Voting with one's feet
Unraveling urban centers attraction using visiting frequency

Cats, Oded; Ferranti, Francesco

DOI
10.1016/j.cities.2022.103773

Publication date
2022

Document Version
Final published version

Published in
Cities

Citation (APA)
Cats, O., & Ferranti, F. (2022). Voting with one's feet: Unraveling urban centers attraction using visiting frequency. *Cities*, *127*, Article 103773. https://doi.org/10.1016/j.cities.2022.103773

Cities 127 (2022) 103773

Contents lists available at ScienceDirect

Cities

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

Voting with one's feet: Unraveling urban centers attraction using visiting frequency

Oded Cats [a,b,*], Francesco Ferranti [b]

[a] Department of Transport and Planning, Delft University of Technology, Delft, the Netherlands
[b] Division of Transport Planning, KTH Royal Institute of Technology, Stockholm, Sweden

ARTICLE INFO

Keywords:
Activity centers
Urban structure
Public transport
Clustering
Smart card data

ABSTRACT

Urban and regional areas worldwide exhibit a complex and uneven distribution of activities with certain areas attracting more people during different time periods. In this study we systemically classify different parts of the urban area which are most attractive as measured by their ability to attract visitors. A weekly visiting profile is constructed for each travel demand zone and thereafter clustered to identify areas with common attraction patterns. We leverage on the availability of longitudinal individual mobility traces in the form of smart card data transactions. We apply our method to the case study of the multi-modal public trans-port system of the Stockholm urban agglomeration area. The results of our clustering based on the weekly visiting profiles reveal four distinctive types of visiting attraction based on the intensity and temporal distribution of activities performed. The results of this study can be used to inform planners and decision makers about the main activity locations of travellers and how their temporal patterns vary across the metropolitan area and the design of related policies.

1. Introduction

Urban and regional areas worldwide exhibit a complex distribution of activities. The seemingly infinite number of combinations of geographical features, transport infrastructure and the resulting accessibility, land-use development policies and the location of key amenities result with each area having its unique spatial organisation. At the same time, common patterns in urban economics, travel behavior and transport network structure result with some similarities in how human activities are geographically distributed (Ahmed & Stopher, 2014; Schläpfer et al., 2021). In this study we aim to systemically identify and classify different parts of the urban area which are most attractive as measured by their ability to attract visitors.

In the last decade, a plethora of studies have analysed human activity patterns using data that contains geo-location traces. This includes the analysis of mobile phone data, social media posts and travel-related records (e.g. Jacobs-Crisioni et al. (2014); Cats et al. (2015); Lee et al. (2018); Perlman and Roy (2021); Ye et al. (2021); Xie et al. (2021)). This stream of research, sometimes referred to as 'Social Sensing' (Liu et al., 2015) and the analysis thereof involves the application of methods ranging from natural language processing and temporal signature analysis to the inference of travel semantics (Liu et al., 2021).

The increasing availability of human mobility traces enables a data-driven analysis of urban activities. In particular, the analysis of social sensing data enables to unravel the temporal dynamics of cities and identify how activity patterns vary throughout the urban area. Consequently, it enables measuring urban structure based on people flows rather than static geographical features. Social media data has been used to extract global statistical properties of urban travel and activity patterns (Jurdak et al., 2015; Wu et al., 2014) as well as for characterising the temporal features of activity patterns at different urban resolution levels (Shen & Karimi, 2016; Wu et al., 2018). Similarly, mobile phone data can be used to establish general properties of activity patterns (Louail et al., 2015; Widhalm et al., 2015) and the analysis of how those vary spatio-temporally (e.g. Lee et al. (2018)), possibly also enhanced by combining it with social media data (Tu et al., 2017).

Since multi-modal public transport networks are playing a critical role in urban mobility, the analysis of passively collected smart card data enables the analysis of key travel flows. Roth et al. (2011) and Ozus et al. (2012) were pioneering in identifying activity areas using public transport mobility data. Luo et al. (2017) developed a method for determining travel demand zones based on both flow and spatial distance information using smart card data. Past studies using smart card data with the goal of analyzing urban activity patterns have mostly

* Corresponding author at: Stevinweg 1, 2628 CN Delft, the Netherlands.
E-mail address: o.cats@tudelft.nl (O. Cats).

https://doi.org/10.1016/j.cities.2022.103773
Received 29 July 2021; Received in revised form 4 May 2022; Accepted 20 May 2022

focused on classifying individual stations (as opposed to geographical areas) based on the land-use patterns and places of interest in proximity to these stations or characteristics of their respective locations (Gan et al., 2020; Kim, 2019; Zhang et al., 2018; Zhang et al., 2021; Zhao et al., 2019). For example, Zhang et al. (2019) analysed the temporal properties of start and end times of activities at metro stations, albeit without examining the resulting urban structure. Gan et al. (2020) clustered stations based on the hourly ridership profiles and examine the relations between the results of the k-means cluster analysis and local land-use characteristics. The majority of these studies clustered stations based on the temporal properties of passenger flow distribution. Conversely, Wang et al. (2021) identified urban functional areas in Beijing by employing a Gaussian mixture model on smart card data. Majority of these studies have utilised (only) metro data, hampering the analysis for areas which might be under-served by the metro. Moreover, while these studies have either identified groups of stations that exhibit similar patterns or identified homogeneous areas and have discussed their observed characteristics, none of those has systematically clustered identified areas in terms of their travel patterns. Such an analysis can offer insights on the urban structure which extend beyond the identification of its sub-components into the quantification of their inter-relations and their different functions. The results of such an analysis can support the assessment of urban vibrancy and allow identifying recurrent patterns across the urban region.

Several past studies went beyond identifying urban centers into systematically characterising those. Cats et al. (2015) proposed a two-step approach for identifying and characterising sub-centers. First, sub-centers are identified using a hierarchical method. Subsequently, these centers were clustered by using an agglomerative hierarchical method based on the temporal profile of the boarding and alighting passenger volumes. They applied this approach for identifying and characterising activity centers in Stockholm and analysed the extent to which it has evolved into a polycentric structure. Zhang et al. (2021) applied a community detection technique to identify urban areas based on smart card data and then characterised those in terms of how the share of intra-flow and degree of compactness of passenger movements has evolved using data from 2013 to 2017 for London. The intra-flow was defined based on the origin and destination of each trip. Notwithstanding, none of the above-mentioned studies has investigated the extent to which different parts of the region attract travellers from other areas. We address this gap by proposing a method to characterise and cluster zones in terms of their temporal visiting profile while distinguishing between residents and visitors. The zones used are constructed based on observed travel patterns while ensuring that they constitute geographically contingent areas.

The objective of this study is to identify and classify urban areas based on how attractive they are as activity centers. We do so by constructing a weekly visiting profile for each travel demand zone within our study area. We leverage on the availability of longitudinal (a year-long) individual mobility data in the form of smart card data transactions that allow us to infer the home-zone of each card-holder. In our analysis, we cluster zones in relation to their weekly visiting profile so as to allow identifying areas that attract visitors for different activity types (e.g. morning peak commuters vs. recreational midday weekend trips). In the following, we describe the proposed approach along with the sequence of analysis steps (Section 2), apply it to the case of the multi-modal public transport system in Stockholm, Sweden (Section 3), discuss the obtained results (Section 4) and discuss the key findings and potential applications (Section 5).

