5.2 Visual Analytics for Characterizing Mobility Aspects of Urban Context

Gennady Andrienko, Natalia Andrienko, Fabian Patterson, Siming Chen, Robert Weibel, Haosheng Huang, Christos Doulkeridis, Harris Georgiou, Nikos Pelekis, Yannis Theodoridis, Mirco Nanni, Leonardo Longhi, Athanasios Koumparos, Ansar Yasar, and Ibad Kureshi

Gennady Andrienko, Natalia Andrienko, Fabian Patterson, Siming Chen
Fraunhofer Institute IAIS
Sankt Augustin, Germany
Email Gennady.andrienko@iais.fraunhofer.de

Robert Weibel, Haosheng Huang
University Zurich
Zurich, Switzerland

Christos Doulkeridis, Harris Georgiou, Nikos Pelekis, Yannis Theodoridis
University Piraeus
Piraeus, Greece

Mirco Nanni
CNR
Pisa, Italy

Leonardo Longhi
Sistematica
Terni, Italy

Athanasios Koumparos
Vodaphone
Athens, Greece

Ansar Yasar
University Hasselt
Hasselt, Belgium

Ibad Kureshi
Inlecom Group BVBA
Brussels, Belgium

Abstract. Visual analytics science develops principles and methods for efficient human-computer collaboration in solving complex problems. Visual and interactive techniques are used to create conditions in which human analysts can effectively

utilize their unique capabilities: the power of seeing, interpreting, linking, and reasoning. Visual analytics research deals with various types of data and analysis tasks from numerous application domains. A prominent research topic is analysis of spatio-temporal data, which may describe events occurring at different spatial locations, changes of attribute values associated with places or spatial objects, or movements of people, vehicles, or other objects. Such kinds of data are abundant in urban applications. Movement data are a quintessential type of spatio-temporal data because they can be considered from multiple perspectives as trajectories, as spatial events, and as changes of space-related attribute values. By example of movement data, we demonstrate the utilization of visual analytics techniques and approaches in data exploration and analysis.

5.2.1 Introduction

The science of visual analytics (Thomas and Cook, 2005) develops principles, methods, and tools to enable synergistic work between humans and computers through interactive visual interfaces. Such interfaces support the unique capabilities of humans (such as the flexible application of prior knowledge and experiences, creative thinking, and insight) and couple these abilities with machines' computational strengths, enabling the generation of new knowledge from large and complex data.

In this chapter, we describe visual analytics approaches that are related to the study of urban mobility data and discuss how visual analytics can support analysis of such data and informed, justifiable decision making. We address different stages of the urban data science process, including data quality assessment, data transformation, exploration and analysis, and indicate possibilities for model building, evaluation, and refinement. We conclude this chapter with a summary of achievements, unsolved problems, and future research directions.

We demonstrate the utilization of visual analytics techniques in a process of exploration and analytical reasoning using a real world data set. In the EU-funded Track&Know project[1], one of industrial partners collects Europe-wide tracks of passenger cars. The data are collected for insurance purposes under vehicle owners' informed consent, aiming at enabling transparent pricing and facilitating analysis of accidents. For these purposes, it is necessary to have an understanding of the context in which the vehicles move, which includes the traffic around. There are several questions that require answers for understanding traffic: What are the major flows and their properties? How do they vary over time? What is the composition of the types of the cars appearing on streets? What are regular and irregular trips and how are they distributed in space and time? etc. Answers to these questions can be valuable for a variety of practical applications such as assessing which part of traffic can

---

[1] Track&Know, grant agreement 780754: https://trackandknowproject.eu/

be potentially served by publicly shared vehicles or by electric cars, evaluating applicability of various car sharing schemes, identifying and assessing different driving styles, and investigating events, such as traffic accidents, in their context.

5.2.2 State of the art

Batty (2013) considers a city as a system composed of flows (between locations and between activities) and networks of relationships and interactions among various entities. For understanding these factors of the urban context, a variety of different data sources is considered. There are studies (e.g. Kesting and Treiber 2013) based on stationary sensors such as traffic counters that record aggregated characteristics (how many cars passed a given street segment during some time interval and what was their speed). Such sensors record aggregates but do not allow tracing vehicles. Another kind of stationary sensors is docking stations for rental bicycles (or, potentially, other kinds of shared vehicles). Usually these sensors provide only general characteristics (overall capacity, numbers of docked bicycles and empty slots) and their aggregates over time intervals. However, sometimes more detailed data are released, enabling analysis of the moves of the vehicles between the docking stations (Beecham and Wood, 2014). Some researchers approximate mobility from space- and time-referenced social media records. A prominent example is provided by Lansley and Longley (2016) who studied in detail the distribution of the message topics in space and their variation over time. Itoh el al (2016) studied data of smart card usage in local trains together with social media records for reconstructing temporal characteristics of major flows and understanding abnormal situations.

Several review papers discussed visual analytics approaches to analyzing mobility and transportation. A review by Andrienko and Andrienko (2013a) considered approaches from the data processing perspective: looking at trajectories, clustering trajectories, transforming times in trajectories, and studying attributes, events and patterns in trajectories, followed by generalization and aggregation of trajectories and tracing derived flows. A more recent review on visual analytics of mobility and transportation Andrienko et al. (2017) outlines approaches used for following problems: understanding details of individual movement, studying variety of taken routes, assessing movement dynamics along a route, linking origins and destinations, characterizing collective movement over a territory, detecting events and studying their distributions, contextualizing movement, and studying impacts and risks.

Markovic et al. (2019) present a viewpoint of a road transportation agency, mentioning the following problems of interest: demand estimation, modeling human behavior, designing public transit, measuring and predicting traffic performance, assessing impact on environment, and improving road safety.

The reviews indicate the need to consider the movement data from multiple perspectives. We follow this approach in our work.

5.2.3 Mobility Data: properties and problems

To demonstrate the data analysis workflow, we use trajectories of 4,521 passenger cars within the Greater London area that were recorded during two regular weeks in winter 2017; 4,284,493 position records in total. Each position record consists of an anonymized identifier of a vehicle, time stamped geographic coordinates, and attributes such as momentary speed and heading, GPS signal quality etc. Transport for London estimates the number of all cars registered in London as about 2.6 million[2]. Respectively, our data set covers about 0.2% of the active "population" of the passenger cars. Figures 1 and 2 show the spatial and temporal distributions of the recorded trajectories. From the map (Fig.5.2.1), we can recognize the major roads and populated areas.
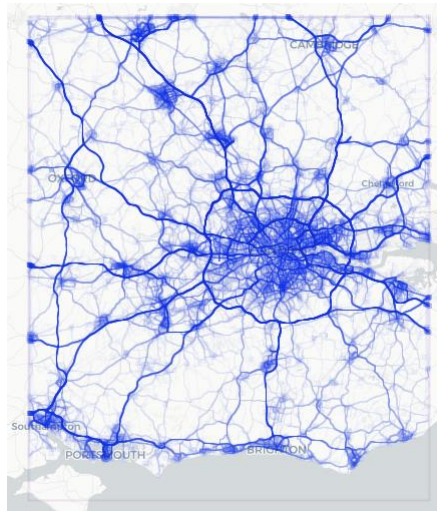

Fig. 5.2.1. Spatial footprint of all trajectories in the data set.

The time histogram (Fig.5.2.2) reflects the distribution of the counts of distinct cars per hour, starting from Sunday midnight: 2 weeks x 7 days x 24 hours = 336 hours in total. The time histogram clearly shows the weekly cycle and distinct profiles of weekdays and weekends.

---

[2] http://content.tfl.gov.uk/technical-note-12-how-many-cars-are-there-in-london.pdf
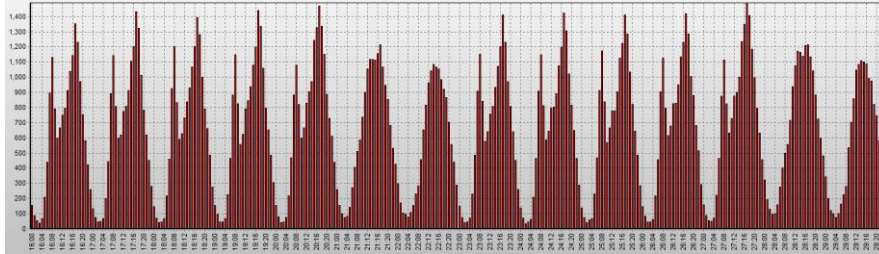
Fig. 5.2.2. Temporal profile of the data: the bars represent the car counts per hour.

