Graph-based mobility profiling

Author(s):
Martin, Henry; Wiedemann, Nina; Reck, Daniel Jan; Raubal, Martin

Publication date:
2023-03

Permanent link:
https://doi.org/10.3929/ethz-b-000584320

Originally published in:
Computers, Environment and Urban Systems 100, https://doi.org/10.1016/j.compenvurbsys.2022.101910

Computers, Environment and Urban Systems 100 (2023) 101910

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Graph-based mobility profiling

Henry Martin [a,b,*,1], Nina Wiedemann [a,1], Daniel J. Reck [c], Martin Raubal [a]

[a] Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland
[b] Institute of Advanced Research in Artificial Intelligence (IARAI), Vienna, Austria
[c] Institute for Transport Planning and Systems, ETH Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:
Mobility as a service
Clustering
Individual mobility behavior
Mobility graphs

ABSTRACT

The decarbonization of the transport system requires a better understanding of human mobility behavior to optimally plan and evaluate sustainable transport options (such as Mobility as a Service). Current analysis frameworks often rely on specific datasets or data-specific assumptions and hence are difficult to generalize to other datasets or studies. In this work, we present a workflow to identify groups of users with similar mobility behavior that appear across several datasets. Our method does not depend on a specific clustering algorithm, is robust against the choice of hyperparameters, does not require specific labels in the dataset, and is not limited to specific types of tracking data. This allows the extraction of stable mobility profiles based on several small and inhomogeneous tracking data sets. Our method consists of the following main steps: Representing individual mobility using location-based graphs; extraction of graph-based mobility features; clustering using different hyperparameter configurations; group identification using statistical testing. The method is applied to six tracking datasets (Geolife, Green Class 1 + 2, yumuv and two Foursquare datasets) with a total of 1070 users that visit about 3′000′000 different locations with a total tracking duration of over 200′000 days. We can identify and interpret five mobility profiles that appear in all datasets and show how these profiles can be used to analyze longitudinal and cross-sectional tracking studies.

1. Introduction

Individual motorized transportation is a major contributor to global greenhouse gas (GHG) emissions [Chapman, 2007; Creutzig et al., 2015] and linked to additional problems such as the creation of microplastics [Evangeliou et al., 2020], injuries, an increase of impervious cover for infrastructure [Gössling, 2020], and more traffic and congestion which already results in high economic costs [Reed, 2019].

Tackling these problems requires to cover people's growing mobility needs using less resources like energy, cars, or space. This is the goal of several novel mobility concepts such as mobility-as-a-service (MaaS) but apart from being more sustainable, these services will need to be comparable with personal cars in terms of comfort and flexibility in order to convince people to change their mobility behavior.

This challenge will require knowledge about the mobility behavior of people and the ability to predict it in the near future in order to optimally allocate shared mobility resources. With the recent success of machine learning algorithms [LeCun et al., 2015], research in computational movement analysis [Long et al., 2018] shifted towards using machine learning methods to support data interpretation (e.g., labeling [Toch et al., 2019], clustering tasks [Ben-Gal et al., 2019, Jonietz et al., 2018]) or prediction tasks [Luca, Barlacchi, Lepri, & Pappalardo, 2023; Kumar and Raubal, 2021; Kreil et al., 2020].

While large tracking datasets of human movement have become available in recent years, they are oftentimes unlabeled [Chen et al., 2016], thereby preventing the use of supervised machine learning methods. Furthermore, available datasets are often different in key properties such as the duration of a tracking study, the deployed tracking technology, and its spatio-temporal resolution of trackpoints, and sample biases.

A particularly difficult problem in this situation is the identification of stable groups of users with similar mobility behavior, which are comparable across datasets. Finding such mobility types can help us to enhance our understanding of mobility behavior [Pappalardo et al.,

Abbreviations: MaaS, Mobility as a Service; GNSS, global navigation satellite system; GPS, global positioning system; TG, treatment group; CG, control group; CDR, call detail record; LBSN, location based social network; GHG, greenhouse gas.

* Corresponding author at: ETH Zurich, Institute of Cartography and Geoinformation, HIL D 54.3, Stefano-Franscini-Platz 5, CH-8093 Zürich, Switzerland.
E-mail address: martinhe@ethz.ch (H. Martin).

[1] Authors contributed equally.

https://doi.org/10.1016/j.compenvurbsys.2022.101910
Received 4 March 2022; Received in revised form 17 August 2022; Accepted 19 October 2022
Available online 30 November 2022

2015], to measure regional similarities between cities McKenzie and Romm [2021] or neighborhoods Calafiore et al. [2021], and to detect changes in mobility behavior over time [Jonietz et al., 2018; Hong et al., 2021]. However, existing solutions are often based on dataset-specific features and can either not be applied to different datasets, or results from different datasets are not comparable.

In this work, we develop an approach to identify mobility types with minimal dataset specific assumptions, which facilitates the application to different datasets. The problems stated above are solved with a graph-based approach that uses a compact representation, does not require labeled data, and allows to easily merge different datasets. Our method is tested on six datasets to demonstrate its general validity independent of the specifics of one tracking study.

In summary, our contributions are the following:

- We propose a set of features that are based on a compact graph representation. They describe integral dimensions of individual mobility behavior and are robust to dataset properties such as tracking duration or spatio-temporal resolution of trackpoints.
- We develop a method that uses statistical testing on multiple clustering results of the same dataset and yields stable user groups.
- We apply our method to six tracking datasets and extract five mobility profiles that appear in all datasets. These profiles are robust against the parameters and initialization of the clustering algorithm.
- We demonstrate in two applications how to use graph-based mobility profiles to analyze longitudinal and cross-sectional tracking studies.

The remainder of this paper is structured as follows: In Section 2 we describe related research on human mobility profiling, clustering, and graph representations. In Section 3, our graph features and the clustering approach are explained. Next, in Section 4 the data and preprocessing steps are outlined, and the results and applications are presented and discussed in Section 5. Section 6 includes further experiments that validate our proposed methodology. Finally, we summarize our conclusions in Section 7.

2. Related work

2.1. Representing individual human mobility

In transport planning, human mobility is commonly modeled based on the hypothesis that travel demand is derived from the need to perform different activities at different locations [Jiang et al., 2017]. This activity-based perspective interprets travel demand as a result of people's decisions whether, where and when to perform activities [Axhausen and Gärling, 1992; Castiglione et al., 2015]. In practice, data about human mobility are often collected passively to avoid asking users to perform time consuming labelling tasks. Therefore, additional information such as activity labels are often not available in datasets [Chen et al., 2016].

To circumvent this problem, most approaches fall back on available tracking information such as the activity location as proxy for the *true* activity. A common way to represent an individual's mobility behavior is based on a sequence of visited locations such as the concept of location history mentioned in [Yu et al., 2009] or the concept of lifeline beads introduced in [Hornsby and Egenhofer, 2002]. In this case the movement profile of a person is a list of locations, ordered by the time of visitation. Depending on the definition, the model can include context data for each visit such as temporal information like start time or duration, spatial information such as coordinates, or semantics such as an associated POI category. Some further variations of this model can be found in [Bhattacharya and Das, 2002].

A major downside of this representation is that it grows quickly in size because the raw data are appended to the sequence for every visit. Furthermore, this representation is privacy sensitive as it contains information such as the time and duration of each individual visit. Representing individual movement profiles using a location graph of visited locations can solve these problems as it can be stored and processed efficiently. In such a graph, nodes correspond to visited locations (as a proxy for activities) and edges correspond to the transition count between two locations. Alessandretti et al. [2018] showed that people only visit a limited set of locations that slowly evolves over time and Schneider et al. [2013] demonstrated that our daily mobility can be described by a small set of sequential location visiting patterns (motifs). Furthermore Yan et al. [2017] created a model based on a graph of visited locations that reproduced important scaling laws of human mobility. This provides evidence that a personalized graph that is based on the visited locations can parsimoniously represent individual human mobility.

Graphs based on visited locations of individual persons have already been explored in the past. For example, [Yu et al., 2008] transformed GPS tracking data into a graph representation to support the prediction of the transport mode of transitions between nodes. [Rinzivillo et al., 2014] transformed a large GPS tracking data set of about 150 k vehicles from Tuscany into individual graph representations. They then combined structural features extracted from the graphs and classical features, such as length or duration, to show that including graph features increases the performance of trip purpose classification. Furthermore, [Martin et al., 2018] used graph representations of individuals in combination with graph neural networks to predict the distribution of activity labels at visited locations. Even though all these examples show promising results, the literature in this area is still sparse, especially with regards to unsupervised learning applications such as the identification of groups with similar mobility behavior based on graphs.

2.2. Clustering based on mobility behavior

Research on the identification of similarities based on movement data is mostly used for the discovery of previously unknown patterns and insights. Studies that do not focus on individuals often analyze movement at a city scale, such as in [Yuan and Raubal, 2012] where CDR data enriched with demographic data are used to classify different urban areas in a city in China. Their approach allows to identify areas of the city in which people move alike. Similarly, Ratti et al. [2006] analyze urban activities from mobile phone data in Italy, and Sulis and Manley [2018] use a combination of twitter data and smart card data to cluster places in London according to their travel activity patterns, which can be used to analyze the *daily rhythms* of places in a city.

Studies that focus on the movement of individuals usually present workflows that are used to mine patterns from specific situational datasets such as in [El Mahrsi et al., 2016] where the authors use public transport smart-card data to cluster users by their travel behavior with respect to time and frequency of trips. They identify 13 different passenger clusters which they further analyze to identify fine grained commute patterns. Xin and MacEachren [2020] present a methodology to extract mobility patterns to characterize different groups of football fans from twitter data. These studies are insightful; however, the methods often rely on very specific features such as the mode of transport of a trip or the content of a twitter message related to a trip destination. This makes the methods difficult to apply to different datasets or other types of tracking data where the required information might be unavailable. There are exceptions such as [Pappalardo et al., 2015] who group individuals in *returners* and *explorers* based on their (k-) radius of gyration. These groups can be found across many datasets, but their work does not contain an approach to identify novel groups based on mobility behavior that generalize over different datasets, and further research showed that the results may depend on dataset properties such as the study duration [Wang et al., 2021]. A potentially generalizable clustering approach is presented in Ben-Gal et al. [2019] who develop a lifestyle-based clustering method. They identify five patterns, namely *home, sweet home*, *working 9 to 5*, *traveling salesmen*, and *commuters* based on a large CDR dataset. Their activity-based approach could be applied

