J geovis spat anal (2021) 5:3
https://doi.org/10.1007/s41651-020-00068-1

SparseTrajAnalytics: an Interactive Visual Analytics System for Sparse Trajectory Data

Xinyue Ye[1] · Jiaxin Du[1] · Xi Gong[2] · Ye Zhao[3] · Shamal AL-Dohuki[4] · Farah Kamw[5]

Accepted: 10 November 2020

Abstract
Sparse trajectory data are trajectories that cover a relatively large geographic area with infrequent samplings of movements, such as human migration and hurricane trajectories. Compared with dense trajectory data (e.g., massive vehicle movements in a metropolitan area), sparse trajectories represent a more general form of movement data due to the data availability and less sensitivity. However, there is a lack of open source software for sparse trajectory data analytics. We designed an online system with a self-adaptive heatmap and multi-spatial-temporal-view charts to analyze sparse trajectories both visually and interactively. With our system, users can (1) store and manage sparse trajectory data on a cloud service; (2) analyze sparse movement behaviors by querying consolidated and live data sets using keywords, spatial location, and time constraints; and (3) explore query results and associated data through web-based thematic maps, tables, and charts. Results of the survey among domain experts show that the open source system is intuitive and user-friendly.

Keywords Sparse trajectories · Visual analytics · Human migration · Hurricane trajectories

✉ Xi Gong
xigong@unm.edu

Xinyue Ye
xinyue.ye@tamu.edu

Jiaxin Du
jiaxin.du@tamu.edu

Ye Zhao
zhao@cs.kent.edu

Shamal AL-Dohuki
saldohuk@uod.ac

Farah Kamw
farah.kamw@cuaa.edu

1 Department of Landscape Architecture & Urban Planning, Texas A&M University, College Station, TX, USA

2 Department of Geography & Environmental Studies, University of New Mexico, Albuquerque, NM, USA

3 Department of Computer Science, Kent State University, Kent, OH, USA

4 Department of Computer Science, University of Duhok, Zakho way, Duhok, Iraq

5 Department of Computer Science, Concordia University Ann Arbor, Ann Arbor, MI, USA

Introduction

A trajectory is composed of a group of geographic locations between its origin and destination with associated timestamps. Based on sampling frequencies and geographic coverage, we can categorize trajectories into sparse trajectory data and dense trajectory data. Sparse trajectory data are trajectories that cover a relatively large geographic area with infrequent samplings of movements, such as human migration and hurricane trajectories; while dense trajectory data consist of sample points that are collected more frequently in a relatively small geographic area, e.g., taxi trajectories in a city (Table 1). Compared to dense trajectory data, sparse trajectory data have fewer sample points that are scattered in a larger area during a given time period.

Analyzing sparse trajectory data could help researchers better understand the geographic phenomenon in larger spatial-temporal scales, which could in turn contribute to better decision-makings. Human migration trajectories are typical sparse trajectory data documenting the movements of people (Willekens et al. 2016). The data come from censuses, population registries, administrative data collections, and national sample surveys (Mather and Cortes 2016). One hundred eighty-three out of 193 United Nations countries have migration data available for

Published online: 07 January 2021

Table 1 Differences between dense trajectory data and sparse trajectory data

Categories	Dense trajectory data	Sparse trajectory data
Sampling time interval	< 30 s	> 30 s
Spatial distribution	Citywide	From citywide to nationwide
Examples	Taxi trajectory	Migration data, hurricane trajectory data

analysis (Bell et al. 2015). Integrating migration data with extensive coverage of geographic features could reveal the spatial-temporal human dynamics. This approach is widely used in studies related to economic development and transportation (Skeldon 2012). Hurricane trajectories are another type of commonly used sparse trajectory data, which track hurricanes in large spatial-temporal scales (Landsea and Cangialosi 2018). The hurricane trajectory data have been recorded by the National Oceanic and Atmospheric Administration of the USA (NOAA) since 1851 at the global level (Landsea and Franklin 2013). Investigating the historical hurricane trajectory data could provide insights of the hurricane trends (Weinkle et al. 2018). The knowledge could help researchers predict the hurricane movements to a certain degree, which would be very beneficial in reducing and mitigating the possible severe impacts on the communities along the hurricane trajectories (Landsea and Cangialosi 2018; Zheng 2015).