For assessing the quality of the data set, we follow the approach proposed by Andrienko, Andrienko and Fuchs (2016). Possible problems in movement data include problems of coverage and accuracy that may occur in all components of the data, namely space, time, identifiers, and attributes. Respectively, we assess properties of all data components and their combinations.
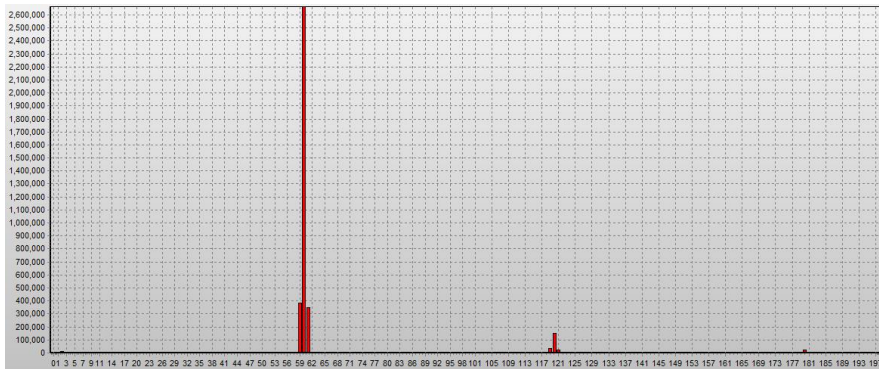
Fig. 5.2.3. Sampling rates.

For the temporal component, we start with examining the sampling rates, i.e., the time intervals between consecutive position recordings for the same car. The statistics (Fig.5.2.3) demonstrates that the most frequent sampling rate is around 1 minute (59-61 seconds). A much smaller subset of points is characterized by the sampling rate of about 2 minutes, and only a few points have 3 minutes intervals to the next points. All other intervals appear in the data infrequently. Next, we checked if the sampling rate of 1 minute is typical for all cars. For this purpose, we calculated the median sampling rate for each car. The results demonstrate that more than 98% of the cars have the median sampling rate of 1minute +/- 1 second. However, we have identified a few outliers: about 100 cars that had only a few positions recorded and, correspondingly, rather arbitrarily sampling rates; 9 cars with many recorded positions but the median sampling rates of 3 to 5 minutes; and 2 cars with very high sampling rates (13 seconds). Such outliers need to be separated in further analysis.

We have also identified several thousands of duplicate pairs of an identifier and a time stamp and excluded the duplicates.
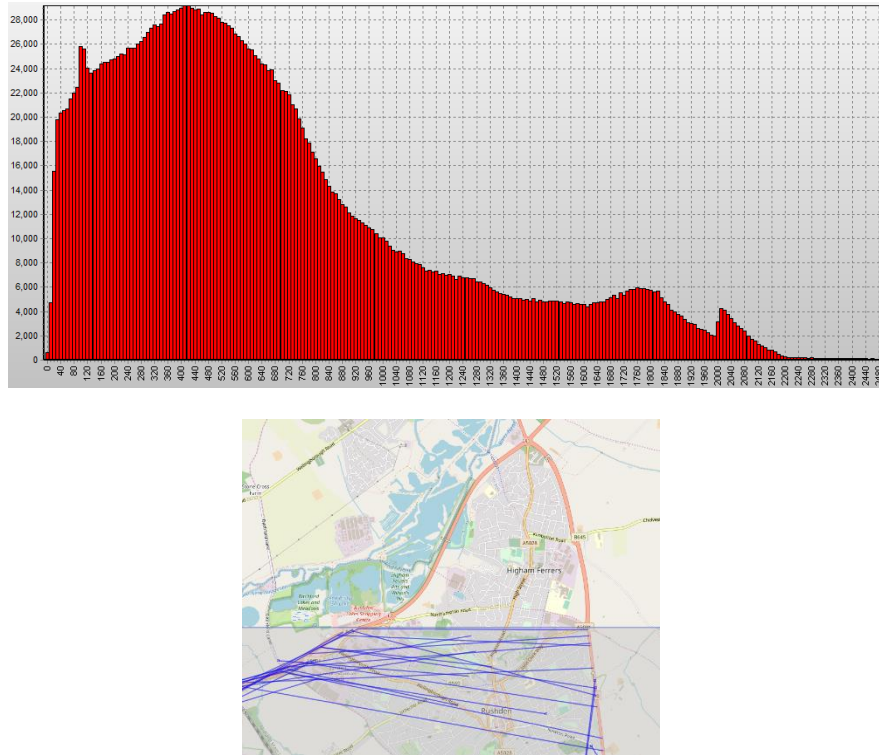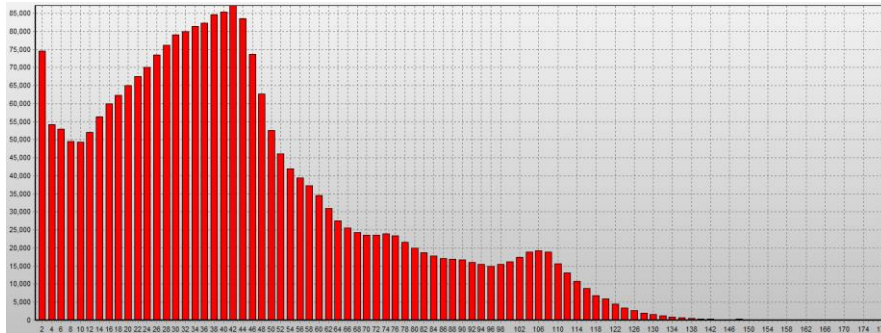

Fig. 5.2.4. Top: frequency distribution of the distances between consecutive points of trajectories. Bottom: long distances between consecutive points are caused by selecting data that fit in a chosen bounding rectangle (border effects).

Fig.5.2.4 shows the frequency distribution of the distances between consecutive position records, with the bins corresponding to 10m intervals. We can observe major peaks at 420m and 1760m. Since the typical sampling rate is 1 minute, these peaks correspond to displacement speeds 25.2 and 105.6 km/h. We also observe narrower peaks at 100m (6 km/h) and 2000m (120 km/h). The former may correspond to small displacements caused by waiting at street intersections. We inspected the second peak separately. Such distances between points appear either at highways and may mean that some points were not recorded (e.g., due to bad satellite connection), or at the borders of the studied area (Fig.4 bottom). These large displacements at the area boundaries are artifacts of data selection by a bounding rectangle.

Figure 5.2.5 presents the frequency distribution of the instant speed values in the positional data after excluding numerous (about 778,000) stationary points and a

[Type here]

few outliers with speeds higher than 180 m/h. The clearly visible peaks roughly correspond to the speed limits on different categories of the UK roads.

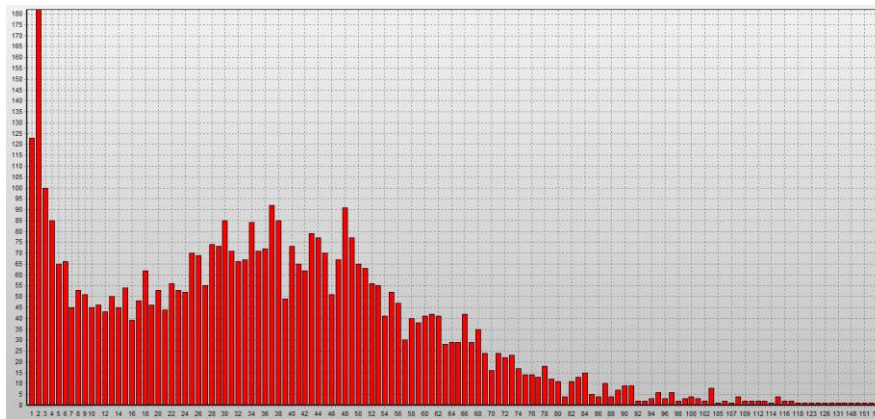**Fig. 5.2.5.** Frequency distribution of the speeds after removing stationary positions and outliers.