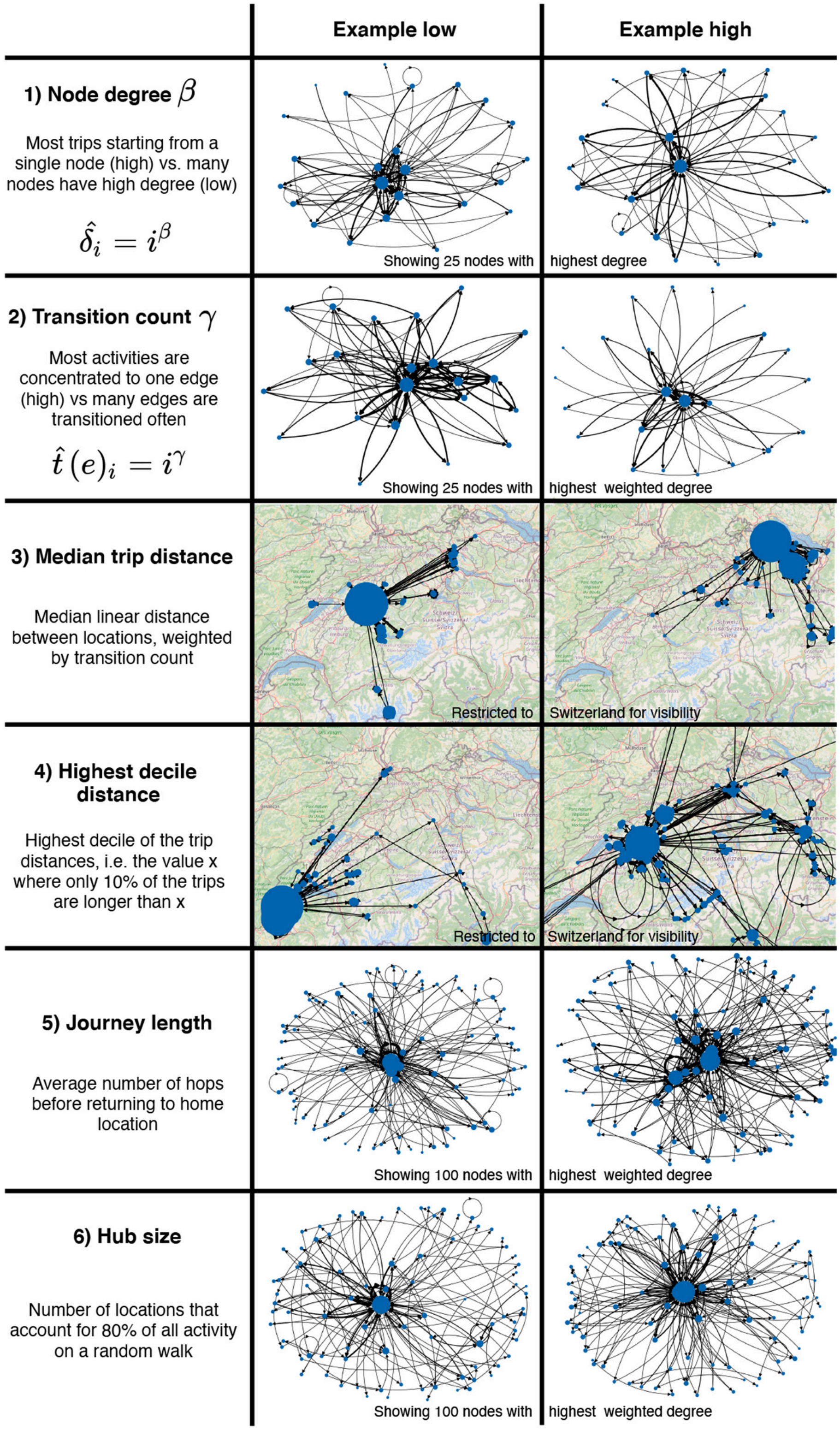


Fig. 1. Overview of graph features with an example graph from the Green Class 1 dataset for which the feature is rather low or rather high. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

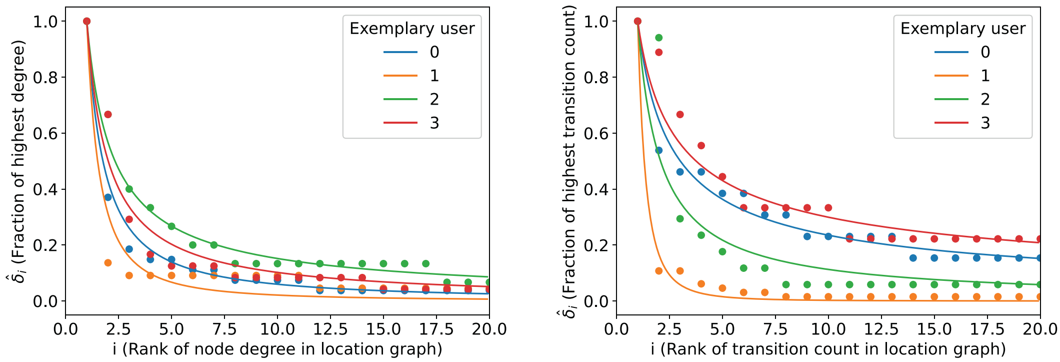
to different datasets but they do not provide any analysis in this regard.

We fill this gap by presenting a workflow that allows to identify groups of individuals based on their mobility behavior. Our approach only requires minimal assumptions on tracking data, no labels and it permits to integrate different tracking datasets to identify their overall user groups.

3. Methods

3.1. Location based graph representation

In contrast to sequential tracking data, graph representations are compact, privacy-preserving, easy to process and still rich in



(a) Fit node degrees (b) Fit transition counts

Fig. 2. Power law fit for location graphs. The degrees are normalized by the highest degree found and ranked, and a power law is fitted.

information. Motivated by these properties, we choose a graph-based representation of individual human mobility. The graph is constructed based on the location sequence of a person, where a location is seen broadly as a place of interest that a person visited to perform an activity. The location sequence of a user $L_{seq} = [l_1^1, l_2^2, l_1^3 \ldots, l_m^n]$ is a list of visits at locations ordered by their visitation time. L_{seq} contains n visits to m unique locations and l^i is defined as the i^{th} element of L_{seq}. Based on the location sequence we define the set of all visited locations as $L = \{l_1, l_2, \ldots, l_m\}$ as the collection of all visited locations without repetition, therefore $|L| = m$.

The graph is constructed using unique locations as nodes and the number of direct transitions between pairs of nodes as weighted and directed edges. More formally, given the location sequence and the set of all visited locations of a person we define the weighted directed *individual location graph* as the pair (G_L, W) where

$$G_L = \{L, E\}, \text{with}\, e = (l_i, l_j) \in E(G_L)\ \ \forall\ \left(l_i^k, l_j^{k+1}\right) \in L_{seq}\ \mid\ k < n \qquad (1)$$

and the elements w of the weight matrix $W \in \mathbb{R}^{|L| \times |L|}$ are

$$w_{ij} = \sum_{k=1}^{n-1} \theta \quad \text{with} \quad \theta = \begin{cases} 1 & \text{if } \left(l^k, l^{k+1}\right) = (l_i, l_j) \\ 0 & \text{otherwise.} \end{cases} \qquad (2)$$

Examples of individual location graphs are given in Fig. 1, where the transition count is proportional to the edge line width. Creating the graph representation only requires L_{seq}, it does not require specific label or context information. It can therefore be applied to datasets that differ in properties such as the data collection methods (e.g., global navigation satellite system (GNSS) vs. CDR). A notable exception would be a significant bias in the sampling of visited locations as it might be present for example in public transport smart card data, where only visits at public transport stations are recorded and important locations such as the home location are missing systematically. This would lead to a systematically different graph structure that is incomparable to graphs based on other collection methods.

The graph representation compresses the location sequence of an individual significantly as we are mostly revisiting known locations [Gonzalez et al., 2008; Alessandretti et al., 2018]. However, despite the compression we know from previous work that the topology of location graphs is highly unique for each individual [Manousakas et al., 2018] and that human mobility can be well represented by substructures of such a graph [Schneider et al., 2013].

3.2. Graph-based mobility features

In order to characterize human mobility, we leverage the topology of the graph representation. Network characteristics extracted from the individual location graph can yield insights into a user's mobility behavior, despite relying on a compressed version of the raw movement data.

We propose a set of non-redundant and interpretable features that each represent separate dimensions of human mobility behavior. The features are motivated by a set of questions that address individual mobility behavior along the dimensions of the *role of base locations*, the *complexity*, the *regularity* and the *geometry* of individual mobility behavior. The numbers in square brackets link to the corresponding features from Fig. 1 that relate to the specific question.

Role of home bases:

- Does a person have a single home base where he starts her trips from or several such bases? [1, 6]
- How home-centered is the person's behavior? Does he return home after each activity, or rather move from place to place? [1, 5]

Complexity:

- Are the activities of the person focused on few locations and trips, or distributed over many? [1, 2, 6]
- Are most trips of the user between the same locations? [2]

Regularity and geometry:

- Is the user flexible, or does he have a very regular mobility behavior? [5, 2]
- How far does a user usually travel? How far does he travel exceptionally? [3,4]

3.2.1. Node degree β

The feature *node degree* β measures whether users start most of their trips from a single location or have several base locations. This will allow us to distinguish users who prefer to return home before visiting a new location from users who are more flexible, e.g., go to different places directly after being at work.

The (unweighted) out-degree of a node u in the location graph is defined as the number of locations that are visited starting from node u. If a single node has a very high out-degree compared to the other nodes, the user starts most of his trips from the same location. If multiple nodes have a high out-degree, the user has several *base locations* from where he starts his trips.

The importance of locations ranked by visitation frequency follows a power law distribution [Gonzalez et al., 2008]. We therefore propose to fit a power law to the distribution of ranked node out-degrees. Let $\delta_1, \delta_2, \ldots, \delta_n$ be the degrees of a graph with n nodes, sorted in descending order ($\delta_1 \geq \delta_2 \geq \ldots \geq \delta_n$). The values are normalized by the highest degree: $\hat{\delta}_i = \frac{\delta_i}{\delta_1}$. We then fit a simple power law following the rule $\hat{\delta}_i = i^{-\beta}$, where

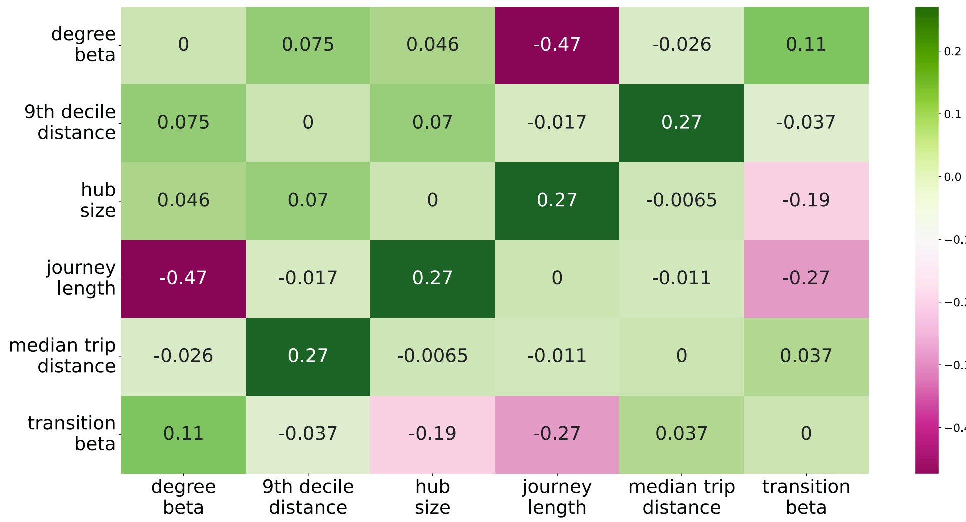


Fig. 3. Correlation between features.

$0 \le \beta \le 5$ ($\beta > 5$ is not realistic in our data and was thus excluded to accelerate the optimization). Note that we defined $\widehat{\delta}_1 = 1^{-\beta} = 1$.

A small β therefore describes graphs that have a multi-hub behavior, meaning that users start trips from several base locations. A high β characterizes user behavior that is centralized to few base locations. Fig. 2a shows the power law fit for exemplary location graphs and Fig. 1 shows example graphs of a user with high and a user with low *node degree* β.

3.2.2. Transition γ

This feature measures whether most trips (i.e., transitions between two locations) of a user are between the same locations or between many different locations. This information provides insights about the variety of a person's mobility behavior and visiting patterns. The transitions are stored as edge weights w_{ij} in the location graph, and their rank distribution follows a power law. Thus, we can measure this feature with the parameter of a power law distribution fitted on the sorted transition weights, where the weights are normalized by the maximum ($\max_{i,j} w_{ij}$). Examples are shown in Fig. 2b. In the commuter example, we expect high *transition* γ, in contrast to low *transition* γ for a salesman because trips are very distributed.

3.2.3. Median and 9th decile trip distance

The *median trip distance* describes how far users usually travel in their day-to-day mobility while the 9th *decile trip distance* describes the travel behavior for non-everyday trips. The distribution of trip distances follows a power law [Brockmann et al., 2006]. It is therefore important to use robust measures such as the median or the 9th decile as metrics.

The median or 9th decile are computed over the distances between each pair of nodes, weighted by their transition counts w_{ij}. Note that we use the Haversine distance between node coordinates and no map-matched distances, since the latter cannot be recovered from the location graph alone and would require information on the transport mode.