Sparse trajectories are usually stored as large amounts of tabular data, which would easily overwhelm the data users with limited computing capability. Simply providing descriptive statistics of the sparse trajectory data summarized in a predefined time interval (e.g., yearly) and spatial scale will not offer adequate information about the movement patterns. To gain deeper insights into the movement phenomenon, researchers need to conduct iterative explorations using their domain knowledge. Interactive visualization is a key component in the process. However, visualizing sparse trajectory data is technically challenging in the following aspects. First, because sparse trajectories are unevenly distributed across space and time (Meng et al. 2016), they cannot be visualized in a fixed spatial-temporal scale. Second, since the trajectories are sparsely distributed across a large area, simply connecting the sample points with straight lines may be misleading (Du et al. 2020) and lacks cartographic aesthetics. Last, sparse trajectories cover larger geographic area and longer time period than dense trajectories, so more geographically referenced data need be retrieved to analyze sparse trajectories. Hence, it poses a computing challenge to the robustness of an online system. Even simple operations, such as smoothly displaying the heatmaps of thousands of sparse trajectories with frequent zoom-in/out of regions and changes of time periods, might collapse the system. Such operations expect spatial and temporal data aggregations and visualizations with random access patterns, which need to adopt advanced computational technologies. Therefore, visualization tools for general trajectories could not address these challenges. To visualize sparse trajectory data in an efficient and interactive way, intuitive tools specifically designed for sparse trajectories are needed.

In general, an ideal visual analytic software system for sparse trajectories should offer the following capabilities: (1) a powerful computing platform for domain researchers to accomplish their tasks over most commonly owned devices such as desktop computers for office use; (2) an easy-access gateway for trajectory retrieval, analysis, visualization, and sharing; (3) a scalable data storage and management system supporting a variety of data queries with immediate responses; and (4) an exploratory visual interface which is informative, intuitive, and interactive (AL-Dohuki et al. 2019).

Based on the open source visual analytics tool for intensive movements in metropolitan areas (TrajAnalytics) (Zhao et al. 2017), we developed SparseTrajAnalytics, a sparse trajectory visual analytics system. The highlights of the system are:

1. **Optimized performance for sparse trajectory data.** The back-end of the system is optimized based on characteristics of sparse trajectories. Spatial and temporal indexes are integrated into the data preprocessing for faster data query and integration performance.
2. **Self-adaptive heatmap.** A self-adaptive color rendering algorithm is designed to present sparse trajectories as a heatmap. The heatmap can clearly visualize the relative density differences of trajectory points in various view scales, even when the point densities are low.
3. **Interactive multi-spatial-temporal-view charts.** Panels in the multi-spatial-temporal-view charts are interrelated and provide multiple views of the same data at different spatial-temporal scales simultaneously. So users can easily explore the sparse trajectory patterns visually and interactively.

The open source system provides tools and a web platform for researchers and policy-makers to identify and understand movement patterns of research objects in their interested area through visually analyzing their sparse trajectories. The codes and sample data of our open source system are available at https://github.com/dujiaxin/TrajVis. If users are concerned about the data security of sharing trajectory data on a public cloud service, they can download the source codes and deploy the system in their local environment.

Related Works

Trajectory data visualization is widely explored in not only computer science and geo-spatial science research but also policy-making and business intelligence. In order to display the spatial dynamics of people and vehicles intuitively, a variety of tools have been developed to visualize trajectory data. Among them, programming libraries (e.g., D3.js (Bostock et al. 2011)) and advanced systems for data analysis (e.g., Microsoft Excel, Tableau, and ArcGIS (Basiri et al. 2016)) provide flexible data preprocessing and visualization functionalities for various data types. However, these software packages are not specifically designed for visualizing trajectory data. Creating customized graphics and charts for trajectory data requires substantial training of certain program languages or a good command of the abundant functionalities in a full-featured graphical workspace (Xia et al. 2018). Trading flexibility for simplicity, many online systems (e.g., Mapbox (Cadenas 2014)) enable non-expert users to visualize their trajectory data in just a few steps (e.g., loading data, selecting a template, and specifying geolocation and timestamp columns to display). However, the functionalities of these applications are limited to general visualizations. The integration of geocoding and spatio-temporary query are usually limited or not supported.