Figure 5.2.6 shows the frequency distribution of the measured vehicle headings in the non-stationary points. There are two strange pits around the values 90 and 270 degrees. It is quite unlikely that these directions were really much less frequent than the others. The pits may be due to the method that is used by the tracking devices for determining the vehicle heading. The method may calculate the angle based on the ratio of the x- and y-differences between two consecutively measured positions (of which the second position is not recorded) and fail in cases when the y-difference equals zero. Whatever the reason is, the measured heading values cannot be trusted.
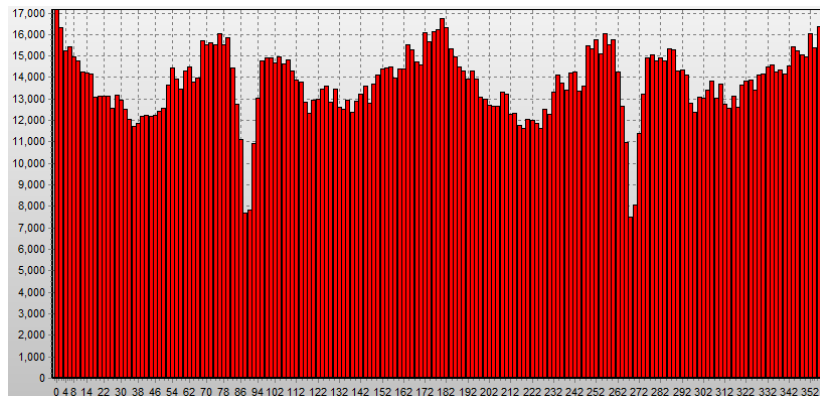
Fig. 5.2.6. Frequency distribution of the measured vehicle headings.

For human mobility studies, it is important to divide trajectories into trips, e.g., between places of significant stops (Andrienko et al., 2013a). There exist different criteria for separating trips: by positional attributes (e.g. taximeter is switched on or off), by temporal cycles (e.g. daily trips), by substantial displacement (e.g. if the

next point is at least 5km away) and by temporal gaps between points (no movement for at least 15min). We used the latter criterion. For tolerating position measurement errors, the periods when positions remained within a small area during a time interval of a chosen length (15 minutes) were also treated as stops. In this way, we acquired 164,644 sub-trajectories, from which 3,943 consisted of single points and were excluded from the further consideration. The remaining sub-trajectories were treated as representing trips. Figure 5.2.7 presents the frequency distribution of the trip counts per one car. About 300 cars had only 1 or 2 trips during the two weeks. Many cars performed from 30 to 50 trips, and only a few cars had more than 80 trips.

Fig. 5.2.7. Frequency distribution of the trip counts per car.

Figure 5.2.8 presents an example of all trips of a single car during two weeks. A map on the left shows the spatial footprint. A space-time cube (Hagerstrand, 1969; Kraak, 2003) shows the same trips in the space and time simultaneously. The vertical axis represents the time of the day. The colors encode the weekdays (green) and weekends (red). Generally, such a visualization may enable identifying the person whose track is shown; therefore, we have masked the locations on the map and will avoid disclosing any further potentially privacy-sensitive details in the text or illustrations.

[Type here]
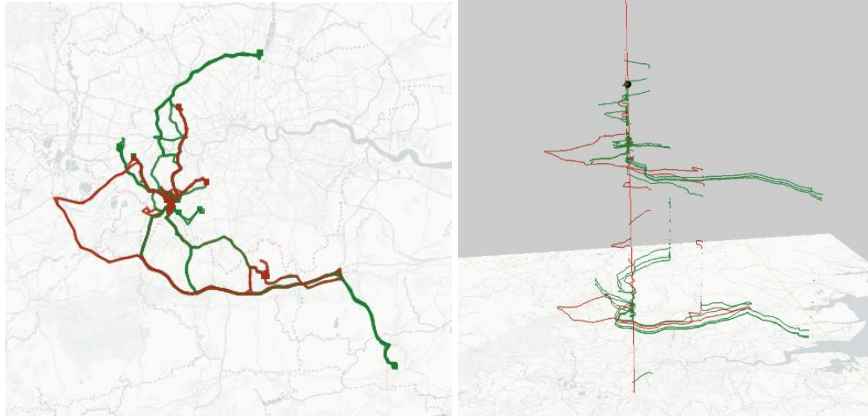
Fig. 5.2.8. Trips of a single car are represented on a map (left) and in space-time cube (right), in which the trips have been temporally aligned within the daily time cycle. The colors denote whether the trips took place on weekdays (green) or weekends (red).

After performing the investigation of the data properties and cleaning the data by excluding incomplete tracks and incorrect values, we can proceed with analysis.

5.2.4 Data types: Events, trajectories, spatial time series, situations

There exists a range of transformations that can be applied to movement data for analyzing them in various ways and extracting different kinds of information. First of all, each recorded position is a *spatial event*, which is specified by a reference to the moving object *id*, time stamp *t*, and coordinates *x* (longitude) and *y* (latitude). An event may also have attributes:

$$id, t, x, y, attributes.$$

The events of moving objects being at specific spatial positions at particular times can be called *position events* for distinguishing from other kinds of spatial events. Integration of chronologically arranged position events of the same moving object produces a *trajectory* of this object (Fig.5.2.9). Such integration allows computation of derived attributes based on the positions of consecutive points: displacement distance and direction, time difference, speed estimate etc. These derived attributes can be used for extracting secondary events from trajectories (e.g. stops) and dividing trajectories into smaller subsets (e.g. trips between stops). We applied these transformations when investigating the data properties.

Both trajectories and events can be spatially aggregated by a set of places. As a result, the places are characterized based on the visits by moving objects (e.g., counts of the objects and the visits, statistics of the duration of object presence in the area, etc.) or the events that occurred in them (e.g., counts of events of different kinds). The aggregation can be performed by time intervals producing *place-based*

time series of the visits and presence. Additionally, trajectories can be aggregated according to the moves (transitions) between areas. The transitions link the areas, and these links can be characterized based on the number and properties of the transitions, such as the number of distinct objects that moved and the statistics of the speeds and durations. Aggregated transitions between places are usually called *flows*. The aggregation can also be made by time intervals resulting in *link-based time series* of flow characteristics.
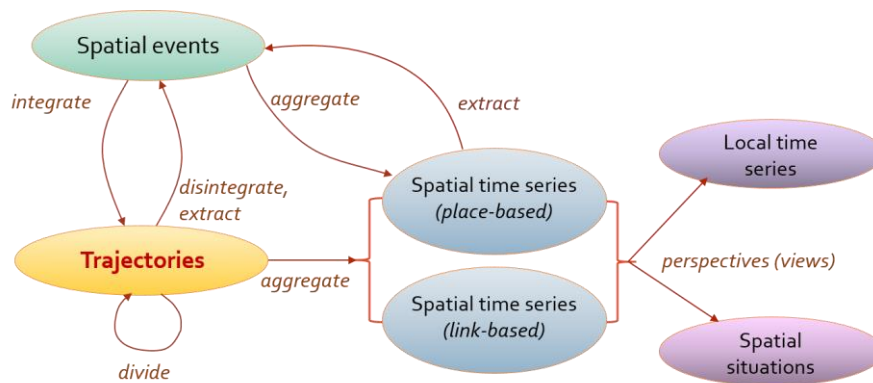
Fig. 5.2.9. A general scheme of movement data transformations.

Spatial time series can be viewed in two complementary ways. On the one hand, they consist of sequences of values associated with individual places or links, which can be called *local time series*. Respectively, the places or links can be characterized and compared based on the temporal variation of the respective values. On the other hand, for each time step, there exists a particular distribution of the values over the set of places or links. This distribution can be called a *spatial situation*. The whole spatial time series can be seen as a sequence of such spatial situations. Respectively, the temporal variation of the spatial situations can be studied and characterized.

Further events (e.g., occurrences of extreme values) can be extracted from place- or link-based spatial time series.

Data transformations support investigation of different aspects of mobility phenomena. As our goal is characterization of urban context, we expect that transformations will allow us to enrich the context by different kinds of relevant information.

5.2.5 Context acquisition from movement data

Traffic and mobility are important parts of the overall urban context. Information concerning movements of vehicle and people in an urban area may be relevant in

studying various phenomena, such as air quality, noise, or disease spread, and events, such as traffic accidents, crimes, or disruptions in the work of public transport. Movement-related context information that can be extracted from trajectory data includes place visiting context, flow context, time context, trip context, and personalized semantic context. We shall consider a selection of the listed aspects in detail in the following sections.

5.2.5.1 Place visiting context

For describing the context in terms of place visits, it is necessary to have a suitable set of places. When there are no predefined places suiting the goals of an intended study, the places need to be appropriately defined. One possible way to do this is taking the neighborhoods of some positions of interest, e.g., circles of a chosen radius around the positions of studied events. Places relevant to transportation studies can be defined based on the street segments and intersections. However, the resulting level of detail and amount of data can be excessive for the envisaged spatial scale of the intended study. For studies of human mobility behaviors, places can be defined based on identifying areas of different kinds of human activities.