2. Method

2.1. Analysis framework

We analyse zonal visiting patterns using disaggregate human mobility data. Our analysis requires individual mobility records for each traveller $i \in N$, where N is the set of all travellers considered in the analysis performed. Any user-based journey dataset can be used as input for this method, from travel diary surveys to travel app tracking data. In the following we will refer to the analysis of public transport travel demand data which is used in our application. The overall workflow of the method proposed in this study is presented in Fig. 1. As can be seen, three datasets need to be made available as input to this analysis, i.e. stops, travel transactions and vehicle movements. The stops dataset contains a list of all possible start and end locations of public transport trips and their corresponding geo-locations. The travel transactions dataset contains information on all travel records with their corresponding geo-locations, time stamps and a unique identifier of the card holder (traveller). Finally, A dataset of vehicle movements contains information on the arrival and departure times for each vehicle at each of the stops along the respective service line.

The method consists of six processes or modules as shown in Fig. 1: (i) Trip Destination Inference, (ii) Transfer Inference, (iii) Zone Generation, (iv) Home-Zone Inference, (v) Zonal Visiting Profile and (vi) Zonal Attraction Clusters. The first three are required in order to obtain the intermediate databases of Journeys and Zones, both of which are given as input in order to perform the latter three modules. The following subsections detail each of the modules.

2.2. Trip destination inference

Depending on the public transport system considered, destinations may or may not be empirically known for all trips. Many public transport fare validation schemes worldwide are either fully or partially based on tap-in only. Consequently, trip destinations are not directly observable but rather need to be inferred. There are by now well-established techniques for inferring tap-out locations based on tap-in only data (Munizaga & Palma, 2012; Trépanier et al., 2007). The key principle is that for each recorded tap-in location, we search for the most likely tap-out location based on the tap-in location of the subsequent trip conducted by the same card-holder. More specifically, we search for the closest downstream stop along the same line as the one boarded by the passenger (in the case of an on-board validation) or along one of the lines that could have been boarded by the passengers (in the case of a gate-based validation). Feasibility constraints related to service availability based on vehicle movement records, walking distances between stops and the respective time window between two subsequent tap-in transactions are enforced (Kholodov et al., 2021). The corresponding time-stamp for the inferred tap-out location is then obtained from the vehicle movement data. This module results in populating for each tap-in record the inferred tap-out location and time stamp, in case those are not directly available.

2.3. Transfer inference

For the analysis of zonal attraction, information on travellers' journey destination is needed. It is therefore necessary to ensure that journeys' destinations rather than intermediate trip destinations such as interchange locations are considered. Individual trips may be bundled into passenger journeys by means of performing a transfer inference procedure. We identify transfers by applying an upper time threshold for the maximum allowable transfer time. We follow the approach proposed in Seaborn et al. (2009) in setting threshold levels that are based on the analysis of the cumulative distribution function of the time gap between inferred tap-out time and the next recorded tap-in time. This was performed and specified separately for each mode transfer combination (e.g. metro-train, metro-bus, bus-tram, bus-bus etc.). The outcome of this module is a database containing a detailed travel diary for each card-holder throughout the analysis period. Each entry in the database corresponds to a journey performed by a certain card-holder and comprised of a set of trips with their tap-in and tap-out locations and time stamps.

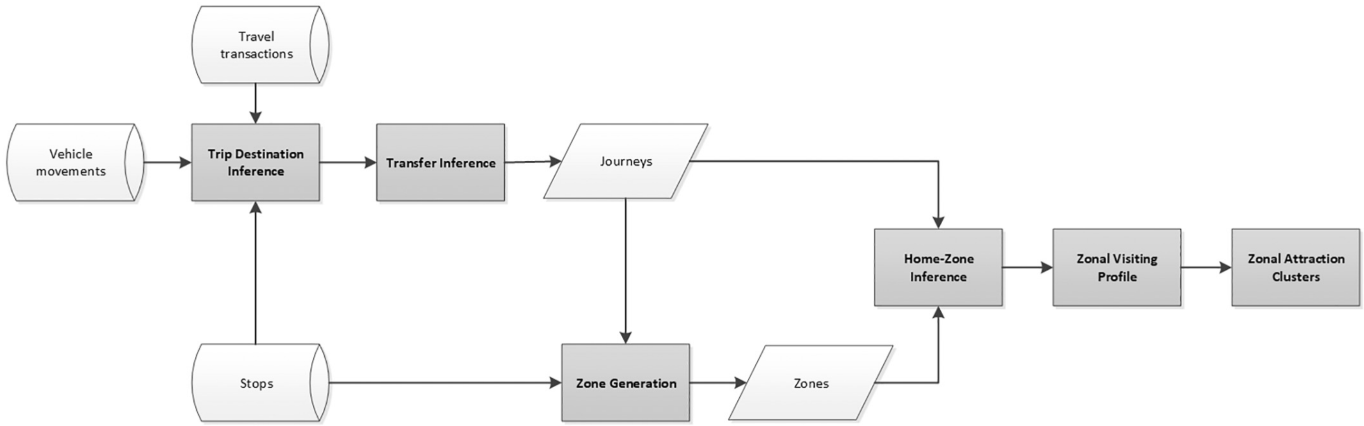


Fig. 1. Overview of the workflow with elliptic tubes, rectangles and parallelogram representing input datasets, modules and intermediate databases, respectively.

2.4. Zone generation

Another ingredient needed for this analysis is a partitioning of the analysis area into zones by means of aggregating stops, see Fig. 1. The set of zones Z and the geographical demarcation of each $z \in Z$ can be based on a pre-determined definition available from the official central bureau of statistics (e.g. postcodes), by overlying a grid and defining equal size grid cells or by applying a data-driven zonal generation method. The latter approach is adopted in this study.

We adopt the data-driven k-means clustering method detailed in Luo et al. (2017) for partitioning all the stops $s \in S$, where S is the set containing all public transport stops that are situated within the analysis area. The partitioning is performed such that S is divided into subsets that constitute zones that form geographically contiguous areas and exhibit coherent demand patterns. Stops and Journeys data are given as input in order to identify clusters. Clusters are generated so that differences in terms of journey destinations for journeys originating from each zone are minimized within zones and these differences are maximized between zones. Hence, it generates geographically compact zones which exhibit similar origin/destination travel demand patterns. The output of this module is a list of mutually exclusive and collectively exhaustive zones and the corresponding sets of stops per zone, i.e. $S_z \in z$.

2.5. Home-zone inference

Our analysis of zones attraction is based on the consideration of visits by non-residents. It is therefore essential to infer travellers' place of residence, or so-called home-zones. This allows us to then label each journey based on its destination and the traveller undertaking it in terms of whether it is performed by a resident or a non-resident of the destination zone. The journey origins and destinations indicated in the individual mobility data (e.g. in the form of coordinates or public transport stops) are then mapped into the zoning obtained in the previous module. We denote traveller $i \in Z$ if $z \in Z$ is the home-zone of traveller $i \in N$. In this analysis, following Kholodov et al. (2021) we conduct an adaption of Sari Aslam et al. (2019). We identify the most likely home zone per traveller based on the frequency of each zone serving as an origin for journeys performed by each card holder. We count the number of first journeys of the day (starting after 5 am) that originate from each of the zones in Z for each card-holder. The most common origin zone for the first trip of the day is then considered to be the home zone.

