic intensity in the two provinces, we highlight some common trends. It emerges the tendency towards a decentralization of private mobility, with a reduction of trips' flows directed to the main employment hubs. At the same time, with our model we detect a decrease in the use of private vehicles to reach workplace locations.
Lastly, in the analysis of individual driving pat-

terns, we propose the application of a network-based clustering algorithm through which we identify six different driving behaviour profiles characterised by different Individual Mobility Network features. The identification of these profiles in our work contributes then in reflecting the state of private mobility, but in a broader sense it can be exploited to provide mobility solutions and services tailored to the driver's specific driving characteristics.

References

[1] 19° rapporto sulla mobilità in Italia. Technical report, ASSTRA - associazione trasporti, 2022.

[2] Jillian Anable, Llinos Brown, Iain Docherty, and Greg Marsden. Less is more: changing travel in a post-pandemic society. *CREDS Report Available online. last accessed 18th April*, 2022.

[3] Claudia Caballini, Matteo Agostino, and Bruno Dalla Chiara. Physical mobility and virtual communication in italy: Trends, analytical relationships and policies for the post covid-19. *Transport Policy*, 110:314–334, 2021.

[4] Valentina Marin, Carlos Molinero, and Elsa Arcaute. Uncovering structural diversity in commuting networks: global and local entropy. *Scientific Reports*, 12(1):1684, 2022.

[5] Henry Martin, Nina Wiedemann, Daniel J Reck, and Martin Raubal. Graph-based mobility profiling. *Computers, Environment and Urban Systems*, 100:101910, 2023.

[6] Salvatore Rinzivillo, Lorenzo Gabrielli, Mirco Nanni, Luca Pappalardo, Dino Pedreschi, and Fosca Giannotti. The purpose of motion: Learning activities from individual mobility networks. In *2014 international conference on data science and advanced analytics (DSAA)*, pages 312–318. IEEE, 2014.

[7] Chen Zhong, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, 28(11):2178–2199, 2014.