A set of places can also be derived by partitioning the territory into compartments based on the spatial distribution of some data, such as positions of stationary objects, events, or points from vehicle trajectories. Andrienko and Andrienko (2011) proposed to divide a territory based on the distribution of characteristic points of trajectories, which include the positions of stops and turns as well as trip starts and ends. The points are extracted from the trajectories and grouped according to their spatial locations. A special method for space-bounded point clustering produces spatial clusters whose radii do not exceed a given threshold. The medoids of the clusters (i.e., the points with the smallest mean distances to the other cluster members) are taken as generating seeds for Voronoi tessellation. When the points are not evenly spread throughout the territory but form dense clusters, the seeds tend to be taken from these clusters, which makes the resulting places meaningful and interpretable. Depending on the chosen maximal radius of a point cluster, the territory is divided into larger or smaller compartments. Hence, an analyst can adjust the partitioning to the spatial scale of the intended analysis and the desired level of detail.

An example of territory partitioning based on trajectory data is shown in Fig. 10. The characteristic points have been grouped in clusters with the maximal radius 2.5 km. As a result, we have obtained 3,535 places (compartments). It can be observed that the geometries and the spatial layout of the places reflect the topology of the major roads. This is the effect of taking seeds for the tessellation from dense concentrations of trajectory points, which mainly occurred along these roads. The places in Fig. 5.2.10 are colored according to the numbers of distinct cars that vis-

ited them. As we mentioned earlier, other characteristics of places that can be derived from movement data are time series of place visits and their durations and aggregate characteristics of the objects that visited the places.
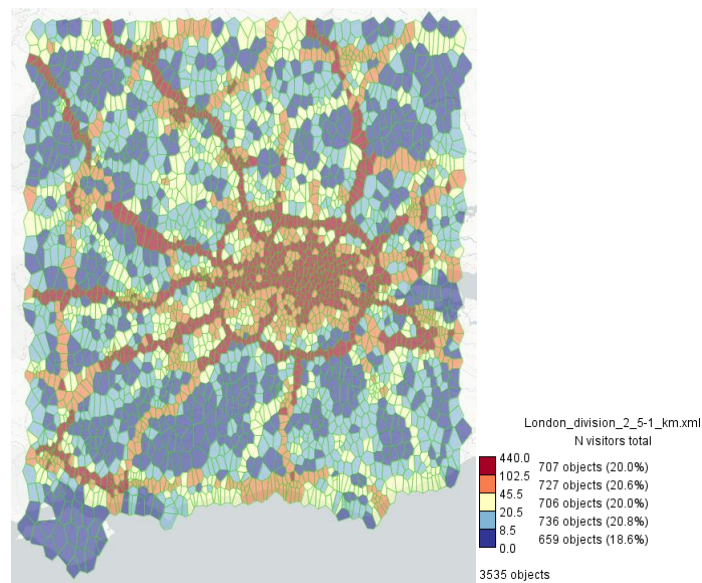

Fig. 5.2.10. Tessellation of the region into 3,535 polygons based on point clustering bounded by a maximal cluster radius of about 2.5 km. Colors represent counts of distinct cars observed in each region, from blue (less than 8) to red (more than 102), using equal class size division.

Thus, our data allow us to characterize the places based on the "population structure" of the cars that visited them. The data set includes car manufacturer information for each anonymized car identifier. Respectively, it is possible to obtain separate car counts for different manufacturers. Using this information, we would like to cluster the places by the similarity of the car population structures. However, a straightforward application of clustering to the absolute counts just separates areas by total car counts, replicating the major patterns visible in Fig.5.2.9. Therefore, it is necessary to normalize the counts by the total numbers of different cars recorded in each compartments, thus obtaining proportional values.

We have clustered the normalized counts using a partition-based clustering method k-Means in combination with a projection of the cluster centroids onto a plane, as suggested by Andrienko and Andrienko (2013). The results are presented in Fig.5.2.11. The positions of the cluster centroids on the projection plane (top left) are used for selecting appropriate clustering parameters and then for assigning colors to clusters reflecting their similarities and differences. The cluster profiles in terms of the proportions of the cars from different manufacturers are shown in a bar chart (top right) and on a map (bottom left).

[Type here]

The clustering results show that the main motorways are dominated by Vauxhall, Ford and VW, while the central London and Brighton are characterized by a mix of everything, with some prevalence of Vauxhalls and Fords. One can find compact "villages" in rural areas populated mostly by Fiat, Ford, SEAT, Peugeot or VW.

Places can also be grouped according to the place-based time series of visits or counts of distinct cars, either in absolute or normalized form. We omit such analysis here due to space restrictions. However, we shall consider link-based time series in the next section.
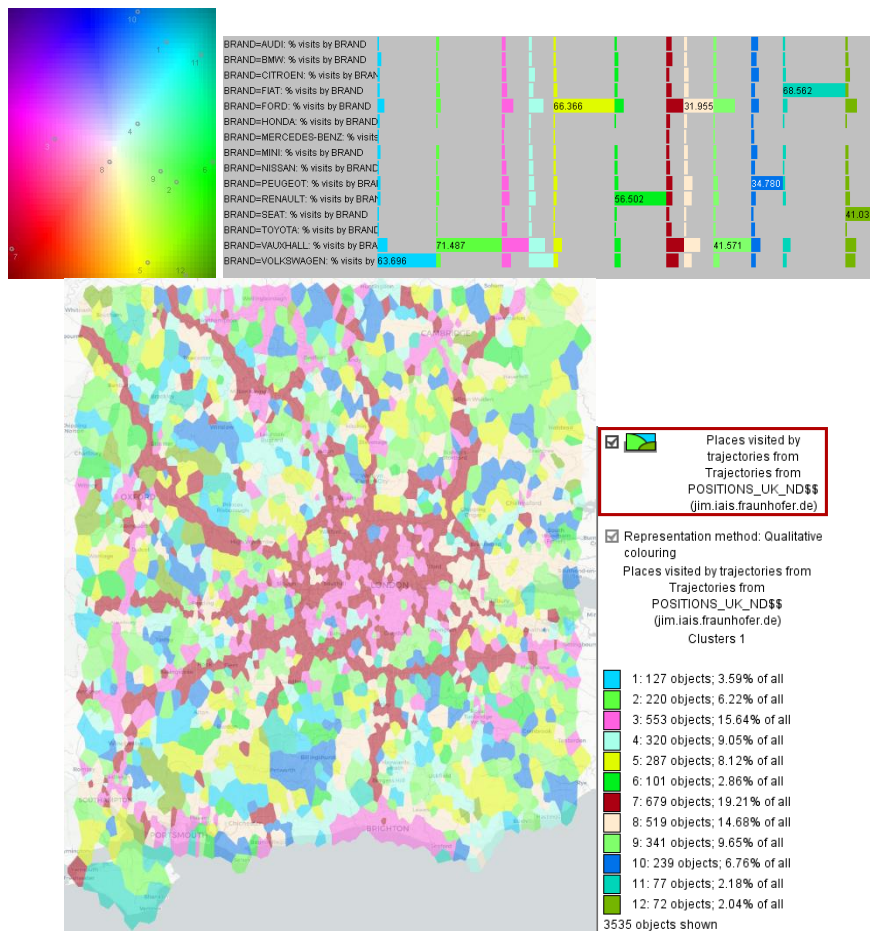
Fig. 5.2.11. Clustering of places by similarity of the car population structure. Top: a 2D projection of the cluster centers (left) and the profiles of the clusters in terms of the attributes involved in the clustering (right). Bottom: a map of the spatial distribution of the clusters (left) and the corresponding legend showing the cluster sizes (right).