2.6. Zonal visiting profile

Next, we utilise the network-wide mobility data with information on journey data records per traveller and their respective home zones to calculate for each zone an attraction metric for each time period. The latter is hereby measured by means of the share of non-residents visiting each zone. To this end, we propose constructing a weekly visiting pattern profile for each zone under consideration. More specifically, we are interested in the share of non-residents among all travellers for each time window (e.g. a 30 min time interval) throughout the week.

We define two flow metrics that are used for constructing our zonal attraction indicator. The first one, non-residential flow, $flow^{NR}(z,t)$, sums all the incoming and outgoing flow from each zone $z \in Z$ which is generated by travellers who do not reside in z during a given time window t.

$$flow^{NR}(z,t) = \sum_{d \in Z \setminus z, i \notin Z} f_{z,d}(i,t) + \sum_{o \in Z \setminus z, i \notin z} f_{o,z}(i,t)$$

where $f_{o,\ d}(i,t)$ denotes the number of journeys performed by traveller $i \in N$ departing from zone $o \in Z$ and destined to zone $d \in Z$ within time window t.

The second flow metric, total flow, corresponds to the total flow to and from each zone during the respective time window:

$$flow^{tot}(z,t) = \sum_{d \in N, i \in N} f_{od}(i,t) + \sum_{o \in Z, i \in N} f_{od}(i,t)$$

For both non-resident and total flows, we consider the weekly average aggregated values for each time window interval. For each zone, we obtain two arrays of the size: number of days of the week × time intervals of a day.

Next, we define *Weekly zonal attraction* vectors as entry-wise ratio of the non-resident flow over the total flow for each zone z for all time intervals t. Formally, the attraction of zone z at time interval t is:

$$A(z,t) = \frac{flow^{NR}(z,t)}{flow^{tot}(z,t)}$$

2.7. Zonal attraction clusters

Finally, we cluster the zones according to their visiting profiles, as measured in the previous module by the share of non-residents visiting each zone for each time window throughout the week. The vectors $A(z)$ are clustered using the K means algorithm. We then obtain clusters of zones that exhibit similar profiles in terms of the share of non-residents visiting them at different times-of-the-day and days-of-the-week.

3. Application

We study the attraction of different parts of Stockholm urban agglomeration area, Sweden, by applying the proposed method. Stockholm is of course the principal and main city in the Stockholm urban agglomeration area. In addition, it includes also the following 10 additional municipalities, all of which are situated in Stockholm County: Järfälla, Sollentuna, Danderyd, Sollentuna and Solna to its north, Nacka to its East, and Haninge, Botkyrka, Huddinge and Tyresö to its south. Fig. 2 provides an overview of the case study area along with references to all the key locations mentioned in the text. The total population residing within our case study area amounts to 1,75 million spread over an area of 381.63 km^2.

Public transport mobility data is used in this study for identifying zonal attractiveness. The average number of public transport trips per person per day in the case study area is 0.8. Moreover, 65% of all weekday trips are carried out by public transport. Consequently, public transport demand patterns are considered to be insightful for analyzing activity areas throughout the case study area.

The multi-modal public transport system consists of bus, tram, metro, commuter train and ferry services. The rail-bound network serves as the backbone of the regional network. The regional public transport network extends almost 10,000 km, 469 of which are rail-bound, serving daily more than 800,000 passengers through more than 5700 stops across the case study area. As can be observed in Fig. 3, the urban agglomeration area has a strong monocentric pattern which is also clearly manifested in the radial public transport network structure. Buses accompany the rail-bound network offering feeding and distributing services and enhance the connectivity by offering additional tangential connections as well as serving lower-density areas. All metro and commuter train lines connect to the inner-city and major interchange hubs are located within or at the edge of urban core. The analysis of Cats (2017) shows that until 1998 network developments focused on increasing its geographical coverage and has since then mostly focused on its densification of and around the network's core. This public transport network development is meant to support the development of sub-centers and their inter-connections (Cats et al., 2015) and thereby enabling to bypass and relief the passenger congestion in the city center and improve overall network robustness (Cats, 2016). The regional public transport authority, Region Stockholm, is responsible for long-term planning, procurement and quality control.

The public transport fare validation scheme in Stockholm is based on tap-in only. Three data sources are used in this study: (i) general transit feed specification (GTFS) data containing stop and timetables data; (ii) automatic vehicle location (AVL) data containing information on scheduled and realised vehicle movements, and; (iii) automatic fare collection (AFC) data with passengers' tap-in records. The latter consists of more than 400 million trips performed throughout the year of 2019, except for the month of July during which the data warehouse was

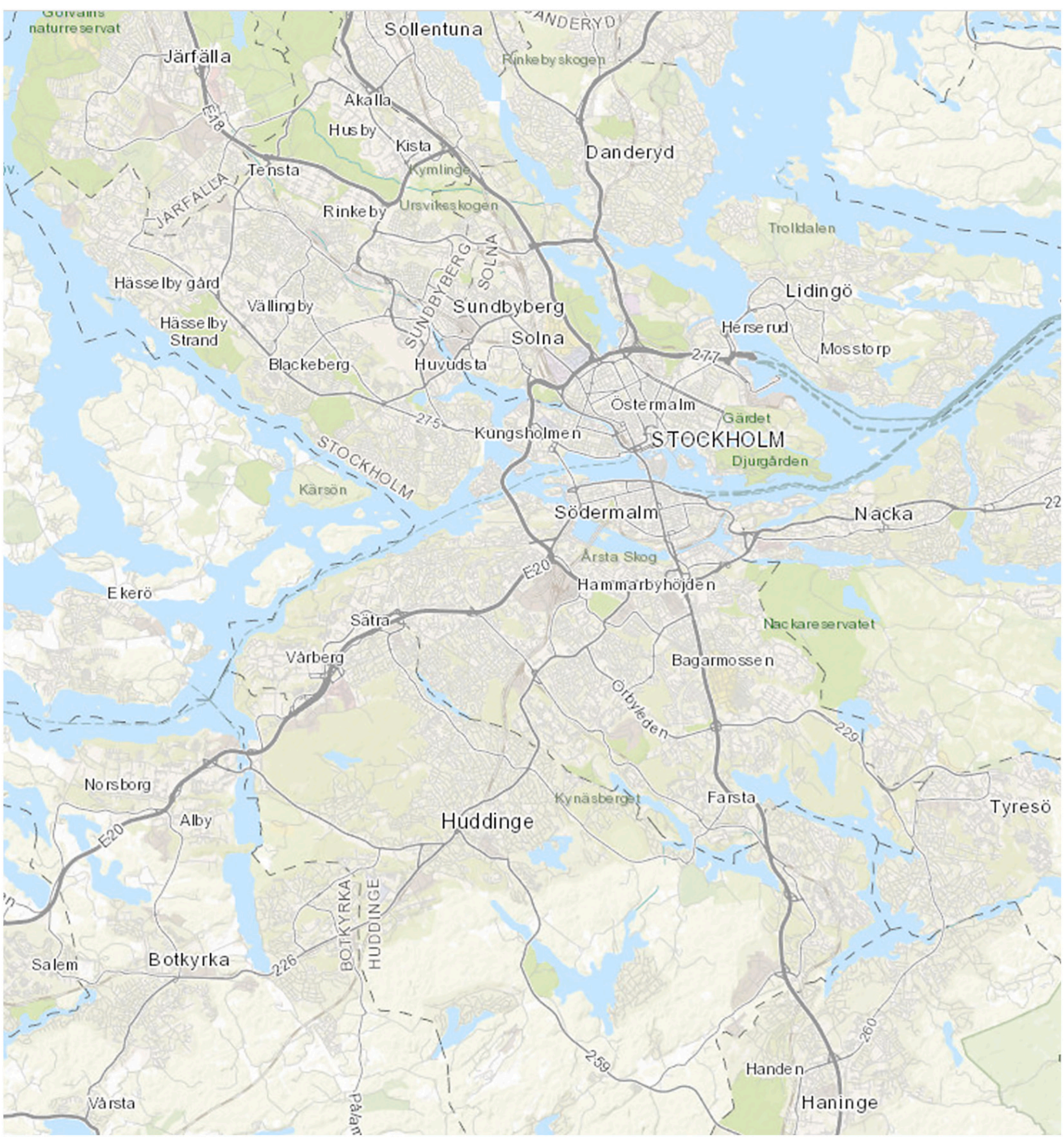


Fig. 2. Map of the Stockholm case study area. Source: Region Stockholm, http://www.rufs.se/kartor/.