3.2.4. Average journey length

The average *journey length* of a user measures how flexible a user is in moving from place to place. If a user oftentimes visits multiple places without returning home, his graph will become more connected and show more (highly weighted) edges. We measure this quantity as the number of visited locations in a journey, where a journey is defined as a simple cycle in the location graph that starts and ends at the home location following the definition of journeys from [Axhausen, 2007]. We propose to approximate the journey length using a random walk in the graph: Starting at the home location node, we conduct a random walk of 5000 steps. Assuming that the current location is l_i, we select the next node j with probability $p(j) = \frac{w_{ij}}{\sum_k w_{ik}}$, i.e., proportional to the transition counts. When reaching a node that does not have any outgoing edges, or only an edge pointing at itself, we reset the random walk to home. The *journey length* is then defined as the number of steps between each consecutive encounter of the home location, excluding the resets.

3.2.5. Hub size

With the *hub size* feature we measure how many locations are visited on a regular basis and thereby account for a significant portion of the user's activity. It is therefore a measure of concentration of the mobility behavior (on few or many locations) and by that of the user's flexibility. In the graph, we measure the *hubbiness* as the number of nodes required to account for at least 80% of the total visits. The feature can be approximated from a random walk, similarly to the *journey length*. A random walk of 5000 steps yields a list of visited locations $L_{random}[l^1, l^2, \ldots, l^{5000}]$. The locations are sorted by their occurrence count (c_l) in L_{random}, such that $c(\widehat{l}_1) \ge c(\widehat{l}_2), \ge \ldots$. The (unnormalized) *hub size* h^* is the required number of locations such that their counts sum up to >4000 occurrences (80% of 5000 steps), formally $h^* = min_h : \sum_i^h c(\widehat{l}_i) \rangle 4000$. Since this number increases with the size of the graph, we normalize the feature by the square root of the total number of locations, $h = \frac{h^*}{\sqrt{|L|}}$. We chose a square root as it has been shown that the number of locations that are important to a person can be characterized with sub-linear exponential growth [Alessandretti et al., 2018]. The importance values assigned to each node by this method correspond to the PageRank value of a node [Page et al. [1999], Schütze et al. [2008]], with the slight variation that the implemented random walk always starts at the home node and restarts at the home node if it hits a dead end.

A visual summary of all features is given in Fig. 1. In addition, Fig. 3 shows the correlation matrix of all features of the users of six datasets combined as described in Section 5. It demonstrates that only few features are significantly correlated. Apart from the obvious correlation between *median* and *9th-decile trip distance*, we find that *journey length* is negatively correlated with *degree* β (−0.47) and *transition* γ (−0.27), and positively correlated with *hub size* (0.27). Intuitively, if there are many nodes with high degree (=low *degree* β) the probability to encounter longer cycles in a random walk (=high *journey length*) increases. Nevertheless, we decided to keep both features to further distinguish users with a flexible mobility behavior. Specifically, a high *average journey length* characterizes users that visit several locations in a row, independent of the locations' node degree. The *journey length* feature is

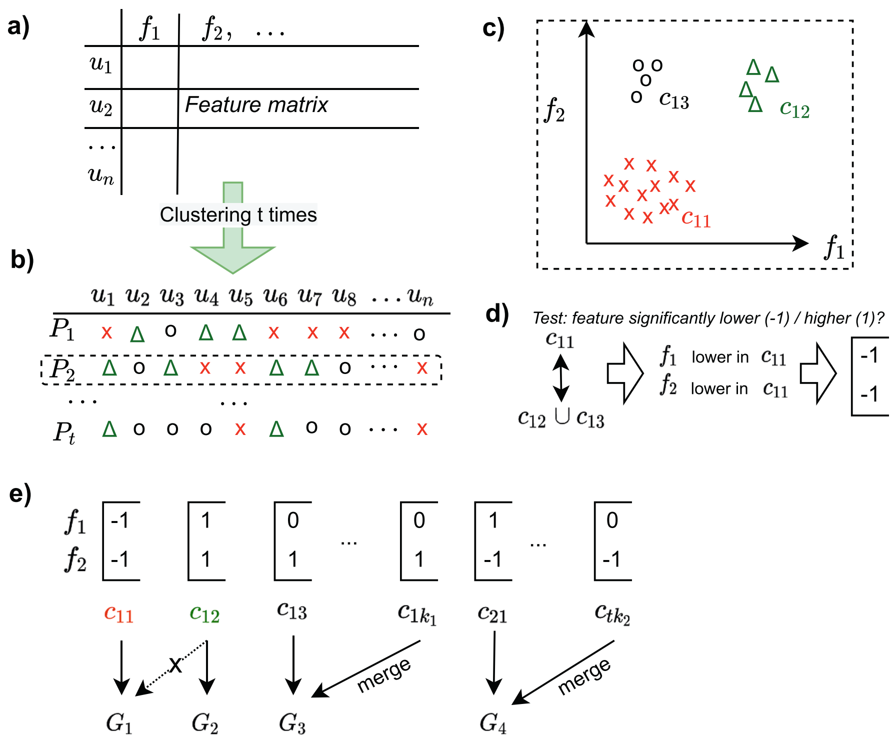


Fig. 4. Workflow of user identification via clustering. a) Features f_i are extracted from the location graphs of each user u_i and form a feature matrix. b) A clustering algorithm is applied t times with different random initialization and parameters, yielding t different partitions of the users. In c), one such partition P is shown schematically for two features. d) By means of a statistical test, we determine for each feature whether the values in one cluster are significantly different from feature values in all other clusters of the partition. e) Based on the significant features, the clusters of all partitions can be merged to an existing group or define new groups. The result is a set of groups G with every cluster assigned to one group.

also included due to its robustness to the tracking duration (cf. Section 6.1).

3.3. Identifying user groups

3.3.1. User group definition based on statistical testing

Given a clustering algorithm and m mobility features $f_1, \ldots, f_m$, we aim to find meaningful user groups of distinct mobility behavior. Unsupervised machine learning methods can identify patterns in high-dimensional feature spaces based on a given distance metric. However, clustering methods are oftentimes sensitive to initialization or to their parameters, e.g., the number of desired clusters. To overcome these problems, we propose a method based on statistical testing that yields stable user groups. The method is explained visually in Fig. 4.

Based on a fixed set of features and users, the clustering algorithm is run t times with different parameter choices or randomization, yielding multiple partitions $P_1, P_2, \ldots, P_t$ where each $P_i = \{c_{i1}, \ldots, c_{is}\}$ is the set of clusters that define the i-th partition (see Fig. 4a and b). One such partition is shown in Fig. 4c. We then consider all clusters in each partition and apply a suitable statistical test to determine which features are significantly different from the other clusters (Fig. 4d). Let $F_j(c_{ik})$ denote the distribution of feature f_j in cluster c_{ik}. We test the hypothesis that $F_j(c_{ik})$ does not differ from $F_j(P_i \setminus c_{ik})$, i.e., the distribution of f_j in all other clusters.

Next, the test results are combined in a m-tuple for each cluster, called $g(c_{ik})$, which defines a potential user group. The entries of $g(c_{ik})$ describe for each feature whether it is significantly lower (-1), higher (1) or not significantly different (0) from the other clusters. For example, given the features height, age and weight of a person, the tuple $g(c_{ik}) = (0, 1, -1)$ would denote that the people in cluster c_{ik} are not of significantly different height, but are significantly older and of lower weight than the people in other clusters.

3.3.2. Merging similar user groups

In the next step, we aggregate clusters to user groups (Fig. 4e). In the aggregation step, the assignment of a cluster c_{ik} is independent of its partition P_i and we therefore drop the partition-index to simplify notation in this section, i.e., a cluster is simply denoted as c_k. Two clusters are of the same user group if their significant features do not contradict each other. Formally, we define a valid merge of two clusters c_k, c_l as the following: Let $[g(c_k)g(c_l)]^-$ be the result of subtracting $g(c_k)$ from $g(c_l)$ element-wise. If any of the values in $[g(c_k)g(c_l)]^-$ is 2 or -2, then the clusters are not merged. The intuition is that if one or more features are opposite to each other (low in one but high in the other), the clusters are dissimilar. However, with this definition it could still occur that, for example, $g(c_k) = (1,1,0)$ is merged with $g(c_l) = (0,0,1)$ because $[g(c_k)g(c_l)]^- = (1, 1, -1)$, even though not a single significant feature corresponds. Thus, as a second requirement, they are only merged if they share at least θ_{minf} significant features, i.e., at least two elements of $[g(c_k)g(c_l)]^-$ are zero. By definition of this merging process, the maximum number of resulting groups is 2^m.

3.3.3. Iterative group finding and assignment

In practice, we distinguish an iterative group-finding phase and an assignment phase. In the group finding phase, we start with $c_{11} \in P_1$ and store $G_1 = g(c_{11})$ as the first user group. For c_{12} it is then checked whether a merge with G_1 is possible; otherwise, a new group $G_2 = g(c_{12})$ is added. All subsequent clusters are assigned to the existing groups or serve as new groups. Clusters with less than θ_{minf} significant features are skipped. Note that this rarely occurs in practice and the effect of θ_{minf} is therefore limited, as analyzed quantitatively in appendix D.

After the groups have been identified, we iterate over all clusters a second time and assign each cluster c_{ik} the group with best correspondence which is defined as the largest overlap between characteristic tuples. Formally, let $g[j]$ denote the entry at the j-th position in the tuple g. Then the group of c_{ik} is assigned to the group with index l^*, $l^* = \arg\max_l \sum_j g(c_{ik})[j] \cdot G_l[j]$. Finally, each user is assigned to the group that occurred most often in her clusters. Two groups can occur equally often. However, as ties were not frequent (<5%), this case was disregarded, and ties are solved by randomly assigning a user to one of the groups.

Table 1
Overview of the datasets used in this study. Column *users* shows the number of participants used in the study after filtering and the total number of available users. The columns *tracking days* and *visited locations* show the average and standard deviation over users.

Study	Tracking type	Users	Tracking days	Visited locations
Geolife	GPS tracker	66/177	301±392	168 ± 187
GC1	GNSS via app	134/139	401±60	756 ± 249
GC2	GNSS via app	47/50	321±63	718 ± 318
Foursquare home	LBSN checkins	88/100	477±94	219 ± 146
Foursquare random	LBSN checkins	82/100	413±116	88 ± 43
yumuv	GNSS via app	653/871	99±27	184 ± 86

4. Data and preprocessing

All preprocessing steps are performed using Python and the Trackintel movement data processing library[Martin et al. [2022]] which provides functionality to extract staypoints, triplegs, trips and locations. *Triplegs* are defined as continuous movement trajectories, *staypoints* are periods of stationary behavior. Staypoints are defined as activity if their duration is longer than 25 min or if there exists a non-trivial purpose label (any purpose except *wait* or *unknown*). *Trips* are defined as the collection of all movement and idling between two activities [Axhausen [2007]]. Trips with gaps longer than 25 min are considered to have an unknown destination as a person could have performed an activity in between, Locations are extracted using DBSCAN with 1 sample required for a cluster and a search radius of 30 m for GNSS datasets. See Table 1 for an overview of the used datasets.

4.1. Yumuv

The yumuv dataset was recorded during the roll-out of a new Mobility as a service (MaaS) offer in Zurich, Switzerland, called *yumuv* to study the impact of mobility bundles on mobility behavior [Martin et al., 2021]. The total study duration was three months and participants were either part of the treatment group (TG), which had access to the new MaaS offer available via the yumuv app after one month of pre-tracking, or part of the control group (CG). Both groups had to install the app MyWay,[2] a GNSS based tracking app, on their phone to record their mobility behavior. The app already provided staypoints and triplegs. The participants labeled triplegs with the used mode of transportation and staypoints with an activity label. Additionally, all participants took part in an online survey before and after the study period and provided person and household specific data such as socio-demographic information or mobility tool ownership. The dataset contains a total of 871 users (161 TG, 710 CG) of which 498 (71 TG, 427 CG) finished the study.

We additionally separate the dataset into four weeks *before* and four weeks *after* getting access to the MaaS bundle via the yumuv app. The exact dates and durations of the before and after period are slightly different for every user depending on when users started tracking and installed the yumuv app. For the users of the TG we use each individuals start and end date as defined above. As the users of the CG do not get access to a MaaS bundle, we use the average start and end date of all TG users as the start and end date for the CG.