5.2.5.2 Flow context

While place-based time series characterize a territory in terms of the spatio-temporal variation of the presence of moving objects or events, link-based time series complement the characterization by describing the volumes and characteristics of movements (flows) between the places. In this section, we present an example of analyzing the flows between the same places as in Figs. 5.2.10 and 5.2.11. For the set of 3,535 places, we obtain 13,153 directed links when we use the original trajectories and 12,654 links when we use the trajectories corresponding to the trips (resulting from dividing the original trajectories based on stops for 15 minutes or more). The divided trajectories are more appropriate for characterization of movement speeds.
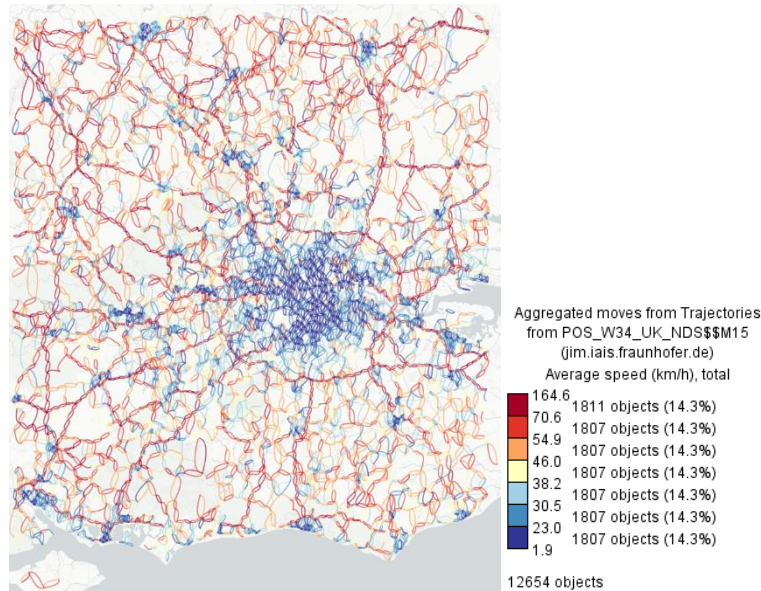

Fig. 5.2.12. Average speeds of the flows between the places.

Figure 5.2.12 presents a map where the links are represented by curved lines colored according to the average speeds during the transitions between the places. Similarly to Fig. 5.2.10, this map reflects the properties of the road network and the spatial distribution of the urban areas. Each pair of places is connected by two lines reflecting movements in opposite directions. We can notice that for the majority of the location pairs there is no substantial difference between the average speeds in the opposite directions. However, aggregates that reflect the temporal variation, such as the hourly flow volumes over the two weeks, may reveal asymmetry between the flows in opposite directions.

In Fig. 5.2.13, we have applied k-means clustering to the flow volumes normalized by the each link's mean value after exclusion of the links with very low flows

[Type here]

(less than 50 moves in total during the 2 weeks period). As in the previous section (Fig.11), the parameters for the clustering were selected by inspecting the positions of the clusters centroids in the projection space, and the projection was also used for assigning colors to the clusters. Clusters whose centroids are close in the projection space due to the similarity of the respective attribute values receive similar colors. In the map in Fig. 13, we can observe the consistency of cluster affiliation along chains of links following the major roads; hence, the traffic has common patterns along the major transportation corridors formed by the most important motorways. We can also notice pairs of opposite links that were put in distinct clusters, which means that the temporal patterns of the respective flows differ.
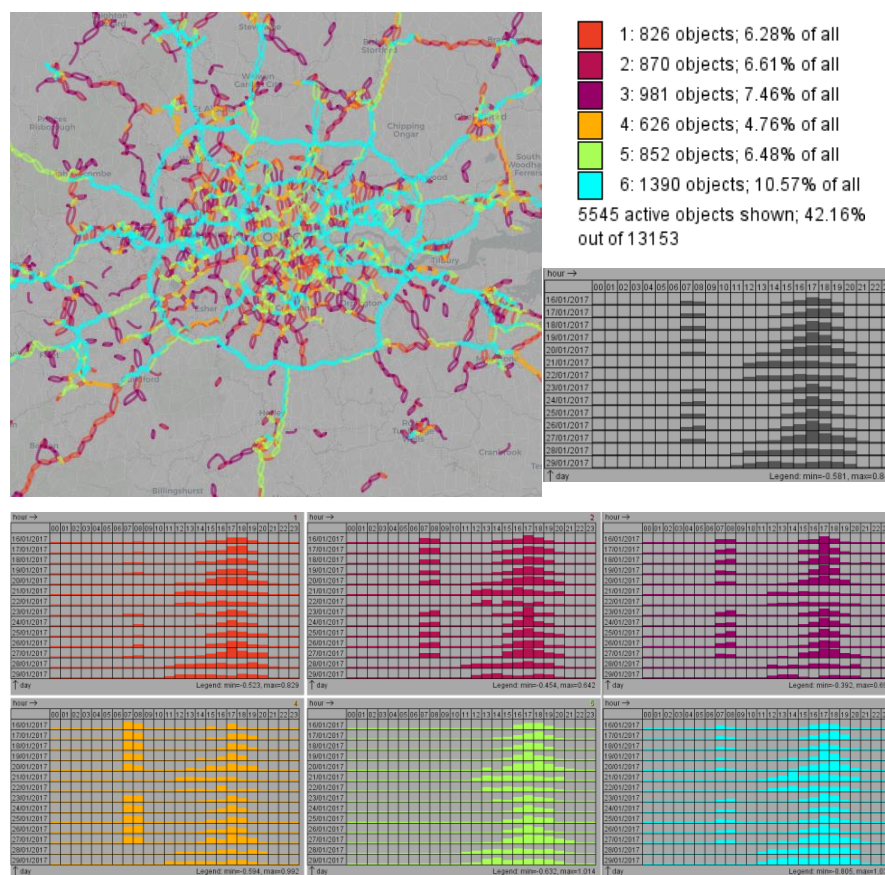

Fig. 5.2.13. The links clustered according to the similarity of the normalized time series of flow volumes. Top: a map with the links colored according to their cluster affiliation; the legend shows the cluster sizes. Bottom: the cluster profiles are represented in an aggregated form in 2-dimensional histograms with the rows corresponding to days and columns to hours. The heights of the colored bars in the cells are proportional to the mean normalized hourly values for the clusters. The 2D histogram with the dark gray bar shows the average temporal variation for all links.

5.2.5.3 Time context

Mobility is essentially a temporal phenomenon; thus, the distribution of people and vehicles over a territory and their movements from place to place vary over time. As human activities are cyclic in general, we can expect temporal cycles to appear in aggregated representations of mobility, and we have observed them in the 2D histograms of the aggregated flows in Fig. 5.2.13.

As shown in Fig. 5.2.9, spatial time series can be viewed from two complementary perspectives: as spatially distributed local time series and as temporally varying spatial situations. Fig. 5.2.13 corresponds to the former perspective: we applied cluster analysis to the local time series associated with the links. Now we are going to take the other perspective and apply clustering to the time steps of the time series. We cluster the time steps according to the similarity of the spatial distributions of the car presence (Figs. 5.2.14 and 5.2.15) and flow volumes (Figs. 5.2.16 and 5.2.17). The aggregates representing the presence have been obtained from the original (undivided) trajectories, to take stationary vehicles into account, and the link-based aggregates have been obtained from the divided trajectories representing the trips.
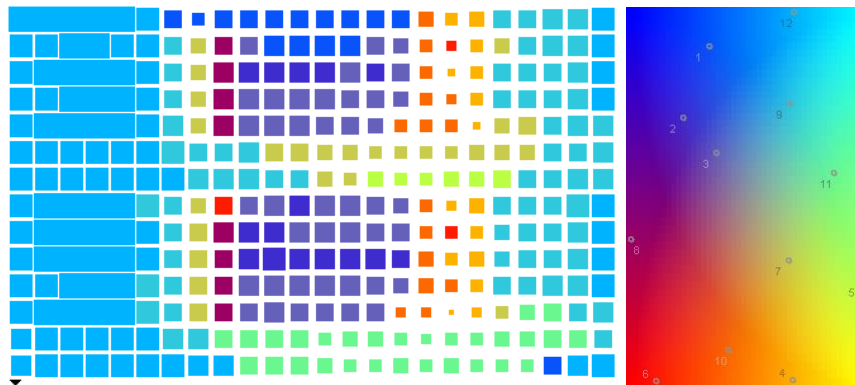

Fig. 5.2.14. Left: a calendar display of the clusters of the hourly time steps according to the distribution of the car presence over the set of places. The columns correspond to 24 hours of the day and the rows to the 14 days from Monday (top) to Sunday of the next week (bottom). The colors correspond to different clusters, and the sizes of the colored rectangles represent the closeness of the cluster members to the cluster centroids (the closer, the bigger). Right: the colors for the clusters have been chosen by projecting the cluster centroids onto a continuously colored plane.

The calendar view in Fig. 14, left, shows the daily and weekly patterns of the spatial distribution of the car presence, where the night hours are similar across the days, the morning and evening rush hours of the weekdays appear quite different from the mid-day times, and the weekend patterns are distinct from the weekday ones. The patterns on Friday evenings differ from the other weekdays by later beginnings of the evening- and night-specific distributions.