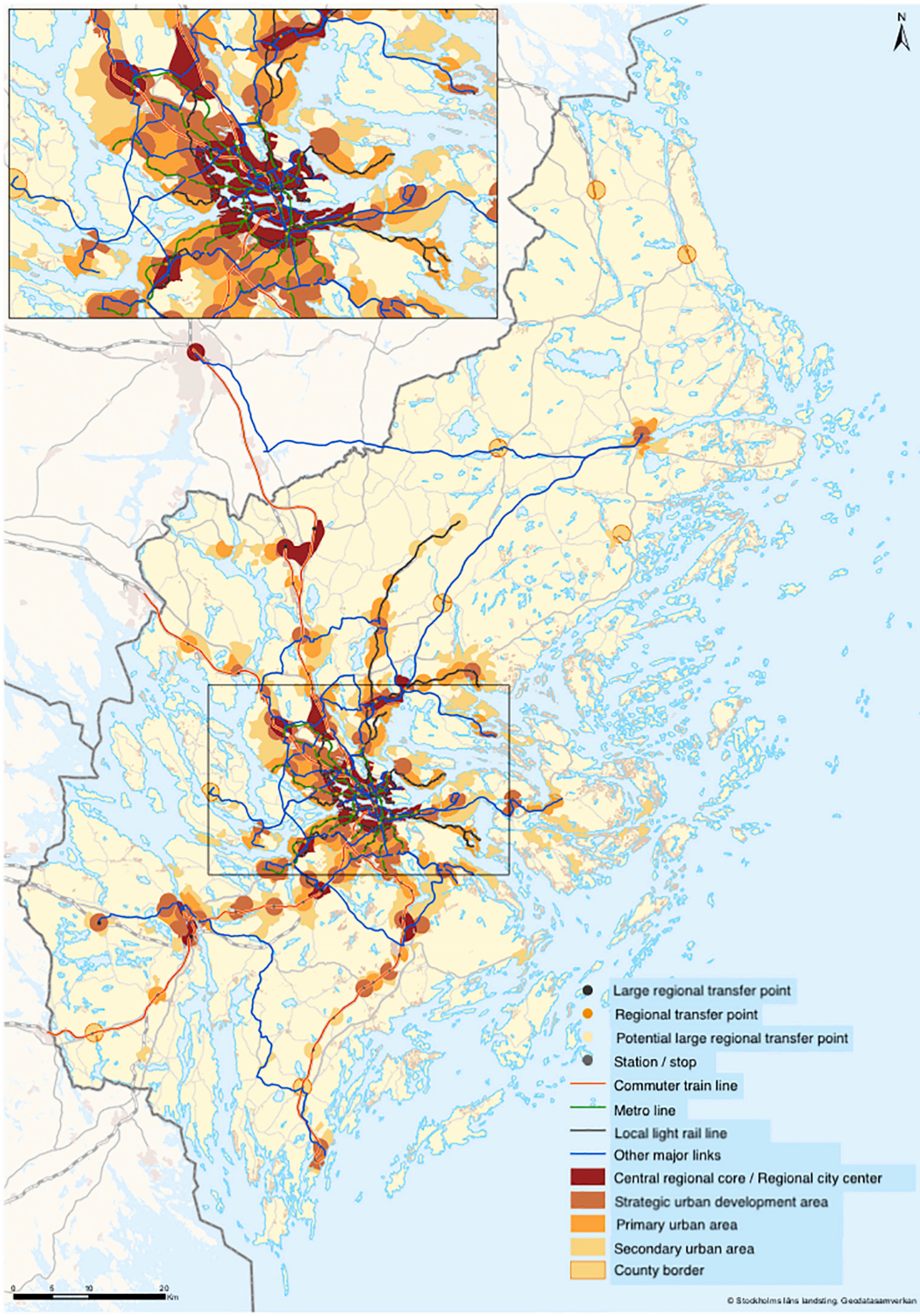


Fig. 3. Geographical overview of Stockholm area public transport network. Source: Region Stockholm, Regional utvecklingsplan för Stockholmsregionen: RUFS 2050. Stockholms läns landsting. http://www.rufs.se/globalassets/h.-publikationer/2017/rufs2050_utstallning_170630.pdf.

under maintenance works. The GTFS and AVL datasets were obtained for the same analysis period. Since public transport services have a limited service span in some parts of the case study area, we limit our analysis to the 6:30 am to 9:30 pm period and we use half-an-hour departure interval as the temporal unit of analysis.

We apply the sequence of inferences described in the previous section in order to compose passenger journeys. Using the *Trip Destination Inference* algorithm, we were able to estimate a tap-out location for 80% of our tap-in records, on par with results reported in other similar studies (e.g. Trépanier et al. (2007); Munizaga and Palma (2012)). We then apply the *Transfer Inference* algorithm and obtain a total of 341,457,079 journeys performed by 4,423,783 (anonymous) card-holders.

Next, we apply the data-driven *Zone Generation* technique. Applying this method to the Stockholm County resulted in clustering the more than 5700 stops which are located within the county into 137 zones, 59 of which are located within the case study area of Stockholm urban agglomeration. Fig. 4 presents the zones obtained.

We then turn into performing the *Home-Zone Inference* procedure. We are able to confidently infer a home zone for 70% of the cards, which account for 95% of all journeys. Finally, we compose the *Zonal Visiting Profile* based on the share of non-residents visiting each zone within each 30-minutes time period and thereafter perform a K means clustering