4.2. Green class

The Swiss Federal Railways (SBB) conducted two large-scale 1-year pilot studies to evaluate the use of a comprehensive *all you can travel* mobility package [Martin et al., 2019]. In the pilot studies the participants had access to a general public transport pass valid in Switzerland, access to popular car- and bike-sharing programs and taxi vouchers. Additionally, participants of the first pilot study referred to as Green Class 1 (GC1) had access to a personal battery electric vehicle whereas participants of the second pilot study referred to as Green Class 2 (GC2) had access to a premium electric bike. Participants agreed to be tracked via the MyWay app and provided socio-demographic information in surveys.

4.3. Geolife

The Geolife GPS trajectory dataset was collected by Microsoft Research Asia over a span of three years [Yu et al., 2009]. Employees were provided with different global positioning system (GPS) loggers and GPS-phones that were used to passively track their everyday movement continuously. The dataset does not systematically provide additional label or socio-demographic information; however, it is still one of the few publicly available large-scale tracking datasets and is included to allow an easy reproduction of the results of this study. Staypoints are generated with Trackintel using the staypoint detection algorithm from [Li et al., 2008]. We used the parameters proposed by the authors for this dataset and we additionally added a threshold that excludes periods without trackpoints for >24 h as gaps.

4.4. Foursquare

The global scale Foursquare-dataset[3] presented in [Yang et al., 2015, 2016] is a vast collection of publicly available check-in data from the location based social network Foursquare. We chose to include the Foursquare dataset to showcase the possibility of the graph-based approach for non-GPS datasets. The full dataset contains check-ins of 144′704 users all over the world collected over the course of 18 months. Users track their movements by checking in at venues (e.g., points of interests). Data quality varies highly between users and not all users provided check-ins at their home location. Especially the second issue is problematic, as the structure of the location graph with a missing home location would be systematically different. We therefore create two subsets of the Foursquare dataset. The *Foursquare Home* subset consists of the 100 users with the most home check-ins in the dataset. As these are some of the most active users in the dataset, we further create the *Foursquare Random* dataset where we randomly draw 100 users from the 27′227 users that have at least 81 check-ins reported (above the 25th percentile), that checked-in at least at 40 different locations (above the 25th percentile) and that have at least 24 check-ins at home (above the 75th percentile).

4.5. Graph generation

The location graph for each person is generated as described in 3.1. Following the definition of trips given in Section 4.1, every trip of a person increases the edge weight between the two activity locations (i.e. nodes) by 1. The graph creation for the Foursquare dataset is slightly different as it uses check-ins without stay duration instead of continuous tracking. Here, all sequential check-ins at locations are used to increase the edge weight between two venues (nodes).

Before creating the graphs for the GNSS based datasets, we filter to include only users with at least 14 days with tracking coverage of

[2] https://play.google.com/store/apps/details?id=ch.sbb.myway

[3] https://sites.google.com/site/yangdingqi/home/Foursquare-dataset#h.p_ID_56

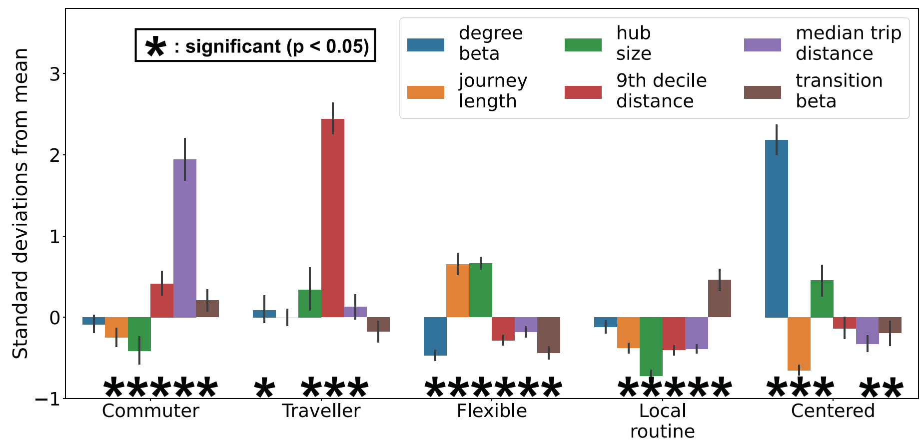


Fig. 5. Feature properties for each user group: The user groups that are consistently found in the data are named based on significant differences in their feature values with respect to the other clusters.

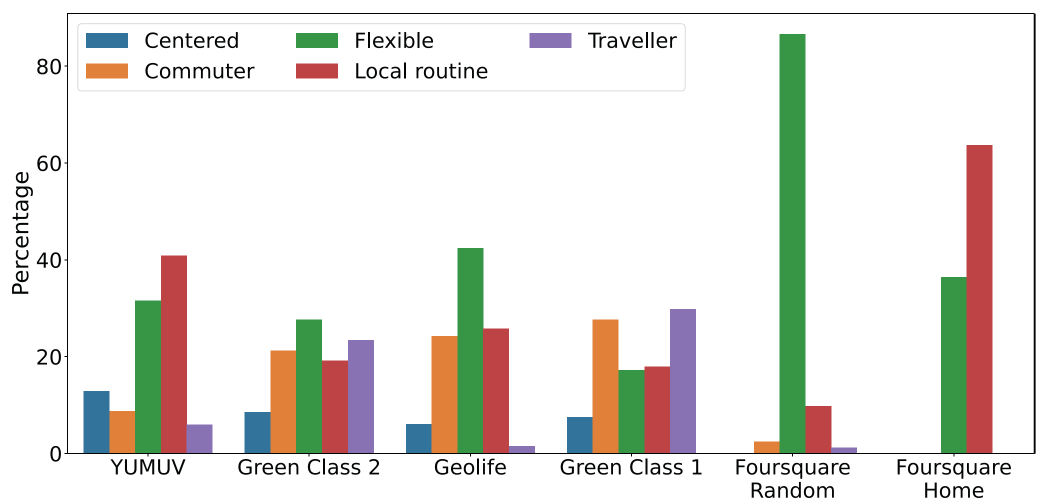


Fig. 6. User groups by study. With the exception of the Foursquare dataset, the user groups are similarly distributed. Differences can be explained by variations in the study target group, e.g., the Green Class 1 offer attracted individuals that cover longer distances. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

>70% of the time of the day).

5. Results and discussion

Given the location graphs for all users, we calculate the features as described in Section 3.2. We then combine the six datasets and classify users by their mobility behavior with K-means clustering and our group identification algorithm (Section 3.3). In the following, we first describe the obtained user groups, secondly, we present cross-sectional and longitudinal studies based on these groups, and finally validate our feature set and the method.

5.1. *Identification of user groups across datasets*

The aim is to study *universal* differences in mobility behavior that appear in diverse tracking studies. Therefore, the six datasets (GC1, GC2, yumuv, Geolife, Foursquare-Random and Foursquare-Home) are combined and processed in an analogous manner. We then exclude users if one or more of their feature values is more than four standard deviations apart from the mean value of that feature. By that 67 users or ~6% of the total users are excluded. The Foursquare-Random dataset has the highest outlier ratio with 18%.

Next, the features are normalized to z-scores.We proceed with the group identification algorithm introduced in Section 3.3. Here, the K-Means++ algorithm [Vassilvitskii and Arthur, 2006] is utilized for clustering, since it enforces compact clusters which are more likely to have significantly different feature values. Although the impact of the initialization of K-Means++ is significantly lower than for the original K-Means algorithm [Lloyd, 1982], we observed different outcomes depending on the initialization. Therefore, we vary both the random initialization as well as the number of clusters k. Specifically, we apply K-means three times for each $k \in [6,7,8,9]$, resulting in $t = 4*3$ partitions $P_1, \ldots P_{12}$. We test for significant difference in the feature distribution with a Mann-Whitney U test [Mann and Whitney, 1947]. Furthermore, we set $\theta_{minf} = 2$, such that clusters with less than two significant features are skipped. In the group finding phase (cf. Section 3.3.3) we identified six user groups. However, in the subsequent group assignment phase, the users were only assigned to five groups. In other words, one of the groups only appeared in few clustering runs and every user was assigned to one of the other five groups more often. The consistency of user-group assignments is analyzed further in Section 6.4.

5.1.1. *Interpretation of user groups*

To further analyze the identified groups, we inspect which feature

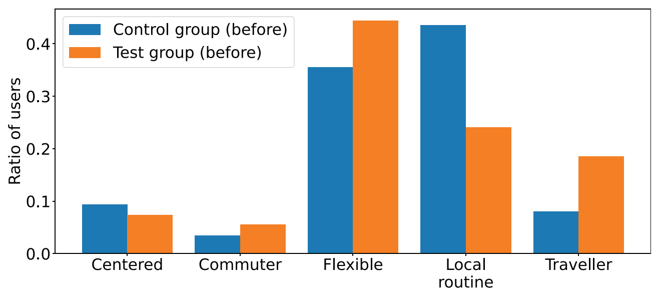


Fig. 7. Characteristics of yumuv users compared to the control group. Users assigned to the groups *Flexible* and *Traveller* are more likely to be interested in the yumuv MaaS offer.

dimensions are distinctive for a group. For that we calculate the deviation of the feature value of a single group from the mean of the distribution of this feature value from all other groups. Fig. 5 shows the deviation for each feature by group. For example, the *median trip distance* of the first group is more than two standard deviations above the average of this feature in all other groups. On the basis of Fig. 5 we interpret the clustering results and summarize each group's mobility behavior in one term. This group-naming is a subjective decision that is dataset- and context dependent; however, it greatly facilitates the communication of results to decision makers and the public. Here, we base our group description on the underlying mobility behavior that leads to a specific layout in the mobility graph as measured by the features described in Fig. 1.

The first two groups are clearly related to trip distances, where we can safely assume that users with a high *median distance* cover much distance on a regular basis (as *commuters* do), whereas the *9th decile* is expected to be high only for users that regularly cover very long distances. The third group is characterized by their *flexibility*, because their activities are highly distributed (high *hub size*), they move from place to place (high *journey length*) and their graphs are less concentrated on single nodes or edges. In contrast, the fourth group's activity is more skewed towards one or few trips (high *transition* γ) and few nodes (low *hub size*) and takes place at a lower radius (low *median distance*). The fifth group uses a single node (or few nodes) with high degree (high *degree* β) as a base and other activities are started from this center of the locations (low *journey length*). For a more detailed visualization of the groups, we refer to the scatterplot matrix of each pair of features in Fig. 11 in appendix A.

5.1.2. Comparison of user groups across datasets

One of the main contributions of this work is the identification of user groups across several datasets that do not depend on technical data set properties such as tracking technology or the study duration. To verify this hypothesis, we analyze the distribution of user groups over the different studies in Fig. 6. We observe that the groups are not study-specific as two groups appear in all studies and the other four groups appear in all studies but *Foursquare-Home*. This rules out the case where the clustering process identifies each dataset as its own cluster. In appendix F, we further use a logistic regression model to show that tracking duration and coverage have very little influence on the graph-based mobility profiles while they strongly influence the mobility profiles generated based on basic features from the literature (cf. Section 6.2).

The variation in group distribution over studies can be explained by actual differences between study populations: For example, the yumuv app attracted mainly young people living in urban areas while the Green Class studies had a focus on suburban professionals. An exception are the two subsets of the Foursquare dataset where only few user groups are present. In general, it is reasonable to assume that persons in the Foursquare datasets are rather young and live in cities, similar to users of the yumuv app. Therefore, it is not surprising that all three datasets share the same two majority classes *Flexible* and *Local routine*. One important difference between the datasets is certainly that many location based social network (LBSN) users do not check-in at their home location. However, even though the group distribution of Foursquare users who do check-in at home (Foursquare-Home) is closer to yumuv as the group distribution of Foursquare-Random, it is still significantly different. Further analysis revealed that Foursquare users rarely cover long distances and have a rather low *degree* β. At this point it is still unclear to which degree these differences can be attributed to a sample bias of the study participants or to a technical bias caused by the characteristics of check-ins compared to GNSS tracking.

Overall, we conclude that the differences in the group distribution over datasets can mainly be explained by differences in the mobility behavior of study participants. The presented method is therefore robust to changes in tracking techniques and reflects actual differences in mobility behavior.