Standalone software systems that focus on trajectory data are also available, e.g., GeoTime (Kapler and Wright 1997), TripVista (Guo et al. 2011), and FromDaDy (Hurter et al. 2009). Such software provides general user interfaces to load trajectory data and visualize the spatial-temporal information. Many of them focus on origins and destinations of trajectories, such as flow maps (Phan et al. 2005), Flowstrates (Boyandin et al. 2011), OD maps (Wood et al. 2010), and TaxiVis (Ferreira et al. 2013). Some systems visualize trajectories using various visual metaphors and interactions, such as vessel movement (Willems et al. 2009), route diversity (Liu et al. 2011), and meteorologic areas (Von Landesberger et al. 2012). Other software systems integrate both multidimensional visualization and map views. For example, the Polorhub (Li et al. 2016) indexes and visualizes data from the temporal, spatial, and topical dimensions. However, most of these existing tools could not address the challenges for sparse trajectory visual analytics mentioned in the previous section.

TrajAnalytics (Zhao et al. 2016) is an open software designed to manage and visualize massive trajectory data. It was designed to match dense trajectory points to roads or grids at the city scale with powerful visualization, so its functionalities work perfectly with a large amount of trajectory data within relatively small area. However, it has limited ability to visualize sparse trajectory data at a much larger spatial-temporal scale. Geocoding trajectory points to zipcode regions across the USA is hardly possible because of the complexity of the iterative matching process. Spatial and temporal scales for sparse trajectories used in different literature are also very different, so we cannot assume the density of data in advance and summarize sparse trajectory data to a single scale, such as weekdays or hours. Therefore, an updated TrajAnalytics system is needed for sparse trajectory visual analytics.

SparseTrajAnalytics System

We implemented the updated TrajAnalytics as a web-based visualization system, because web-based systems are more flexible and require virtually no setup process by end-users (Coscia et al. 2012). The SparseTrajAnalytics system (Fig. 1) includes two major parts. The back-end runs on a server and provides data services, including data pre-processing, geocoding, and data query; and the front-end runs on web browsers to provide interactive visualization for users. The developing and testing of the system were conducted on a personal laptop (Intel® Core™ i7-5600U, 16GB RAM, 500GB SSD disk).

Sample Data Sets for Demonstration

To demonstrate the back-end service performance and the front-end visualization, two sample data sets were used, including a migration data set and a hurricane data set.

The migration data set was derived from the survey data collected by the Survey Research Lab (Yu and Shaw 2013), Bureau of Business and Economic Research at The University of Montana–Missoula. It contains 100 person's migration trajectories during 1940–2001. The original file is stored as a personal geodatabase (.mdb file). We extracted only key information about migration sample points (person ID, year, latitude, and longitude) from the original data (Table 2). The person ID indicates whom the trajectory sample point belongs to.

The hurricane data set is the Atlantic hurricane database (HURDAT2) for year 1851–2019 from the NOAA (Landsea and Franklin 2013), which was updated on 28 April 2019. This data set has 6-hourly information about the locations, maximum winds, central pressures, and sizes (starting from 2004) of all known tropical cyclones and subtropical cyclones. Table 3 shows a few sample records in the data