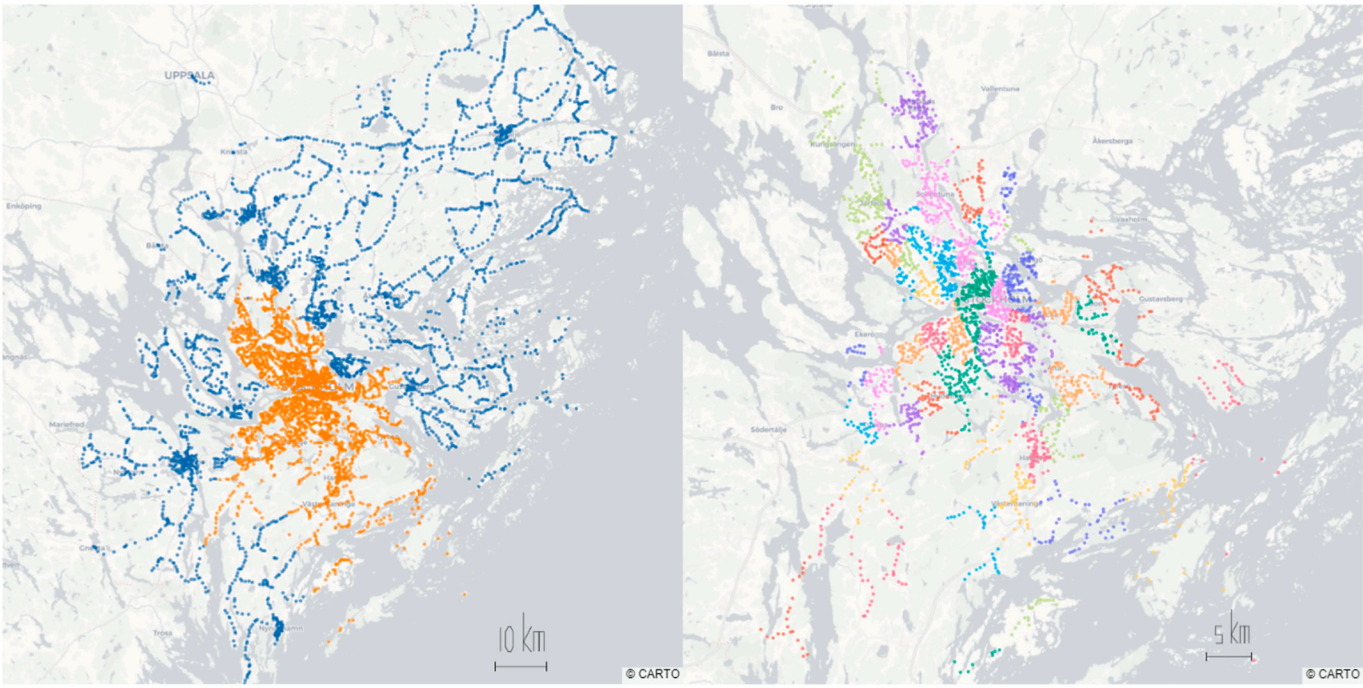


Fig. 4. All public transport stops in Stockholm County with those situated within the Stockholm urban agglomeration area highlighted in orange and those falling outside this demarcation shown in blue (left) and the zoning used in our analysis with stops grouped in different zones (clusters) shown using different colors (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

algorithm to obtain the *Zonal Attraction Clusters*. The results of which are reported in the next section.

4. Results and analysis

We start by reporting and discussing the results of the clustering *Zonal Attraction Clusters* step to thereafter turn to examining their underlying *Zonal Visiting Profile*.

In order to identify the most suitable partitioning of our zones into clusters, we test for a variety of a number of clusters, ranging from 3 to 15. We examine the silhouette scores shown in Fig. 5. We adopt the elbow rule in selecting to proceed with nine clusters.

Of the nine clusters we obtain with this set up, four are found to be representative of the majority of the zones, whereas the remaining five correspond to one zone each. The four main clusters are geographically depicted in Fig. 6. These four clusters are labeled based on their *Zonal Visiting Profile* that is discussed below. Each of these clusters consists of a number of zones out of the remaining 54 defined within the case study area in the *Zone Generation* phase. As can be seen, the single zone clusters are all located in the southern end of the case study area and are quite isolated, including mostly remote island and peninsulas. These areas are characterised by low population density and limited public transport services. Consequently, the total flow associated with them is negligible - amounting to no more than 0.01% of all journeys in our database - and the absolute number of residents in those areas is small, factors altering their weekly attractiveness profile. We therefore focus in the following on the four main clusters.

It is evident from the geographical positions of the four main clusters that can be observed in Fig. 6, that the distinctive zone visiting profiles exhibit a clear radial center-periphery pattern. Fig. 7 provides a zoomed in map focusing on Stockholm municipality.

In Fig. 8 we show the center profile for each of the four main clusters. The trend for the four main clusters follows a notable regularity, i.e. they present a similar outline but different amplitude of attractiveness. A

Fig. 5. KM silhouette scores for the investigated range of cluster numbers.

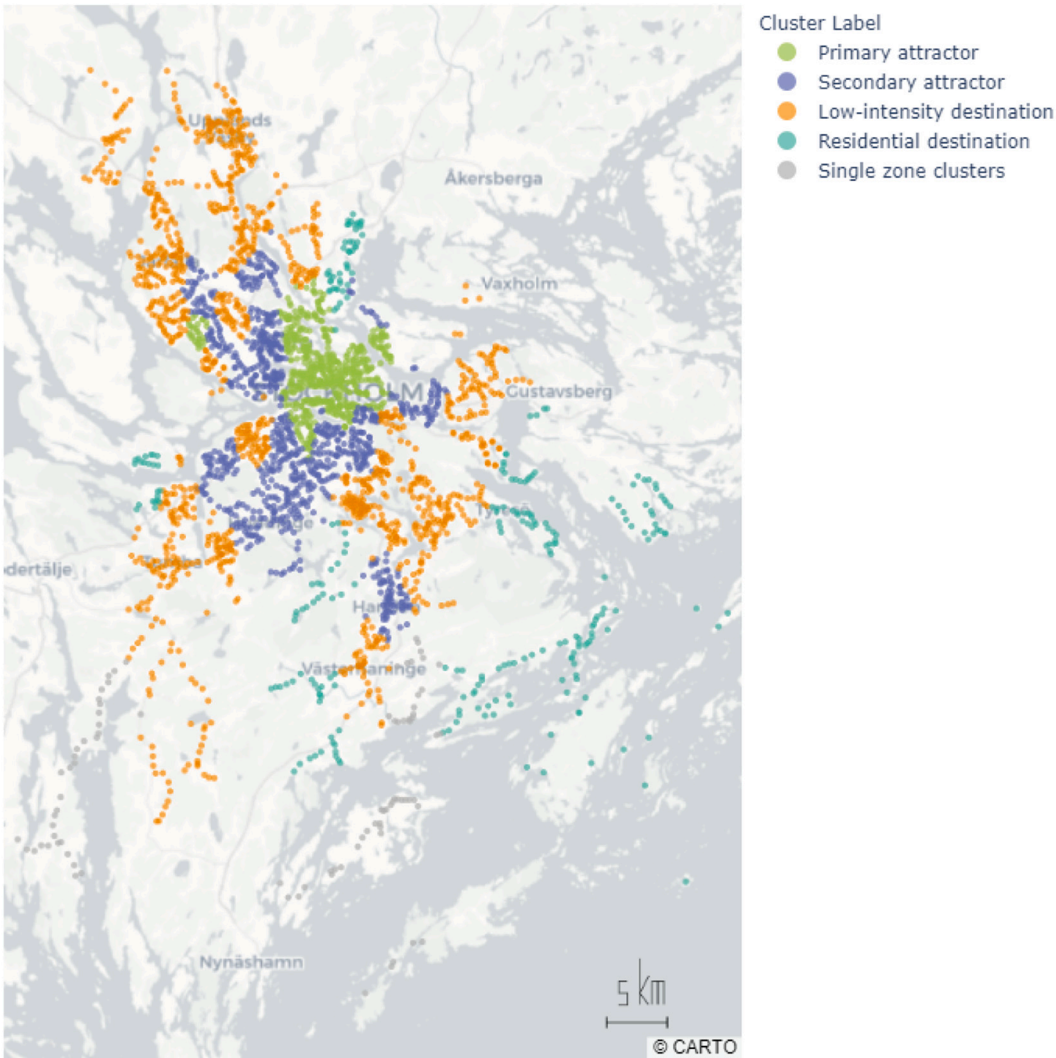


Fig. 6. Spatial distribution of the clusters over the Stockholm urban area. The points in the map correspond to public transport stops, i.e. the origin and destination of the journeys included in the database.

common trend for all four clusters is a peak of attractiveness during office hours on working days and a smoother weekend profile with a lower time-of-day variation.