5.2. Use cases of mobility profiling in MaaS applications

After having established stable mobility profiles, we can use them to answer questions that are typically of interest when analyzing longitudinal or cross-sectional tracking studies. Here, we consider questions that arise around the introduction of a novel MaaS offer:

- What are the target groups for the MaaS offer? (Section 5.2.1)
- How does access to a specific MaaS offer change mobility behavior over time? (Section 5.2.2)

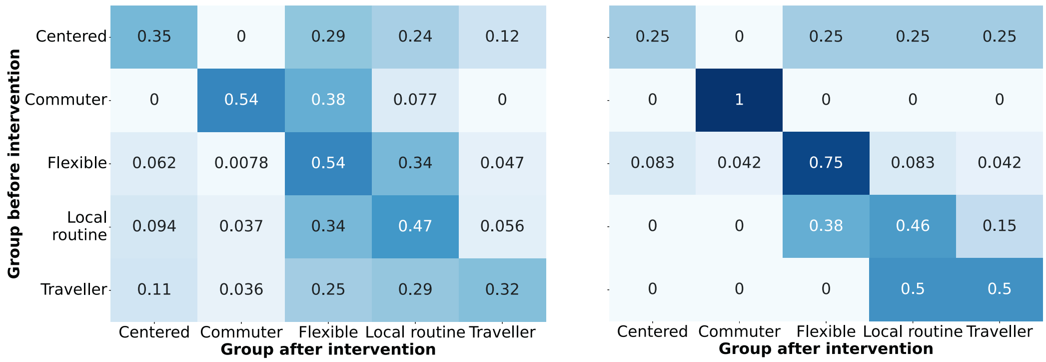
The yumuv dataset is very suitable for a case study due to the availability of distinct control group (CG) and treatment group (TG), and the availability of tracking data before and after access to the app. For the following analysis we split the yumuv dataset into four parts: TG-Before, TG-After, CG-Before and CG-After (cf. Section 4.5). For this analysis, we consider only participants that finished the study; after pre-processing and outlier-filtering the TG consists of 51 users and the CG of 372 users.

These four yumuv subsets were not part of the merged datasets D that were clustered (cf. Section 5.1). Consequently, the graphs of these subsets must be assigned to a user group first. However, the final user groups $G_{final} = G_1, \ldots, G_5$ resulting from the iterative group finding and assignment procedure do not have unique cluster center assigned to them, as they are the result of merging different partitions $P_1, \ldots, P_{12}$. We therefore chose the specific partition P_i out of all partitions $P_1, \ldots, P_{12}$ with the highest correspondence to the final user groups G_{final}, meaning that most users are assigned to the same group.

Here, in the best partition P_4 (with $k = 7$), 95% of the users were assigned to the same as their final user group. The graphs in subsets TG-Before, TG-After, CG-Before and CG-After are now assigned to a user group by finding their closest cluster center in P_4, as it is commonly done in K-Means clustering for new test data. These preprocessing steps yield a user group for all users per subset, such that the before- and after-group of one user may differ.

5.2.1. Cross sectional comparison

The assignment of user groups to each of the subsets (TG-Before, CG-Before) can now be used for a cross-sectional analysis. For mobility service providers and more general the design of MaaS offers it is important to know the target group of their offer. For this purpose, we compare the distribution of groups in the TG with the group distribution in the CG. We do this comparison for the period *before* access to the app in order to exclude possible confounding factors from the app usage that might affect mobility behavior. The comparison in Fig. 7 shows that persons who bought the yumuv offer (TG) are more often assigned to the groups *Flexible* and *Travellers*, whereas the *Local routine* group is more prevalent in the control group (distributions significantly different in χ^2



(a) Control group (b) Treatment group

Fig. 8. User group changes upon intervention (start of the yumuv offer).

Table 2
User group analysis with respect to demographic and mobility characteristics from study questionnaire. The mean values are given and compared to the other groups in a Mann-Whitney U test for continuous variables or a Chi-Squared test for categorical variables. Significant differences are marked bold, and PT denotes public transport. Note that all fields are self-reported in a questionnaire and not measured in the tracking study.

	Commuter ($n = 57$)	Traveller ($n = 39$)	Flexible ($n = 206$)	Local routine ($n = 267$)	Centered ($n = 84$)
[pht] Age	39.11 $p = 0.46$	37.72 $p = 0.351$	**37.86 $p = 0.022$**	40.7 $p = 0.059$	40.65 $p = 0.217$
Money spent on PT (in CHF per month)	**66.5 $p = 0.023$**	78.89 $p = 0.127$	94.19 $p = 0.243$	94.04 $p = 0.39$	107.0 $p = 0.311$
Home office (in days per week)	**0.86 $p = 0.034$**	**1.81 $p = 0.007$**	1.31 $p = 0.168$	**0.93 $p = 0.003$**	**2.19 $p = 0.002$**
Distance travelled by car yearly (in km)	11,840 $p = 0.106$	11,844 $p = 0.097$	**11,039 $p = 0.027$**	**7855 $p = 0.001$**	9656 $p = 0.413$
Satisfaction with PT reachability (%)	**79.12 p = 0.034**	86.25 $p = 0.364$	85.4 $p = 0.205$	87.22 $p = 0.079$	84.69 $p = 0.416$

test with $p = 0.02$). These target and non-target groups can now be characterized using Fig. 5 and using additional information such as demographics if available (cf. Section 5.2.3).

5.2.2. Longitudinal study

One of the main questions with regards to the introduction of MaaS offers is if and how they impact the mobility behavior of users [Hensher et al., 2021]. The assignment of user groups to participants before and after the intervention allows to analyze whether the group assignment of one user changes from the *before* period to the *after* period and allows to compare the changes between CG and TG.

Fig. 8 shows the changes between user groups from the period before intervention to afterwards. Each row is normalized to 1 and each cell shows the percentage of users that were assigned to a specific group before the intervention (row label) and moved to another group after the intervention (column label). We observe that more people in the treatment group switch towards the *Traveller* group than in the control group (cf. rightmost column). Furthermore, it seems that the *Flexible* group is more stable in the treatment group (cf. the values on the diagonal for the row *Flexible*). However, we compared the distributions row-wise with a χ^2 test and due to the small size of the Treatment group (51) there are no significant differences (the lowest p-values are $p = 0.09$ for the changes of the former *Flexible* group and $p = 0.23$ for former *Traveller*). Despite the lack of statistical significance this analysis still serves as a show-case of how to use the identified user groups for a longitudinal analysis. A graph-based visualization of the user group changes can be seen in Fig. 13 in appendix C.

5.2.3. Cluster analysis with respect to labels

The yumuv study also included surveys that collected socio-demographic and household information of the participants. In this section we analyze which of these features are significantly different for a specific group with respect to all other groups and therefore characterize that group. Table 2 shows the replies for each user group for a selection of relevant questions. Many results confirm the assumptions about mobility behavior that determined our naming of the clusters as user groups. For example, *Commuters* are less satisfied with the public transport connections to their home; they oftentimes travel by car and are seldom in home office. Only 37% of the *Commuters* live in cities. *Travellers* and *Flexible* users in contrast are younger, oftentimes live in cities and spend more days in home office. Interestingly, the *Centered* group works from home significantly more often than others. The fact that working behavior such as home office is reflected in the user groups provides evidence for a strong influence of the home and work locations on the graph features. Last, the naming of the group *Local routine* is reflected well in the users' self-reporting of their covered distance.

6. Validation

This section provides analyses to validate the method for extraction of generalizable user groups based on graph representations of individual mobility.

6.1. Feature robustness to study duration

An important factor in the feature selection process is their robustness to dataset properties. Here we investigate how much the feature values depend on the tracking duration. For this experiment, we split the Green Class 1 data into distinct bins of $t = 4,8,12,16,20,24$ and $t = 28$ weeks. Since the participants in the study were tracked for 56 weeks, $t = 4$ yields 14 non-overlapping bins and $t = 28$ is the maximum duration with two distinct bins. Next, we construct the location graphs from the activity of each user in each time bin, and compute the corresponding features. In Fig. 9 the mean feature values for all users per time bin are shown. Our selected features (top two rows) are largely robust to the tracking period, or converge after around $t = 12$ in the case of *degree* β and *transition* γ. In contrast, other considered features show a strong time dependency, such as the mean Eigenvector centrality in a graph or the

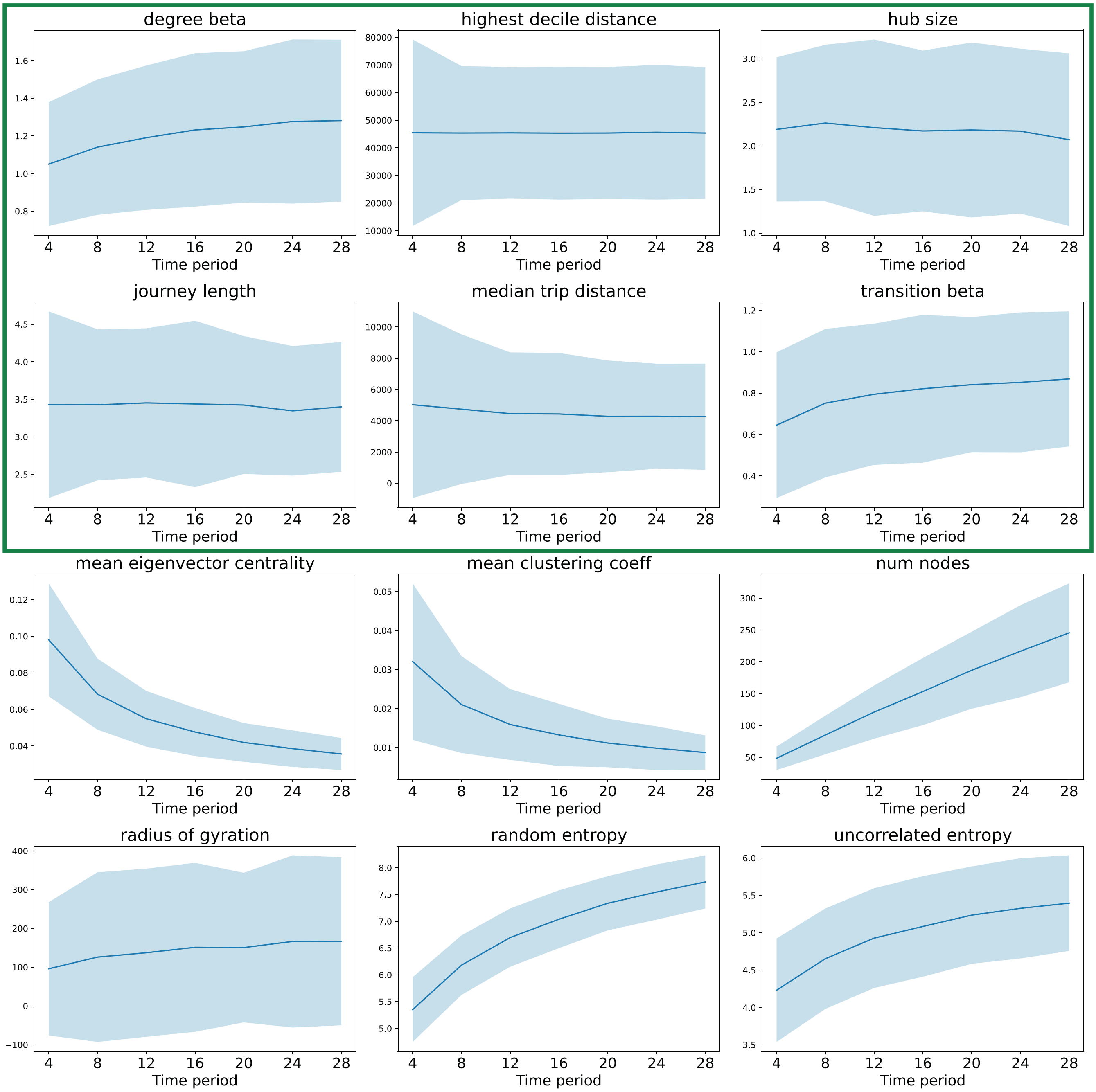


Fig. 9. Mean and standard deviation of different features with respect to tracking period. The selected features are within the green box. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

clustering coefficient as proposed by [Onnela et al., 2005]. Similarly, features of human mobility that are commonly used in the literature usually depend on the number of nodes and therefore change significantly with the tracking duration.