The small multiple maps in Fig. 5.2.15 demonstrate the spatial distribution of the mean volumes of the presence for each cluster. The clusters are arranged according to the succession of their numeric labels (from 1 to 12) in rows from left to right and from top to bottom. We can observe extremely prominent road network patterns, especially during the mass commuting times (e.g., clusters 6 and 10). These patterns do not appear in late evenings and nights (clusters 9 and 12).
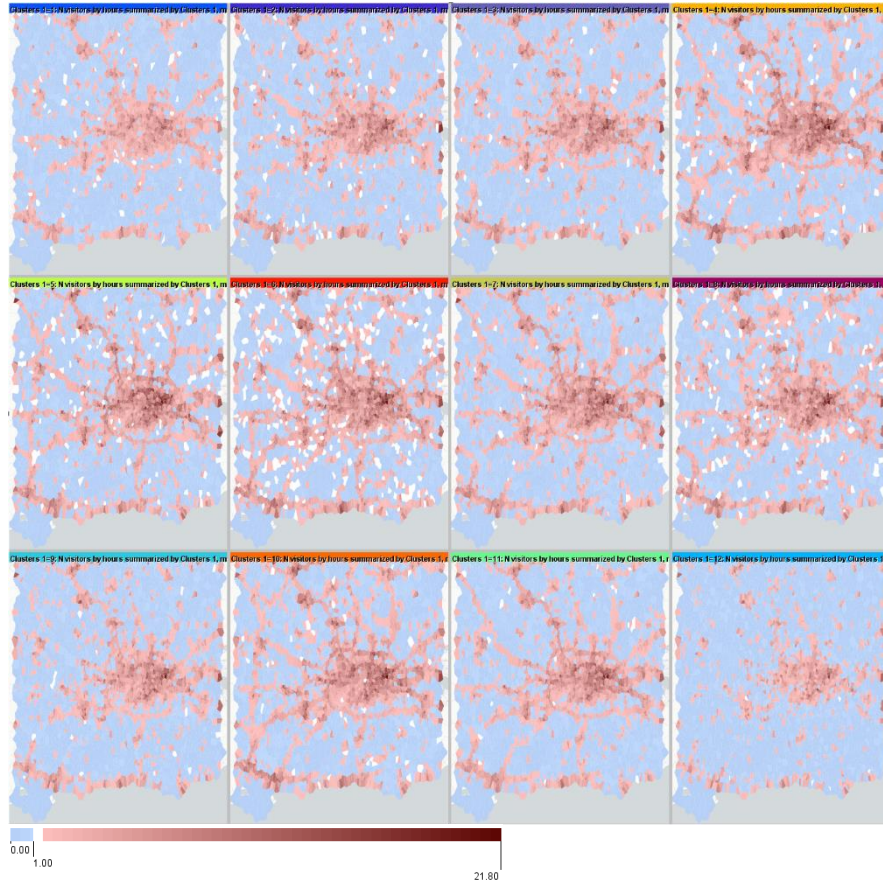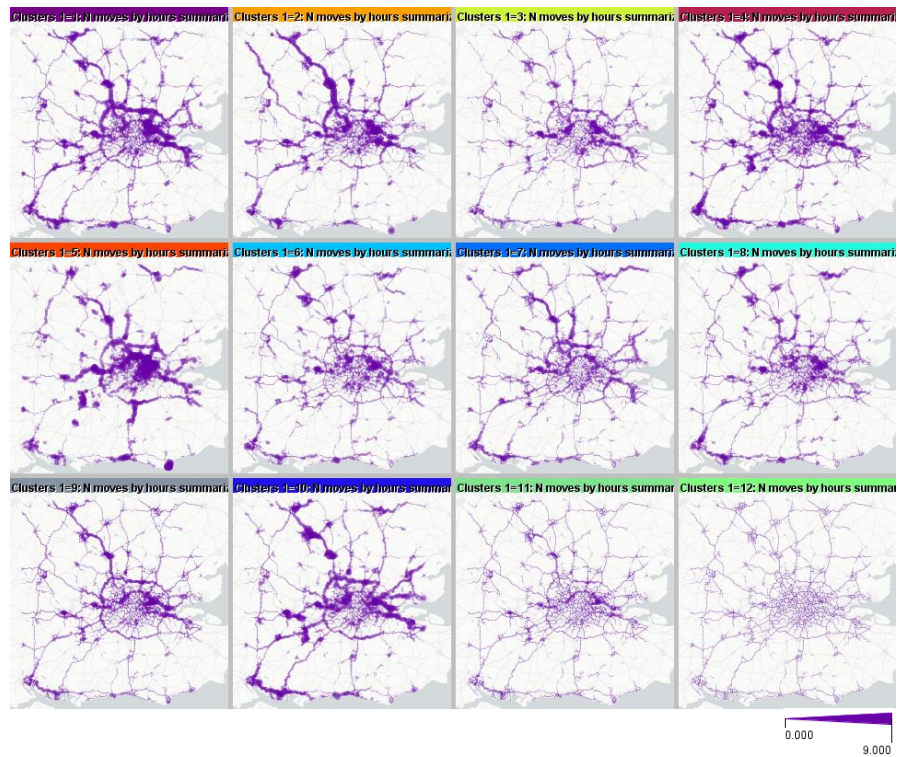

Fig. 5.2.15. The average spatial distributions of the car presence for the time clusters presented in Fig. 5.2.14. The mean car counts are represented by the darkness of the shades of red while light blue corresponds to zero values.

Figures 5.2.16 and 5.2.17 present the results of applying clustering to the time steps of the link-based time series. The times have been clustered according to the similarity of the spatial distributions of the flow volumes. Figure 5.2.16 is analogous to Fig. 5.2.14, and Figure 5.2.17 corresponds to Fig. 5.2.15, but the maps here show the spatial distributions of the mean flow volumes corresponding to the clusters. The volumes are represented by proportional widths of the flow lines.
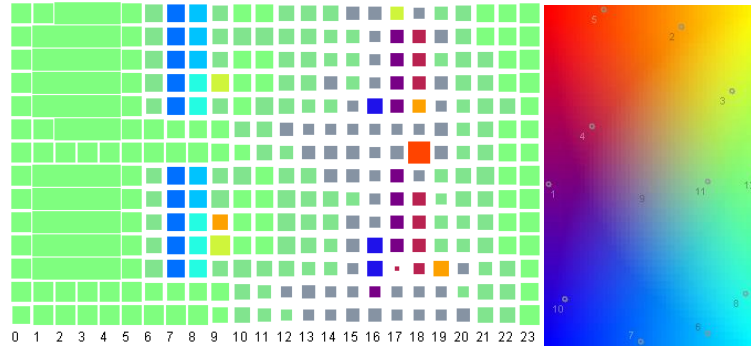
Fig. 5.2.16. Clusters of the hourly time steps according to the spatial distributions of the flow volumes. The representation is analogous to Fig. 5.2.14.

Fig. 5.2.17. The maps show the spatial distributions of the flow volumes, represented by proportional line widths, for the clusters shown in Fig. 5.2.16.

The afternoon clusters 1, 4, and 9 are characterized by intensive traffic on highways while the morning clusters 6, 7, and 8 show higher traffic on local roads and in populated areas. Interestingly, the flow distribution patterns in hours 9 to 14 on

the weekdays are similar to those in the nights. Several clusters consist of only a few or even a single time moment with extraordinary traffic distributions. For example, cluster 5 has a very high traffic on the inner ring of London.

5.2.6 Specifics of episodic movement data

Depending on the temporal resolution and sampling regularity, movement data can be categorized as quasi-continuous or episodic (Andrienko et al. 2013a). The example data used in this chapter can be ascribed to the former category, because the time intervals between the records are quite small and mostly of the same length. In episodic movement data, position measurements may be separated by large time gaps, in which the positions of the moving objects are unknown and cannot be reliably reconstructed. Such data require special approaches to analysis. Thus, like with quasi-continuous data, it is possible to aggregate episodic trajectories to flows between places. However, consecutive positions of a trajectory may fit in non-neighbouring places. Flow maps constructed from episodic trajectories are typically extremely cluttered due to a large number of intersecting flow lines connecting distant places. Moreover, time intervals between consecutive positions may be longer than the time intervals chosen for aggregation. Such trajectory segments must be ignored. It is also not possible to estimate the number of moving objects that were present in a place during a time interval because the exact times of coming to a place and leaving it are unknown.

In interpreting flow maps built from episodic movement data, analysts should keep in mind that they do not represent all movements that really happened. Nevertheless, such flow maps can be useful since there is a chance that mass movements or sufficiently frequent movement patterns can be adequately reflected.

As an example of episodic movement data, Figure 5.2.18 demonstrates 11,671 trajectories reconstructed from georeferenced posts of social media (Twitter) users. Each trajectory consists of a chronological sequence of posts of one user. Similar trajectories can be constructed from data about mobile phone activities, including making calls, sending messages, and accessing Internet.