Based on the visiting profile pattern and the geographical properties we denominate the four clusters as described below. Om the following, we focus on the general features of each cluster and support the geographical description with references to geographical parts within the case study area (for references to local locations, the interested reader may refer to Fig. 2).

- Primary attraction

Areas included in this cluster serve as a primary attraction at the metropolitan level. It can be observed that zones included in this cluster have the highest day peaks of attractiveness with a sharp rise in the share of visitors during office hours. This cluster is also well-visited during weekends with a late afternoon peak on Saturdays that continues into the late evening and an early afternoon peak on Sundays. Even though this cluster consists of the fewest number of zones among the nonsingle ones (Fig. 9a), with only 11.5% of the zones, the share of flows associated with it is the largest, with 55.1% of the total (Fig. 9b). All parts of Stockholm inner-city (consisting of the core districts of Kungsholmen, Östermlam and Södermalm shown in Fig. 2) are included in this cluster. Stockholm inner-city districts are characterised by a very high density and diversity of land-uses and activities including offices, shops, restaurants and cultural venues. In addition, directly neighbouring areas with high job and retail intensity and a wide variety of amenities are also included in this cluster. This includes for example, among others, the north-western district and city, respectively, of Vällingby and Solna. These areas are hosts to major commercial centers and large-scale event facilities. This exemplifies how our method for zonal clustering and characterisation which is agnostic to administrative barriers is able to detect areas of prime attraction that lie outside of the municipal boundaries of the main city.

- Secondary attraction

Areas included in this cluster act as attractions for a limited part of the metropolitan area. This cluster is characterised by a profile similar to the one of Primary attractions, albeit with a considerably less pronounced peak. The share of non-residents is highest during the morning hours of weekdays and is overall lower during the weekends with peaks in the afternoons. This cluster accounts for about the 25% of both the zonal and the flow shares. The zones belonging to this cluster mostly form the first ring of outskirts of the city (often referred to as the inner-suburbs). This includes for example the districts of Hammarbyhöjden

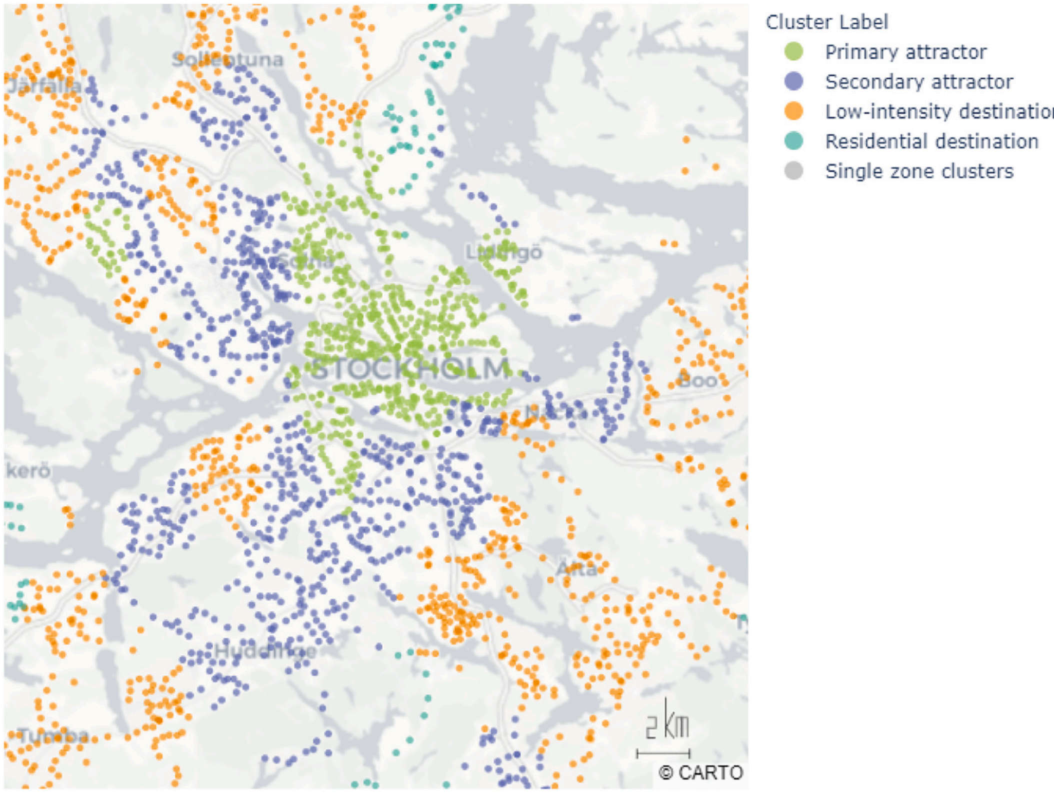


Fig. 7. Close up of the cluster distribution for the municipality of Stockholm.

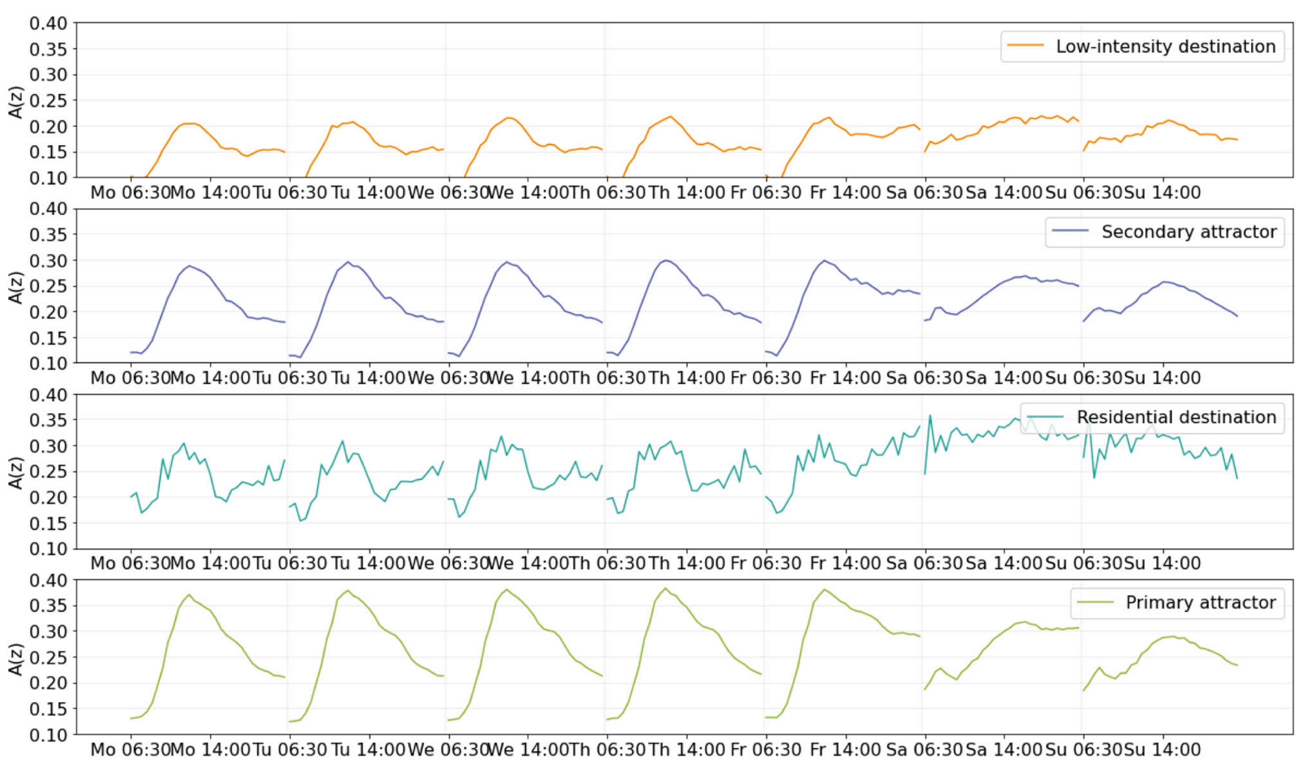
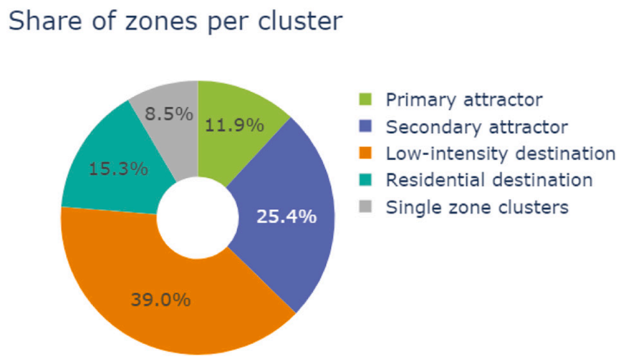