6.2. Relation of graph-based and classical mobility features

Furthermore, we investigate the relation of the proposed graph features to a selection of classical non graph-based mobility features that are commonly used to characterize mobility behavior. These features are directly calculated from the raw data and can therefore make use of information that is lost in the location graph, such as the order of the activities or their duration.

Concentrating on the most widely used measures, we consider the following features, which are termed *basic features* from here on:

- Number of visited locations
- Radius of gyration ([Gonzalez et al., 2008])
- Maximal distance from home
- Random, uncorrelated, and real entropy ([Song et al., 2010])
- Mean trip duration and distance

Except for trip distance and trip duration, we utilize the implementation in the scikit-mobility package [Pappalardo, Simini, Barlacchi, & Pellungrini, 2022].

For this experiment we use only the datasets Green Class 1 and 2,

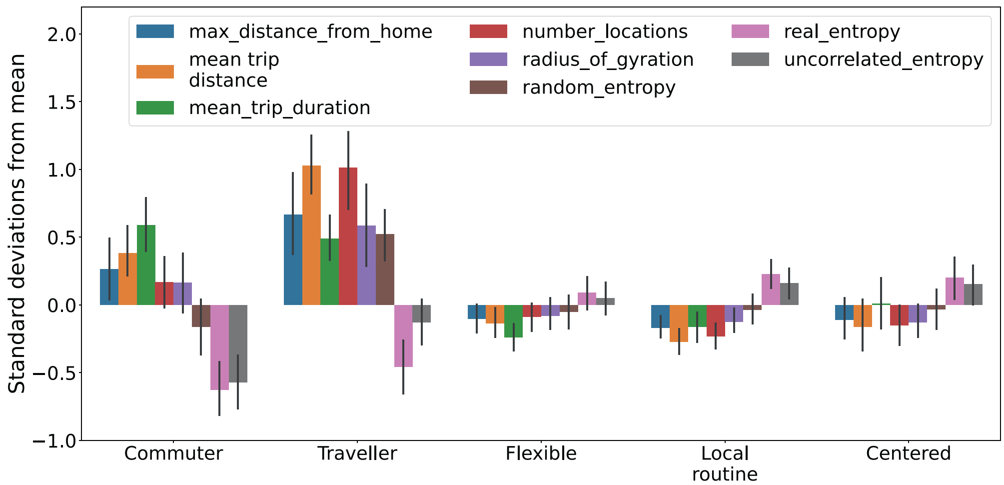


Fig. 10. Distribution of basic features over the identified user groups. On the one hand, differences in the basic features are also reflected in our user groups. On the other hand, our groups seem to identify further differences in mobility behavior that are hardly reflected in the basic features (e.g., group *Centered*).

yumuv and Geolife, because there is no trip information available for the Foursquare datasets (only check-ins). First, we were interested whether a clustering based on the basic features results in similar clusters as our user groups. We cluster the basic features with $k = 5$ and compare the resulting clustering to our user groups with the Adjusted Rand Index [Hubert and Arabie, 1985; Rand, 1971]. Intuitively, the Rand Index is proportional to the number of pairs that end up in the same cluster in both partitions, or in different clusters in both partitions. The Adjusted Rand Index is its normalized version that yields a value between −1 and 1. An index of 0 means that there is no relation between two partitions while the same partitions would yield an index of 1. Here, the similarity to our user groups is 0.08., i.e., the user groups found with graph features are fairly distinct from clusters that are identified with the basic features.

Secondly, Fig. 10 depicts the mean and standard deviation of the raw features in the proposed user groups. It can be observed that *Traveller* obtain significantly higher values also in these basic features, confirming the differences between groups also with respect to these basic measures.

We conclude that using graph features results in different user groups that can then be analyzed with respect to classical features.

6.3. Group robustness to cluster ordering

Last, although it was shown that all clusters of 12 partitions can be reduced to five user groups, we found that the resulting groups still depend on the *order* of considered clusters during the *iterative* group-finding phase. For example, the first cluster is always used as G_1, and other clusters may be merged with this particular cluster. The properties of G_1 thus depend on which cluster is considered first. Nevertheless, we qualitatively observed a strong stability of the resulting groups and their properties.

To quantify this stability, we perform 20 runs of the group-finding and assignment phases with different random seeds and compare the resulting groupings to the main grouping found in Section 5.1 on a per-user basis using the Adjusted Rand Index [Hubert and Arabie, 1985, Rand, 1971]. In our analysis, the Adjusted Rand Index of the pair-wise comparison of the groupings with the main grouping is 0.91 on average. Another initialization would therefore yield a very similar output, where the resulting groups could be named in a similar manner and a large majority of users would be assigned to the same group.

Table 5
Dependence of mobility user groups, derived from basic features, on tracking coverage and duration. The coefficients with p-values in brackets of a multinomial logistic regression model are shown. Significant coefficients are marked bold.

	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
Days tracked	**−30.04 (0.0)**	**32.7 (0.0)**	**32.78 (0.0)**	−18.13 (0.12)	**−13.66 (0.0)**	**−38.51 (0.02)**
Tracking coverage	6.11 (0.3)	**−40.0 (0.0)**	**−32.07 (0.0)**	**−21.11 (0.02)**	−5.39 (0.22)	−6.51 (0.66)
Intercept	−5.25 (0.35)	**31.98 (0.0)**	**24.59 (0.0)**	**17.87 (0.04)**	5.5 (0.19)	4.52 (0.74)

6.4. Consistency of group assignment

In the method presented in (cf. Section 3.3) the clusters resulting from the different runs are assigned to groups. A user can therefore belong to clusters that are assigned to different groups and is finally assigned to the group that most of his clusters are assigned to (cf. Section 3.3.2).

We now analyze how consistent this assignment is by computing a *consistency score* that indicates how often a user belonged to its most dominant group. The score is calculated by counting the number of times a user was assigned to its majority group divided by the total number of assignments.

In our case study, the average consistency score over all users is 0.87, i.e., an average user is assigned in 87% of the runs to its most dominant group. 80% of the users obtain a consistency >0.9.

7. Conclusion

Research on mobility behavior oftentimes suffers from a lack of reproducibility and transferability. Big tracking datasets are inherently noisy and usually unlabeled, and proposed methods do not generalize to other datasets. Here, we have presented an attempt to develop a generic clustering approach that yields stable mobility behavior groups on several diverse datasets. In contrast to previous work, we base our analysis on a compact graph representation of the tracking data, which 1) reduces memory resources needed to store long-term tracking data, 2) facilitates the comparability of different datasets, and 3) captures other aspects of mobility behavior than time-series based basic features. Based on six features that were shown to be particularly robust with respect to the time period, we apply a clustering algorithm multiple times and extract stable and interpretable user groups in an iterative fashion based on statistical testing.

Our analysis showed that five groups could consistently be found in the six datasets, which differ by the complexity, the role of home bases and the geometric extent of their mobility behavior. All user groups can be found in all studies except Foursquare, despite significant differences in the tracking quality, duration, and user demography of the studies. These user groups were also shown to be robust to the clustering parameters, consistent, and seem to depict novel aspects of mobility behavior that are not contained in classical mobility features. It is still unclear to what degree the differences of the Foursquare data are due to a sample bias of highly active urban LBSN users or due to systematic differences between GNSS based and check-in based tracking data. However, the Foursquare datasets did not generate exclusive user groups and could still be described by our framework. Furthermore, it could be shown that differences of the distribution of user groups also reflect differences of the target groups of each study. Such analysis is of interest to providers of MaaS offers to direct their services to the right people. Similarly, it could be shown that the effect of a MaaS offer on mobility behavior can be viewed in the context of user group changes over time. While a detailed analysis of the changes of location-graphs over time is out of scope of this paper, it is an interesting endeavor for future research. While the cluster analysis can be used to describe the change in mobility behavior over time, we noticed that this description of mobility behavior exhibits a higher volatility than expected, i.e., up to 50% of users change their group from one slot to the next. A possible reason for this could be that the mobility behavior did not fully stabilize after the considered tracking duration or clusters are overlapping which may lead to a certain number of samples that lie between two clusters and thus easily switch clusters over time. Further work could explore the possibility to connect our clustering approach with soft assignments where each sample belongs to multiple groups with certain probability.

Finally, it should be stressed that much of the proposed methodology is by no means restricted to mobility research. Some of the proposed features could be relevant in other fields where data is represented in graph structures, such as molecules in biology or computer networks (e.g., hub size as a descriptor of network activity). More importantly, clustering is a popular method used in many fields, and the identification of stable, statistically valid groups is a common problem. Our algorithm is a simple yet effective method to make clustering results more generic and reproducible.

Author statement

Martin, Henry: Conceptualization, Methodology, Software (preprocessing), Data collection, Writing - Original Draft; Writing - Review & Editing.

Wiedemann, Nina: Conceptualization, Methodology, Software (experiments), Writing - Original Draft; Writing - Review & Editing.

Reck, Daniel: Data collection, Writing - Review & Editing.

Raubal, Martin: Supervision, Writing - Review & Editing.

Data availability

The geolife and foursquare datasets are publicly available others can not be shared (yumuv, GC1+GC2). The code for this study is publicly available https://github.com/mie-lab/Graph-based-mobility-profiling.

Acknowledgement

Funding: This work was supported by the ETH Zurich Foundation [MI-01-19].

Appendix A. Feature exploration

For a more in-depth understanding of the distribution of features over user groups, we provide the scatterplot matrix in Fig. 11. The two largest groups are *Flexible* and *Local routine* which differ mostly in the *hub size* feature. The *Flexible* group also has a striking difference from the other groups with respect to the *journey length* feature. *Commuters* and *Travellers* are clearly distinguished by their high *median* and *9th-percentile trip distance* respectively. In contrast, the group *Local routine* has a particularly left-skewed distribution in the distance-based features. Last, the *Centered* profile is clearly characterized by the high node degree β in this group.

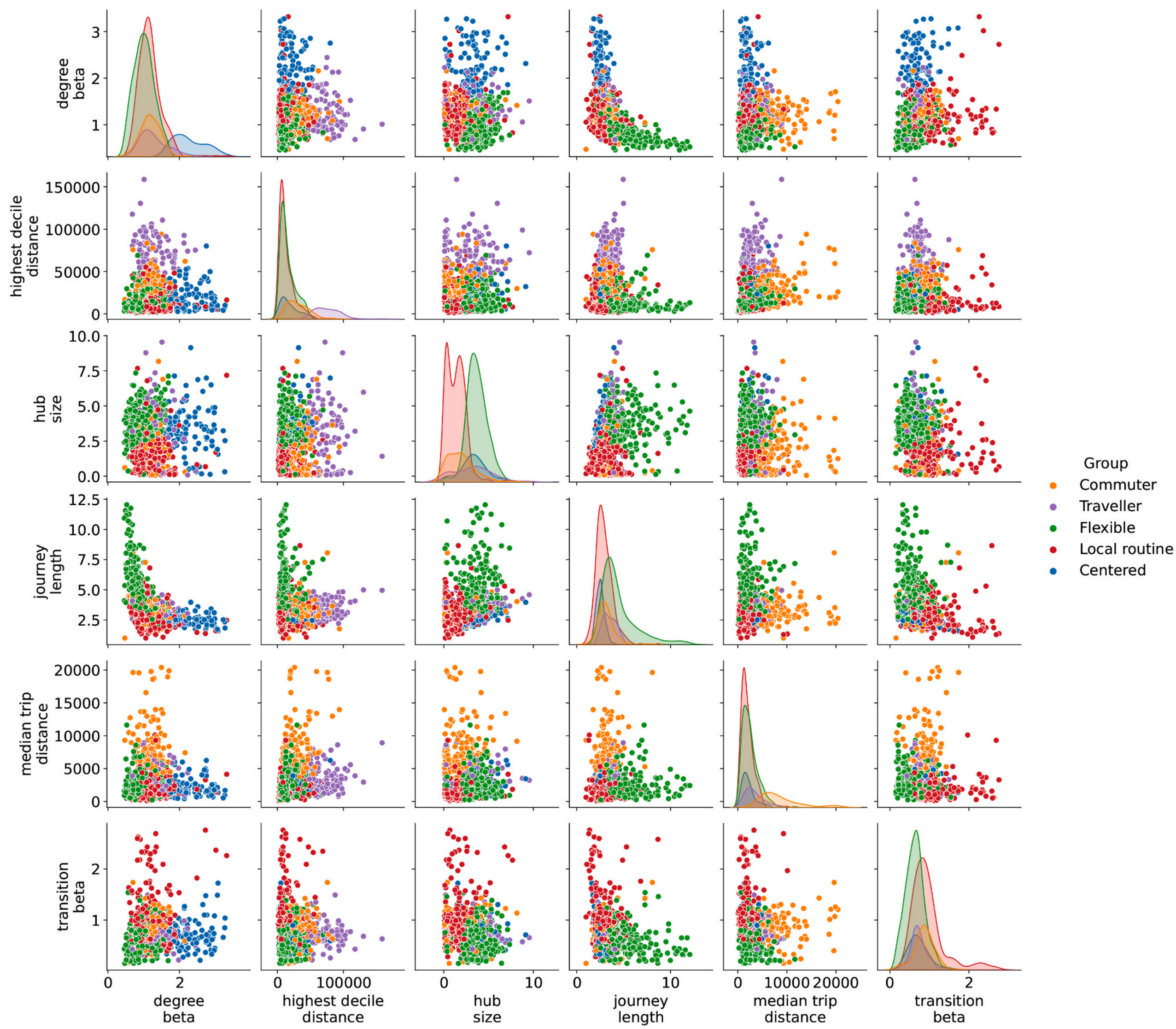


Fig. 11. Scatterplot matrix for the features of all datasets. Outliers were removed beforehand. The user groups can be clearly distinguished on certain features axes.