In Fig. 5.2.18, the locations of the social media posts are connected by lines, which are drawn with 97% transparency. Long lines mean unknown users' paths between the locations of their consecutive posts. In this dataset, which spans over a 28-days period in September, the median time interval between records of the same user is 14 minutes, the third quartile is about three hours, and the maximum is over 24 days. However, in most cases, the distances between the points are small, the third quartile being only 0.26 km. This means that people tend to make repeated posts from the same or nearly the same locations, which are, possibly, repeatedly visited.
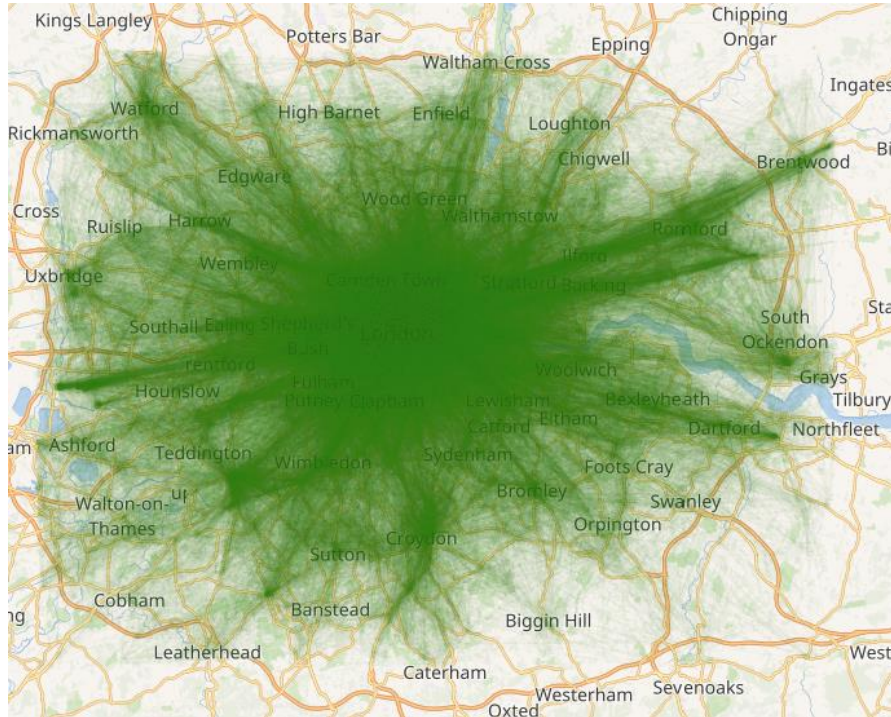
Fig. 5.2.18. Episodic trajectories reconstructed from georeferenced posts of social media users.

Despite all uncertainties, episodic trajectories reconstructed from social media posts or mobile phone use registers can provide valuable information about mobility behaviors of people. Unlike trajectories of personal cars, taxis, or any particular kind of vehicles, these trajectories can reflect movements made with the use of diverse transportation modes. However, because of the uncertainties and inherent biases, such data need to be used cautiously as a complement to other mobility data rather than alone.

As we mentioned, special care needs to be taken in aggregation of episodic movement data. In our example, we partition the territory into spatial compartments using the method described in Section 5.2.5.1, i.e., the same as we used for the vehicle trajectories. We want to aggregate the data by hourly time intervals; therefore, we split the trajectories into trips by time gaps longer than one hour. This means that, when the time interval between two points exceeds one hour, the later point is treated as the beginning of a new trip. Hence, the transition between the points is not used in the aggregation. Additionally, we split the trajectories by spatial gaps of more than 5 km, which is the average radius of a spatial compartment used for the aggregation. The flow map resulting from the aggregation is shown in Fig. 5.2.19. It reveals the importance of the central area of London for people's mobility: not only the major flows occurred in the center, but also there were relatively many

radial movements to and from the central area. Besides, we can see "hubs", such as Camden Town and Wimbledon, with star-like patterns of flows around them.
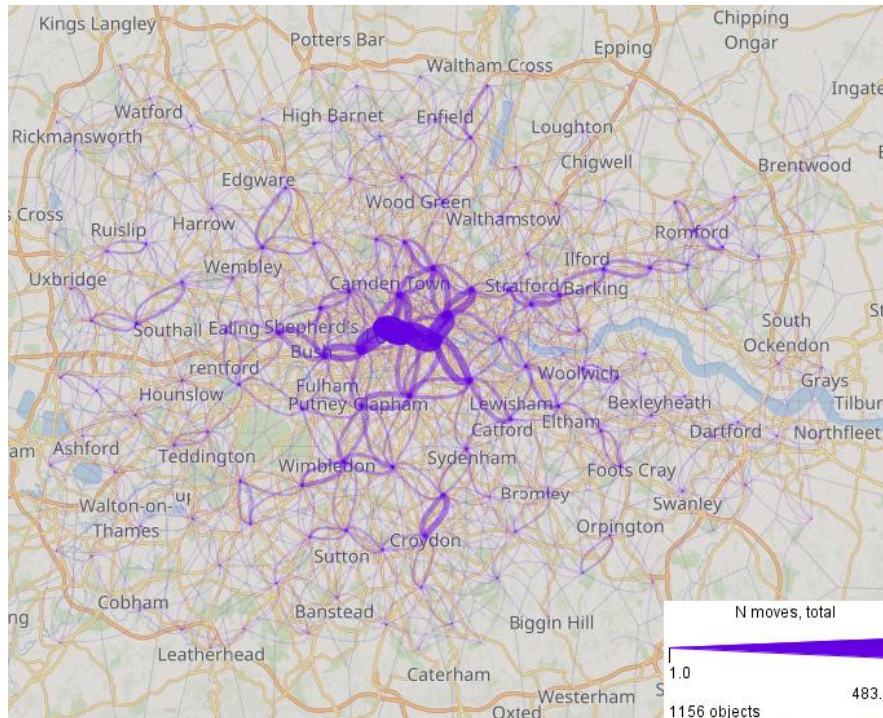
Fig. 5.2.19. Aggregated movements of social media users.

Figure 5.2.20, left, demonstrates the temporal distribution of the aggregated movements of the social media users. In this two-dimensional temporal histogram, the rows correspond to the days, columns to the hours of a day, and the sizes of the squares are proportional to the numbers of moves made in the corresponding hourly intervals. Prominent patterns of more intensive movements in morning hours of the weekdays, with peaks at hour 9, are clearly visible. Many movements also happen in the late afternoons and evenings of the weekdays, while on the weekends the movements are more uniformly distributed over a day starting from late morning. Interestingly, this temporal distribution differs from the temporal distribution of the counts of the posted messages shown on the right of Fig. 5.2.20.
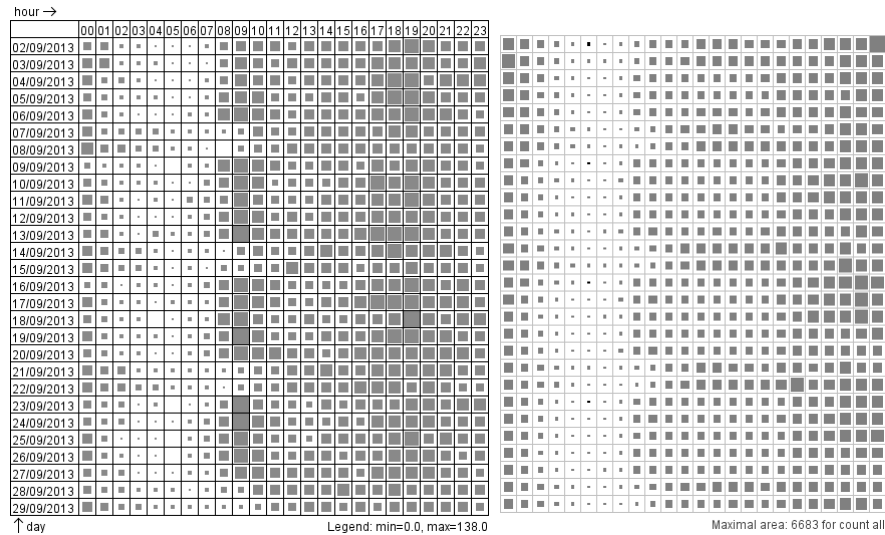
Fig. 5.2.20. Temporal patterns of the aggregated moves of the social media users (left) are compared with the temporal patterns of the number of posted messages (right). The rows correspond to the days, columns to the hours of a day, and the sizes of the squares are proportional to the numbers of moves or messages, respectively.