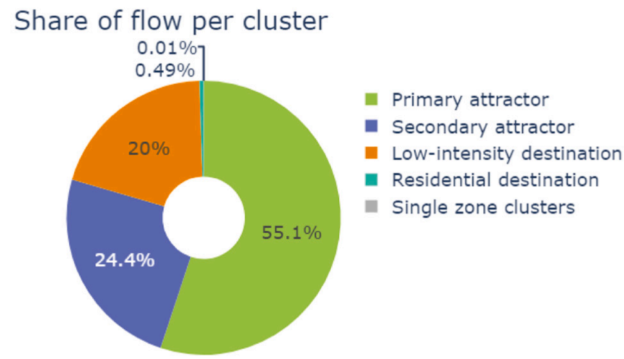
Fig. 8. Center profile for the four most representative clusters.

and Huvudsta (see Fig. 2) as well as parts of Rinkeby, Kista and Tensta, all of which constitute parts of Stockholm municipality, as well as the cities of Sundyberg, Huddinge and Nacka. These areas are characterised by small or middle size office areas and a local retail and commerical function.

- Low intensity destinations



(a)



(b)

Fig. 9. The shares of zones (a) and flows (b) per cluster;

Areas included in this area serve a local function. This cluster consists of those zones that have an overall low visiting frequency by non-residents throughout the week with small peaks around midday on weekdays and in the afternoon on weekends (possibly due to recreational activities). While a plurality of zones belongs to this cluster (39% of the zones), it accounts for only 20% of the total flow. Zones belonging to this cluster are mostly peripheral in relation to our case study area. This includes among others Farsta (a district of Stockholm) and almost all parts of Botkyrka. This illustrates how some areas which are part of the Stockholm municipality can be detected by our method to serve as low-intensity destinations whereas areas which are part of neighbouring municipalities are labeled as either primary or secondary attractors.

- Residential destinations

This last cluster exhibits a less smooth profile which is distinctively different from the three previous ones. The frequency of non-residents visits is highest throughout the weekends. During the week, visiting frequency peaks around midday and after a decrease in the early afternoon increases in the late afternoon and evening. Note that only a small fraction, i.e. 0.5% of the journeys start or end in these zones. Settlements included in this cluster include the municiaplities of Haninge, Tyresö, Danderyd as well as the Stockholm's district of Norsborg.

Combined, these four clusters account for more than the 90% of the zones, and in terms of absolute total flow share they represent almost the totality of the journeys (Fig. 9a and b).

5. Conclusion

We analyse the attractiveness of different parts of a metropolitan area using individual mobility data traces. Our method consists of a series of inference and clustering steps to characterise zonal attraction by analyzing travel patterns across the area of interest. An important feature of our analysis is the identification of the home-zone of each traveller so as to quantify the share of visitors among those attracted to each zone at any time window. We constructed a weekly visiting profile for each part of the case study area by leveraging on longitudinal smart card data transactions. The latter enabled the identification of the share of non-residents among all zonal visitors within each time window.

We applied our method for the case study of the multi-modal public transport system of the Stockholm urban agglomeration area. Our findings demonstrate how smart card data can be used to analyse common features of human activity across an urban region. The results of our clustering based on the weekly visiting profiles reveal four distinctive types of visiting attraction: primary attractions, secondary attractions, low-density attractions and residential destinations. Past studies using smart card data have resorted to a comparable number of clusters and offered insights into the functions of different station across the network. Zhao et al. (2019) have identified five clusters of metro stations for the case of Nanjing, characterised by the authors as residential, central (mix land-use), residential and universities, suburbs with employment areas and major long-distance hubs, based on the land-use functions in proximity to the stations included in each of the clusters. Also studying the network of Nanjing, Gan et al. (2020) identified seven clusters, two of which described as traffic hubs and university influence and all the remaining five as various mixes of employment and residential areas. Similarly, Zhang et al. (2021) identified five communities for the case of London metro stations, each of which corresponding to a geographical section of the Greater London area. In contrast to those works, our method and analysis allow directly characterising the attraction strength and pattern of different parts of the city and similarities and differences among those, that extend beyond static land-use characteristics or station-specific features.

The obtained clusters exhibit a clear center-periphery geographical order. Notwithstanding, our analysis identifies that the boundaries of the primary attractions area extend beyond the inner-city borders to encompass mixed-use neighbouring districts as well and towns as well as a disconnected district characterised by high-intensity retail attractions. Most of the centers of the secondary attractions form a ring of sub-centers within 7–12 km from the city center, inline with the plans of the regional planning authority to stimulate the development of sub-centers in an effort to shift away from its long-term monocentric planning (Cats et al., 2015). From our findings it is evident that our method is capable of identifying and characterising zonal attraction regardless of administrative boundaries by applying a demand-driven zonal generation and the clustering of these zones based on their visiting profiles. Consequently, some of Stockholm's districts are categorised as secondary attractions, low-intensity destinations or even residential destinations, whereas (parts of or the entirety of) some neighbouring municipalities have been labeled as primary or secondary attractions. This result is in line with the findings reported by Zhang et al. (2018) who compared the urban analytic clustering results obtained when using smart card data versus taxi usage data. While in general both travel demand data generate clusters that are spatially cohesive, those produced using by taxi data corresponded more to administrative borders and encompassed larger geographical areas when compares to those obtained from smart card data.