Appendix B. Cross-sectional study

The cross-sectional study on the yumuv dataset in Section 5.2.1 can be conducted in similar form for the Green Class dataset. Since no control group is available in the Green Class studies, we instead compare their user group distribution to the one in all other studies. Fig. 12 shows the group prevalence's. In Green Class 1 and 2, there is a significantly higher share of the *Commuter* and *Traveller* groups compared to the other studies, i.e., these two groups are above average attracted to the Green Class offer. In contrast, less users are part of the *Local routine* and *Flexible* groups. The differences are slightly weaker for Green Class 2. In both cases, these differences in the distribution of the user groups are significant (χ^2 test, $p < 0.01$). These target and non-target groups can be characterized using Fig. 5 and it could be further analyzed with respect to additional information such as demographics (cf. Section 5.2.3).

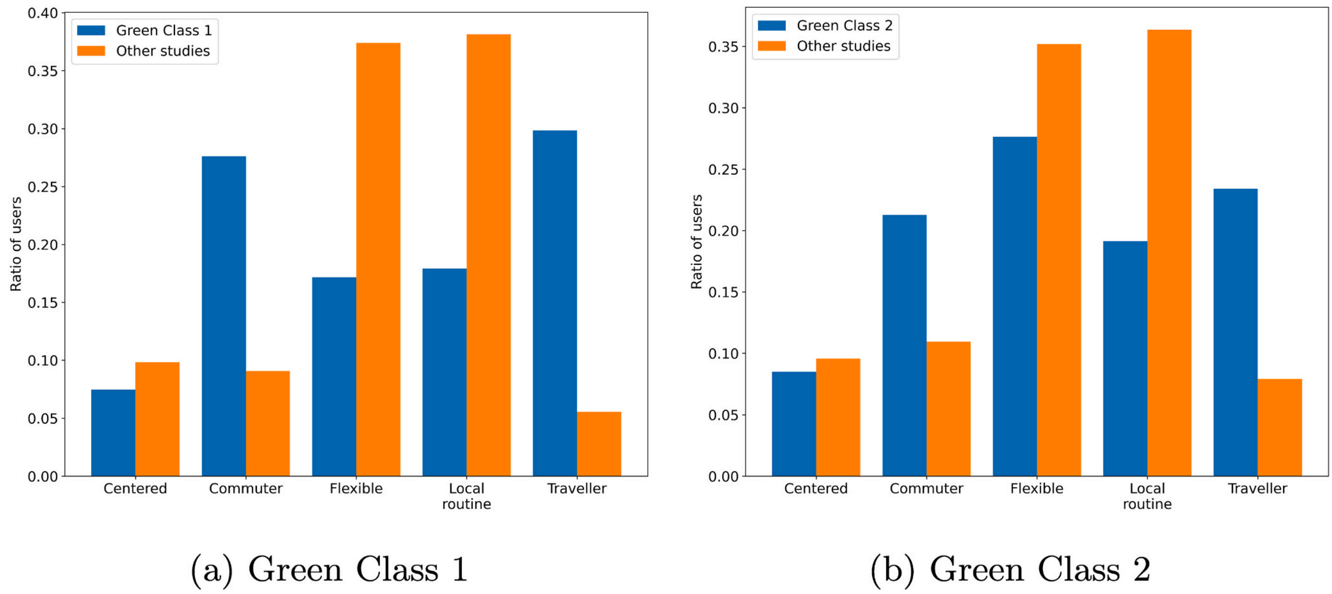


Fig. 12. Cross-sectional study for Green Class 1 and 2: There are more Travellers and Commuters taking part in the Green Class studies compared to the proportion in other datasets.

Appendix C. Longitudinal study

Analogous to Figs. 8, 13 visualizes the movements of users between groups in a network. The width of the edges is proportional to the number of users that are assigned to group A before, but switch to group B during the trial period.

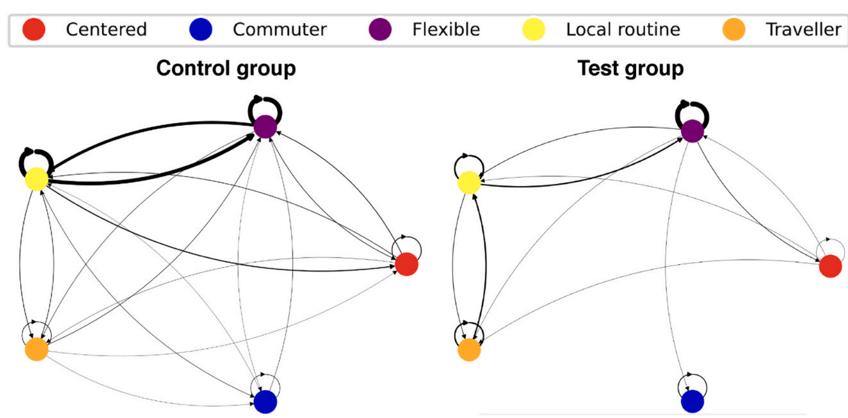


Fig. 13. Changes of user groups from the period before access to the yumuv offer to the period after intervention. The arrow width corresponds to the number of users that change from one group to another.

Appendix D. Dependence on clustering hyperparameters

Although our clustering approach does not depend on a major design choice such as the number of clusters as input, it still uses several hyperparameters. Those either provide more flexibility than before (e.g. the choices of k) or have minor influence on the result, such as the threshold θ_{minf}. The latter is demonstrated with an additional experiment reported in Table 3. The initial partitions (cf. Fig. 4a) with 3 runs for each of $k = [6,7,8,9]$ yield 90 clusters in total. The cluster are merged (cf. Fig. 4e) if no features are significant in contradicting directions and if there are more than θ_{minf} significant features corresponding. Increasing θ_{minf} leads to more unmerged user groups since more and more clusters have an insufficient number of significant features to be merged. Similarly, users can only be assigned to groups with a sufficient number of significant features. Therefore the number of unassigned users increases (see Table 3). However, the resulting user groups that are consistently assigned (see section 3.3.3) are clearly stable and robust to θ_{minf}.

Table 3
Effect of the θ_{minf} threshold, defining the minimum number of significant features to merge user groups. Only when θ_{minf} is set to a high value, the merging process is affected, i.e. clusters cannot be merged due to a small number of significant features.

θ_{minf}	Clusters	User groups (merged)	User groups (assigned consistently)	Unassigned users (%)
1	90	6	5	0.00
2	90	6	5	0.00

(continued on next page)

Table 3 (*continued*)

θ_{minf}	Clusters	User groups (merged)	User groups (assigned consistently)	Unassigned users (%)
3	90	8	5	0.00
4	90	18	6	0.03
5	90	35	7	0.14
6	90	80	2	0.82

As a rule of thumb, a user of the framework should set $\theta_{minf} \leq \frac{m}{2}$, i.e. not more than half of the number of features, and ensure that all users can be assigned to a group.

Appendix E. Mobility profiling based on classical mobility features

To demonstrate the generality of our clustering algorithm, we additionally show its results when applied on a set of classical mobility features, i.e. the *basic features* described in section 6.2. The same parameters are used. Note that we can not include the two Foursquare datasets, as the calculation of some of the features requires trajectory data. In the group finding phase, 10 groups are identified, but the users are only assigned consistently to 7 of them. The average consistency score (c.f. 6.4) is 0.9, meaning that on average a user is assigned to its most dominant group in 90% cases. This gives evidence that our algorithm is in general suitable to derive stable user groups. Fig. 14 shows the user groups based on classical features analogously to Fig. 5 (we omit the step of naming the user groups as it is not the focus of this work). Fig. 15 depicts the groups per dataset and shows strong differences between the studies, indicating a strong influence of tracking period and other technical dataset properties on the user groups (c.f. appendix F for further analysis). We argue that graph features are thus more robust to technical dataset properties and are therefore suitable for comparing mobility behavior of users in different studies.

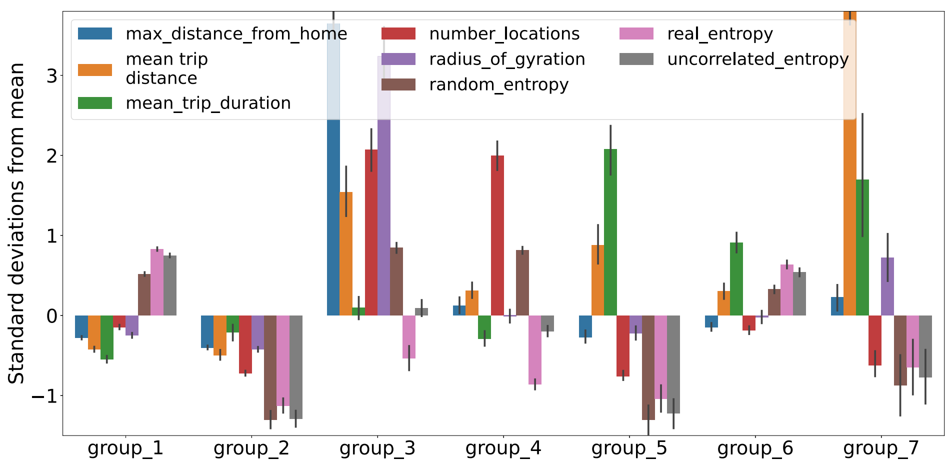


Fig. 14. Characteristics of the identified groups based on classical mobility features.

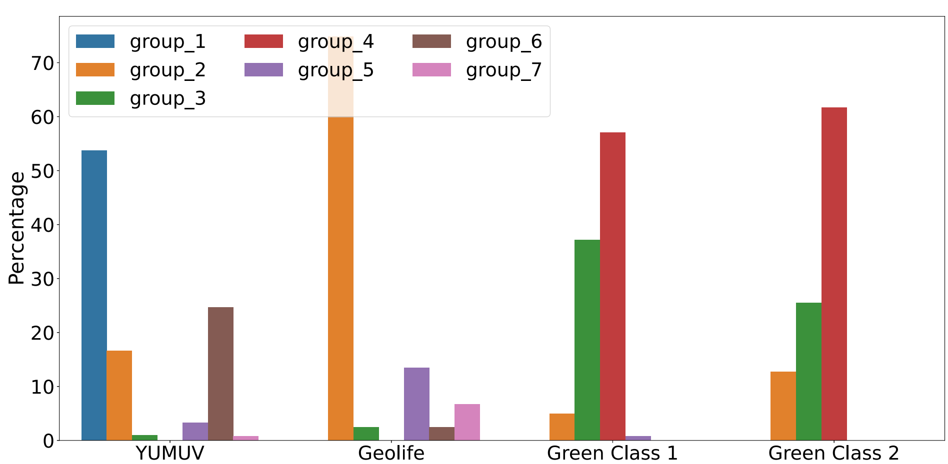


Fig. 15. Distribution of user groups (found based on classical mobility features) over datasets. Clearly, the groups are highly dependent on dataset properties.