This example shows that the approaches presented in this chapter are not specific to GPS tracks of vehicles but can be applied to other kinds of spatio-temporal data collected in various ways. However, the ways of data collection and the properties of the data need to be carefully taken into account in data transformation, analysis, and interpretation of visual displays and computation results.

5.2.7 Discussion and conclusions

Our examples demonstrate how three major aspects of the urban context – places, flows, and times – can be characterized using trajectory data. We proposed methods to define a suitable set of places, aggregate trajectories into place- and link-based time series, and characterize the places, flows, and times taking two complementary perspectives in analyzing the time series. We demonstrated the use of methods of cluster analysis as a means of abstraction and as an aid in coping with large data volumes. Particularly, we showed that clustering by similarity can be applied to local time series, for characterizing places and links, and to spatial distributions, for characterizing times.

Due to the page limit, we shall only briefly outline the potential directions for extraction of further context information from trajectory data. One possibility is to consider attributes along trajectories, such as (Andrienko et al, 2013b):

[Type here]

- measured values, e.g., instant speed and direction, acceleration, turn, fuel consumption, CO_2 emission, etc.;
- spatial context, e.g., road type, land use, distances to stationary objects such as gas stations or other places of interest;
- derived from sequences of positions of the same trajectory, e.g. computed speed and direction, curvature of the travelled path in a sliding time window;
- computed based on trajectories of co-moving objects, e.g., count of trajectories in given space- and time-windows or distance to Nth closest neighbor.

Acquired attributes can be aggregated by places, flows, or along trajectories, enabling selection of locations, connections, or vehicles with particular features. Such vehicles can be visualized on a trajectory wall (Tominsky et al, 2012).

Trajectory attributes can be used for identifying locations that are characterized by particular properties. Thus, density-based clustering of trajectory segments characterized by slow movement can be used for identifying locations of traffic jams and revealing their dynamics (Andrienko et al, 2013b). Scalable methods are developed for identifying hotspots from big data (Nikitopoulos et al. 2019). Considering the parts of trajectories preceding traffic jams, one can study the traffic jam propagation over the street network (Wang et al 2013).

Methods for time series analysis and modeling can be applied to place- or link-based local time series that have been clustered by similarity. The resulting models can be used for predicting traffic characteristics depending on time. Besides, link-based time series of flow volumes and average movement speeds not only can be modelled in separation but also used for representing and modelling the speed-volume dependencies as proposed by Andrienko and Andrienko (2013b). Such models can be utilized for simulation of regular and extraordinary traffic (Andrienko et al 2016b) or for billboard pricing and informed decision making (Liu et al 2017)

Division of trajectories into trips allows extraction of routine movement behaviors (Rinzivillo et al. 2014) and semantic interpretation of locations (Andrienko et al. 2016a). Analysis of semantically-annotated trajectory data (e.g. by state transition graphs, Andrienko and Andrienko 2018) allows finding important behavior patterns without compromising personal privacy.

Our study demonstrates that visual analytics approaches and techniques can support sophisticated analyses for gaining understanding of complex phenomena, such as urban mobility, which is necessary for building explainable models and making informed substantiated decisions. However, we see a need in further advances in visual analytics research and technical developments in the following major directions:

- Stronger support of joint analysis of multiple datasets of diverse structure and quality.
- Dealing with streaming data that are constantly generated and updated.
- More specific approaches for supporting decision making, including development, evaluation, and comparison of decision options and performing what-if scenarios.

References

Andrienko G, Andrienko N, Bak P, Keim D, Wrobel S (2013a) Visual Analytics of Movement, Springer

Andrienko G, Andrienko N, Hurter C, Rinzivillo S, Wrobel S (2013b) Scalable analysis of movement data for extracting and exploring significant places. IEEE transactions on visualization and computer graphics 19(7):1078-1094. https://doi.org/10.1109/TVCG.2012.311

Andrienko G, Andrienko N, Fuchs G (2016) Understanding movement data quality. journal of location based services, 10(1):31-46. https://dx.doi.org/ 10.1080/17489725.2016.1169322

Andrienko G, Andrienko N, Chen W, Maciejewski R, Zhao Y (2017) Visual analytics of mobility and transportation: state of the art and further research directions. IEEE Transactions on Intelligent Transportation Systems, 18(8): 2232-2249. http://dx.doi.org/10.1109/TITS.2017.2683539

Andrienko N, Andrienko G (2011) Spatial generalization and aggregation of massive movement data. IEEE Transactions on Visualization and Computer Graphics, 17(2):205-219. http://doi.ieeecomputersociety.org/10.1109/TVCG.2010.44

Andrienko N, Andrienko G (2013a) Visual analytics of movement: a rich palette of techniques to enable understanding. In: Renso C, Spaccapietra S, Zimanyi E (Eds.) Mobility Data: Modeling, Management and Understanding. Cambridge University Press:149-173

Andrienko N, Andrienko G (2013b) A visual analytics framework for spatio-temporal analysis and modelling. Data Mining and Knowledge Discovery, 27(1): 55-83. http://dx.doi.org/10.1007/s10618-012-0285-7

Andrienko N, Andrienko G, Fuchs G, Jankowski P (2016a) Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. Information Visualization. 15(2):117-153. https://doi.org/10.1177/1473871615581216

Andrienko N, Andrienko G, Rinzivillo S (2016b) Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics. Information Systems, 57(1):172-194. http://dx.doi.org/10.1016/j.is.2015.08.007

Andrienko N, Andrienko G (2018) State transition graphs for semantic analysis of movement behaviours. Information Visualization, 17(1):41-65. http://dx.doi.org/10.1177/1473871617692841

Batty M (2013) The New Science of Cities. MIT Press

Beecham R, Wood J (2014) Exploring gendered cycling behaviours within a large-scale behavioural dataset. Transportation Planning and Technology. 37(1):83-97. https://doi.org/10.1080/03081060.2013.844903

Hägerstrand T (1970) What about people in regional science? Papers of the Regional Science Association, 24:7-21.

Itoh M, Yokoyama D, Toyoda M, Tomita Y, Kawamura S, Kitsuregawa M (2016) Visual exploration of changes in passenger flows and tweets on mega-city metro network. IEEE Trans. Big Data, 2(1): 85–99. https://doi.org/10.1109/TBDATA.2016.2546301

Kesting A, Treiber M (2013) Traffic Flow Dynamics: Data, Models and Simulation. Springer

Kraak MJ (2003) The space-time cube revisited from a geovisualization perspective. In Proceedings of the 21st International Cartographic Conference, Durban, 10 - 16 August 2003. International Cartographic Association:1988-1995

Lansley G, Longley P (2016) The geography of twitter topics in London, Computers, Environment and Urban Systems, 58:85-96. https://doi.org/10.1016/j.compenvurbsys.2016.04.002

Liu D, Weng D, Li Y, Bao J, Zheng Y, Qu H, Wu Y (2017) SmartAdP: Visual analytics of large-scale taxi trajectories for selecting billboard locations. IEEE Transactions on visualization and Computer Graphics 23(1):1–10. https://doi.org/10.1109/TVCG.2016.2598432

Nikitopoulos P, Paraskevopoulos AI, Doulkeridis C, Pelekis N, Theodoridis Y (2018) Hot spot analysis over big trajectory data. IEEE International Conference on Big Data. https://doi.org/10.1109/BigData.2018.8622376

Markovic N, Sekula P, Vander Laan Z, Andrienko G, Andrienko N (2019) Applications of trajectory data from the perspective of a road transportation agency: literature review and maryland case study. IEEE Transactions on Intelligent Transportation Systems (accepted). http://dx.doi.org/10.1109/TITS.2018.2843298

Rinzivillo S, Gabrielli L, Nanni M, Pappalardo L, Pedreschi D, Giannotti F (2014) The purpose of motion: Learning activities from Individual Mobility Networks. International Conference on Data Science and Advanced Analytics. https://doi.org/10.1109/DSAA.2014.7058090

Thomas J, Cook K (2005) Illuminating the Path: The research and Development Agenda for Visual Analytics. IEEE

Tominski C, Schumann H, Andrienko G, Andrienko N (2012) Stacking-based visualization of trajectory attribute data. IEEE Transactions on Visualization and Computer Graphics. 18(12):2565-2574. https://doi.org/10.1109/TVCG.2012.265

Wang Z, Lu M, Yuan X, Zhang J, van de Wetering H (2013) Visual traffic jam analysis based on trajectory data. IEEE Transactions on visualization and Computer Graphics. 19(12):2159-2168. https://doi.org/10.1109/TVCG.2013.228