The results of this study can be used to inform planners and decision makers about the main activity locations of travellers and how their temporal patterns vary across the metropolitan area. Tailored fare products can be designed to cater for different user segments. The findings can also be used to assess policies aimed at stimulating or revitalising the generation of urban centers and attractions. For example, the City of Stockholm has adopted the policy of promoting a more polycentric structure by supporting the development of strategic nodes (Council, 2010). Repeating this analysis over time will allow assessing the extent to which such developments have materialized. Furthermore, future research may investigate how the visiting profiles vary for different user segments in terms of their socio-demographic characteristics. For example, tourists and occasional visitors may exhibit different visiting patterns than those of local residents.

CRediT authorship contribution statement

Oded Cats: Conceptualization, Methodology, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Francesco Ferranti:** Data curation, Formal analysis.

Declaration of competing interest

The authors have no competing interests to declare.

Acknowledgment

This study is funded by Region Stockholm, project “Unravelling travel demand patterns using Access card data” RS 2019-0499. We also thank Region Stockholm for providing the smart card data that made this study possible. The authors also thank Isak Rubensson, Matej Cebecauer and Erik Jenelius for their support in the process.

References

Ahmed, A., & Stopher, P. (2014). Seventy minutes plus or minus 10—a review of travel time budget studies. *Transport Reviews, 34*(5), 607–625.
Cats, O. (2016). The robustness value of public transport development plans. *Journal of Transport Geography, 51*, 236–246.
Cats, O. (2017). Topological evolution of a metropolitan rail transport network: The case of Stockholm. *Journal of Transport Geography, 62*, 172–183.
Cats, O., Wang, Q., & Zhao, Y. (2015). Identification and classification of public transport activity centres in Stockholm using passenger flows data. *Journal of Transport Geography, 48*, 10–22.
Council, S. C. (2010). *The walkable city: Stockholm city plan*. Stockholm: Stockholm City Council.
Gan, Z., Yang, M., Feng, T., & Timmermans, H. (2020). Understanding urban mobility patterns from a spatiotemporal perspective: Daily ridership profiles of metro stations. *Transportation, 47*(1), 315–336.
Jacobs-Crisioni, C., Rietveld, P., Koomen, E., & Tranos, E. (2014). Evaluating the impact of land-use density and mix on spatiotemporal urban activity patterns: An exploratory study using mobile phone data. *Environment and Planning A, 46*(11), 2769–2785.
Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., & Newth, D. (2015). Understanding human mobility from twitter. *PloS one, 10*(7), Article e0131469.
Kholodov, Y., Jenelius, E., Cats, O., van Oort, N., Mouter, N., Cebecauer, M., & Vermeulen, A. (2021). Public transport fare elasticities from smartcard data: Evidence from a natural experiment. *Transport Policy, 105*, 35–43.
Kim, K. (2019). Identifying the structure of cities by clustering using a new similarity measure based on smart card data. *IEEE Transactions on Intelligent Transportation Systems, 21*(5), 2002–2011.
Lee, K.-S., You, S. Y., Eom, J. K., Song, J., & Min, J. H. (2018). Urban spatiotemporal analysis using mobile phone data: Case study of medium-and large-sized Korean cities. *Habitat International, 73*, 6–15.
Liu, Y., Gao, S., Yuan, Y., Zhang, F., Kang, C., Kang, Y., & Wang, K. (2021). Methods of social sensing for urban studies. *Urban Remote Sensing: Monitoring, Synthesis, and Modeling in the Urban Environment*, 71–89.
Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers, 105*(3), 512–530.
Louail, T., Lenormand, M., Picornell, M., Cantú, O. G., Herranz, R., Frias-Martinez, E., Ramasco, J. J., & Barthelemy, M. (2015). Uncovering the spatial structure of mobility networks. *Nature Communications, 6*(1), 1–8.
Luo, D., Cats, O., & van Lint, H. (2017). Constructing transit origin–destination matrices with spatial clustering. *Transportation Research Record, 2652*(1), 39–49.
Munizaga, M. A., & Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies, 24*, 9–18.
Ozus, E., Akn, D., & Çiftçi, M. (2012). Hierarchical cluster analysis of multicenter development and travel patterns in Istanbul. *Journal of Urban Planning and Development, 138*(4), 303–318.
Perlman, J., & Roy, S. S. (2021). Analysis of human movement in the Miami metropolitan area utilizing uber movement data. *Cities, 119*, Article 103376.
Roth, C., Kang, S. M., Batty, M., & Barthélemy, M. (2011). Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PlOS One, 6*(1), Article e15923.
Sari Aslam, N., Cheng, T., & Cheshire, J. (2019). A high-precision heuristic model to detect home and work locations from smart card data. *Geo-spatial Information Science, 22*(1), 1–11.
Schläpfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., Anklesaria, S., Vazifeh, M., Ratti, C., & West, G. B. (2021). The universal visitation law of human mobility. *Nature, 593*(7860), 522–527.
Seaborn, C., Attanucci, J., & Wilson, N. H. (2009). Analyzing multimodal public transport journeys in London with smart card fare payment data. *Transportation Research Record, 2121*(1), 55–62.
Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities, 55*, 9–21.
Trépanier, M., Tranchant, N., & Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems, 11*(1), 1–14.
Tu, W., Cao, J., Yue, Y., Shaw, S.-L., Zhou, M., Wang, Z., Chang, X., Xu, Y., & Li, Q. (2017). Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science, 31*(12), 2331–2358.
Wang, Z., Liu, H., Zhu, Y., Zhang, Y., Basiri, A., Büttner, B., Gao, X., & Cao, M. (2021). Identifying urban functional areas and their dynamic changes in Beijing: Using multiyear transit smart card data. *Journal of Urban Planning and Development, 147*(2), 04021002.
Widhalm, P., Yang, Y., Ulm, M., Athavale, S., & González, M. C. (2015). Discovering urban activity patterns in cell phone data. *Transportation, 42*(4), 597–623.
Wu, C., Ye, X., Ren, F., & Du, Q. (2018). Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in Shenzhen, China. *Cities, 77*, 104–116.
Wu, L., Zhi, Y., Sui, Z., & Liu, Y. (2014). Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PloS one, 9*(5), Article e97010.
Xie, C., Yu, D., Zheng, X., Wang, Z., & Jiang, Z. (2021). Revealing spatiotemporal travel demand and community structure characteristics with taxi trip data: A case study of New York city. *PLoS one, 16*(11), Article e0259694.
Ye, C., Zhang, F., Mu, L., Gao, Y., & Liu, Y. (2021). Urban function recognition by integrating social media and street-level imagery. *Environment and Planning B: Urban Analytics and City Science, 48*(6), 1430–1444.
Zhang, X., Xu, Y., Tu, W., & Ratti, C. (2018). Do different datasets tell the same story about urban mobility—A comparative study of public transit and taxi usage. *Journal of Transport Geography, 70*, 78–90.
Zhang, Y., Liu, L., & Wang, H. (2019). A new perspective on the temporal pattern of human activities in cities: The case of shanghai. *Cities, 87*, 196–204.
Zhang, Y., Marshall, S., Cao, M., Manley, E., & Chen, H. (2021). Discovering the evolution of urban structure using smart card data: The case of London. *Cities, 112*, Article 103157.
Zhao, X., Wu, Y.-P., Ren, G., Ji, K., & Qian, W.-W. (2019). Clustering analysis of ridership patterns at subway stations: A case in Nanjing, China. *Journal of Urban Planning and Development, 145*(2), 04019005.