Appendix F. Dependency of mobility profiles on technical dataset properties

In section 5.1.2 the distribution of user groups over studies was discussed, and the occurrence of all groups in most datasets indicated a certain generality of the mobility profiles. Here, we provide further evidence that the user groups are robust to technical properties of the data. A multinomial logistic regression model is used to quantify the dependence on data properties, namely tracking duration and tracking coverage. We compare the resulting coefficients and p-values for the user groups derived from graph features (c.f. section 5.1), given in Table 4, to the ones for the basic features (c.f. appendix E), listed in Table 5. The logistic regression model for graph-based user groups is hardly better than random, with an accuracy of 0.375 (random is 0.35 for the 5 imbalanced classes) and an R-Squared value of 0.035. The coefficients in Table 4 are also lower than for the basic features and mostly non-significant, with the exception of the Commuter group. Note that the coefficients of the first group (Centered) can not be computed since it serves as the reference group in the model.

Table 4
Dependence of graph-based mobility profiles on tracking period and coverage. The coefficients with p-values in parentheses of a multinomial logistic regression model are shown. Significant coefficients are marked bold.

	Commuter	Flexible	Local routine	Traveller
Days tracked	**5.48 (0.0)**	0.13 (0.93)	−0.69 (0.62)	**6.95 (0.0)**
Tracking coverage	**−13.57 (0.02)**	4.89 (0.37)	9.82 (0.07)	−5.94 (0.39)
INTERCEPT	**12.54 (0.02)**	−3.78 (0.47)	−8.32 (0.11)	4.55 (0.49)

In contrast, the basic-features lead to user groups that are strongly influenced by the number of tracked days and the coverage, as shown by the large and significant coefficients in Table 5. The accuracy is 0.59 (random: 0.47) and R-Squared is 0.279. In summary, while undesired dependencies on dataset-specific properties may still exist, the experiment shows the advantages of our approach with a comparably high robustness of the proposed graph-based feature set for mobility profiling.

References

Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., & Baronchelli, A. (July 2018). Evidence for a conserved quantity in human mobility. *Nature Human Behaviour, 2*(7), 485. https://doi.org/10.1038/s41562-018-0364-x. ISSN 2397-3374.

Axhausen, K. W. (2007). Definition of movement and activity for transport modelling. In *Handbook of transport modelling*. Emerald Group Publishing Limited.

Axhausen, K. W., & Gärling, T. (October 1992). Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems. *Transport Reviews, 12*(4), 323–341. https://doi.org/10.1080/01441649208716826. ISSN 0144-1647.

Ben-Gal, I., Weinstock, S., Singer, G., & Bambos, N. (2019). Clustering users by their mobility behavioral patterns. *ACM Transactions on Knowledge Discovery from Data (TKDD), 13*(4), 1–28.

Bhattacharya, A., & Das, S. K. (March 2002). LeZi-update: An information-theoretic framework for personal mobility tracking in PCS networks. *Wireless Networks, 8*(2), 121–135. https://doi.org/10.1023/A:1013759724438. ISSN 1572-8196.

Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature, 439*(7075), 462–465.

Calafiore, A., Palmer, G., Comber, S., Arribas-Bel, D., & Singleton, A. (2021). A geographic data science framework for the functional and contextual analysis of human dynamics within global cities. *Computers, Environment and Urban Systems, 85*, Article 101539.

Castiglione, J., Bradley, M., & Gliebe, J. (2015). *Activity-based travel demand models: A primer. SHRP 2 Report, (S2-C46-RR-1)*.

Chapman, L. (2007). Transport and climate change: A review. *Journal of Transport Geography, 15*(5), 354–367.

Chen, C., Ma, J., Susilo, Y., Yu, L., & Wang, M. (July 2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies, 68*, 285–299. https://doi.org/10.1016/j.trc.2016.04.005. ISSN 0968-090X.

Creutzig, F., Jochem, P., Edelenbosch, O. Y., Mattauch, L., van Vuuren, D. P., McCollum, D., & Minx, J. (2015). Transport: A roadblock to climate change mitigation? *Science, 350*(6263), 911–912.

El Mahrsi, M. K., Côme, E., Oukhellou, L., & Verleysen, M. (2016). Clustering smart card data for urban mobility analysis. *IEEE Transactions on Intelligent Transportation Systems, 18*(3), 712–728.

Evangeliou, N., Grythe, H., Klimont, Z., Heyes, C., Eckhardt, S., Lopez-Aparicio, S., & Stohl, A. (2020). Atmospheric transport is a major pathway of microplastics to remote regions. *Nature Communications, 11*(1), 1–11.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782.

Gössling, S. (July 2020). Why cities need to take road space from cars - and how this could be done. *Journal of Urban Design, 25*(4), 443–448. https://doi.org/10.1080/13574809.2020.1727318. ISSN 1357-4809.

Hensher, D. A., Ho, C. Q., & Reck, D. J. (2021). Mobility as a service and private car use: Evidence from the Sydney maas trial. *Transportation Research Part A: Policy and Practice, 145*, 17–33.

Hong, Y., Xin, Y., Martin, H., Bucher, D., & Raubal, M. (2021). A clustering-based framework for individual travel behaviour change detection. In *11th International Conference on Geographic Information Science (GIScience 2021)-Part II*.

Hornsby, K., & Egenhofer, M. J. (September 2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence, 36*(1–2), 177–194. https://doi.org/10.1023/A:1015812206586. ISSN 1012-2443.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193–218.

Jiang, S., Ferreira, J., & Gonzalez, M. C. (2017). Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data, 3*(2), 208–219.

Jonietz, D., Bucher, D., Martin, H., & Raubal, M. (2018). Identifying and interpreting clusters of persons with similar mobility behaviour change processes. In *The annual international conference on geographic information science* (pp. 291–307). Springer.

Kreil, D. P., Kopp, M. K., Jonietz, D., Neun, M., Gruca, A., Herruzo, P., … Hochreiter, S. (2020). The surprising efficiency of framing geo-spatial time series forecasting as a video prediction task–insights from the iarai traffic4cast competition at neurips 2019. In *NeurIPS 2019 competition and demonstration track* (pp. 232–241). PMLR.

Kumar, N., & Raubal, M. (2021). Applications of deep learning in congestion detection, prediction and alleviation: A survey. *Transportation Research Part C: Emerging Technologies, 133*, Article 103432.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Li, Q., Yu, Z., Xie, X., Chen, Y., Liu, W., & Ma, W.-Y. (2008). Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems* (pp. 1–10).

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory, 28*(2), 129–137.

Long, J. A., Weibel, R., Dodge, S., & Laube, P. (2018). *Moving ahead with computational movement analysis*.

Luca, M., Barlacchi, G., Lepri, B., & Pappalardo, L. (2023). A survey on deep learning for human mobility. *ACM Computing Surveys, 55*(1), 1–44. https://doi.org/10.1145/3485125, 7.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60.

Manousakas, D., Mascolo, C., Beresford, A. R., Chan, D., & Sharma, N. (2018). Quantifying privacy loss of human mobility graph topology. *Proceedings on Privacy Enhancing Technologies, 2018*(3), 5–21.

Martin, H., Bucher, D., Suel, E., Zhao, P., Perez-Cruz, F., & Raubal, M. (2018). Graph convolutional neural networks for human activity purpose imputation from GPS-based trajectory data. In *NeurIPS 2018 Spatiotemporal Workshop*.

Martin, H., Becker, H., Bucher, D., Jonietz, D., Raubal, M., & Axhausen, K. W. (2019). *Begleitstudie SBB Green Class-Abschlussbericht. 1439*. Arbeitsberichte Verkehrs-und Raumplanung.

Martin, H., Reck, D. J., Axhausen, K. W., & Raubal, M. (2021). *ETH mobility initiative project MI-01-19 empirical use and impact analysis of MaaS. Technical report*. ETH Zurich.

Martin, H., Hong, Y., Wiedemann, N., Bucher, D., & Raubal, M. *Trackintel: An open-source python library for human mobility analysis*. (2022). URL https://arxiv.org/abs/2206.03593.

McKenzie, G., & Romm, D. (2021). Measuring urban regional similarity through mobility signatures. *Computers, Environment and Urban Systems, 89*, Article 101684.

Onnela, J.-P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E, 71*(6), Article 065103.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The pagerank citation ranking: Bringing order to the web. Technical report*. Stanford InfoLab.

Pappalardo, L., Simini, F., Barlacchi, G., & Pellungrini, R. (2022). Scikit-mobility: A python library for the analysis, generation and risk assessment of mobility data. *Journal of Statistical Software, 103*(4), 1–38. https://doi.org/10.18637/jss.v103.i04

Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., & Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature Communications, 6*(1), 1–8.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association, 66*(336), 846–850.

Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning. B, Planning & Design, 33*(5), 727–748.

Reed, T. (February 2019). INRIX global traffic scorecard. URL https://trid.trb.org/view/1456836.

Rinzivillo, S., Gabrielli, L., Nanni, M., Pappalardo, L., Pedreschi, D., & Giannotti, F. (2014). The purpose of motion: Learning activities from individual mobility

networks. In *International conference on data science and advanced analytics (DSAA)* (pp. 312–318). IEEE.
Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of the Royal Society Interface, 10* (84), 20130246.
Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (vol. 39). Cambridge: Cambridge University Press.
Song, C., Zehui, Q., Blumm, N., & Barabási, A.-L. (2010). Limits of predictability in human mobility. *Science, 327*(5968), 1018–1021.
Sulis, P., & Manley, E. (2018). Exploring similarities and variations of human mobility patterns in the city of london. In *, vol. 42. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives* (pp. 51–58). ISPRS.
Toch, E., Lerner, B., Ben-Zion, E., & Ben-Gal, I. (2019). Analyzing large-scale human mobility data: A survey of machine learning methods and applications. *Knowledge and Information Systems, 58*(3), 501–523.
Vassilvitskii, S., & Arthur, D. (2006). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027–1035).
Wang, R., Li, N., & Wang, Y. (2021). Does the returners and explorers dichotomy in urban human mobility depend on the observation duration? An empirical study in Guangzhou, China. *Sustainable Cities and Society, 69*, Article 102862.
Xin, Y., & MacEachren, A. M. (2020). Characterizing traveling fans: A workflow for event-oriented travel pattern analysis using twitter data. *International Journal of Geographical Information Science, 34*(12), 2497–2516.
Yan, X.-Y., Wang, W.-X., Gao, Z.-Y., & Lai, Y.-C. (November 2017). Universal model of individual and population mobility on diverse spatial scales. *Nature Communications, 8*(1), 1639. https://doi.org/10.1038/s41467-017-01892-8. ISSN 2041-1723.
Yang, D., Zhang, D., Chen, L., & Bingqing, Q. (2015). Nationtelescope: Monitoring and visualizing large-scale collective behavior in lbsns. *Journal of Network and Computer Applications, 55*, 170–180.
Yang, D., Zhang, D., & Bingqing, Q. (2016). Participatory cultural mapping based on collective behavior data in location-based social networks. *ACM Transactions on Intelligent Systems and Technology (TIST), 7*(3), 1–23.
Yu, Z., Li, Q., Chen, Y., Xie, X., & Ma, W.-Y. (2008). Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on ubiquitous computing, UbiComp '08* (pp. 312–321). New York, NY, USA: ACM. https://doi.org/10.1145/1409635.1409677. ISBN 978-1-60558-136-1.
Yu, Z., Zhang, L., Xie, X., & Ma, W.-Y. (2009). Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web* (pp. 791–800).
Yuan, Y., & Raubal, M. (2012). Extracting dynamic urban mobility patterns from mobile phone data. In *International conference on geographic information science* (pp. 354–367). Springer.