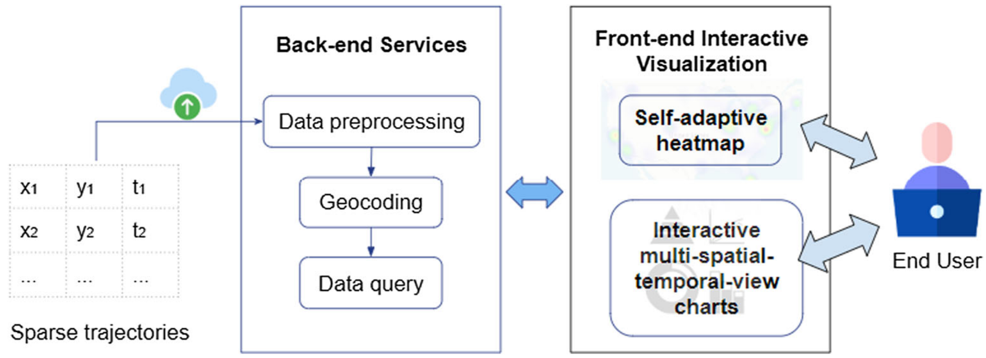


Fig. 1 Basic architecture of the SparseTrajAnalytics system

set. The ID is the unique identifier for a hurricane. The time field contains both the date and hour level information.

Back-end Services

The back-end does the heavy lifting in the system, which includes uploading sparse trajectory data from users, preprocessing the data, and aggregating the data for spatial-temporal queries. There are three major modules in the back-end, including the data preprocessing module, the geocoding module, and the data query module. The back-end is cross-platform and can be deployed on any machine with PHP, Python, and PostgreSQL.

Data Preprocessing

To use our system, users first need to obtain an account on our website. Then, users can upload their data files and define metadata by specifying the meaning of each data column. This software allows the users to upload their raw trajectory data set in comma-separated values (.csv) format. Each data record represents one sample point, which consists of trip ID, geo-location (longitude and latitude), timestamp, and speed (optional). The trip ID indicates which trip this sample point belongs to.

In the data preprocessing module, our system automatically detects missing values in user uploaded data. A record is discarded if its essential fields are missing, such as longitude, latitude, or timestamp. The remaining records are used to create a trip data (TD) table. Trips in the TD table are created by aggregating the sample points of each trip in chronological order based on the trip ID. Three types of indexes are added to the TD table to improve the performance of data queries. The indexes are as follows: (1) B-TREE index over the trip ID which could improve the search performance by trip ID, (2) Generalized Search Tree (GiST) index over the trip geometrical points and linestrings which could improve the performance of spatial queries, and (3) Generalized Inverted Indexes (GIN) index over all time related fields which could improve the performance of trips data retrieval and aggregation queries. The module stores data in a database and builds indexes on the data, so that the system can respond in near real time. In our server, we use the PostgreSQL, an open source database, with its spatial extension named PostGIS (Obe and Hsu 2011) for data storage. The three indexes are added using the built-in implementations of them in the PostGIS and PostgreSQL. The B-TREE and GIN are natively supported by the PostgreSQL. The PostGIS extension also has built-in functions to create spatial indexes, which only need users to specify the type of spatial index and the geometry column in the database. Once the spatial index is created, it will be utilized automatically in spatial queries.

Geocoding

The geocoding module matches trajectory data to zip code regions or user-defined grid regions. The zip code level trajectory data can be overlaid with other data collected by government agencies for further analysis. The zip code boundary data are stored in our system in advance. Users

Table 2 Migration data samples

Person ID	Year	Latitude	Longitude
80027	1962	34.285	−118.261
80027	1969	36.120	−121.059
80148	1933	47.153	−123.753
80148	1940	47.447	−121.716
80148	1945	47.529	−122.698

Table 3 Hurricane data samples (sample records of Tropical Storm Alberto in 2018)

ID	Time	Latitude	Longitude
AL012018	20180525, 1200	18.8N	87.1W
AL012018	20180525, 1800	18.7N	86.5W
AL012018	20180526, 0000	18.9N	85.9W
AL012018	20180526, 0600	19.6N	85.7W
AL012018	20180526, 1200	21.3N	85.6W

can also divide the study area into grids with arbitrary resolutions (e.g., 0.5 km or 100 km). Then, the system automatically matches the loaded trajectory sample points to these regions. This grid region related function is inherited from AL-Dohuki et al. (2019), which enables flexible spatial scale of the study. The reference map data are collected from OpenStreetMap using its online application programming interfaces (APIs).

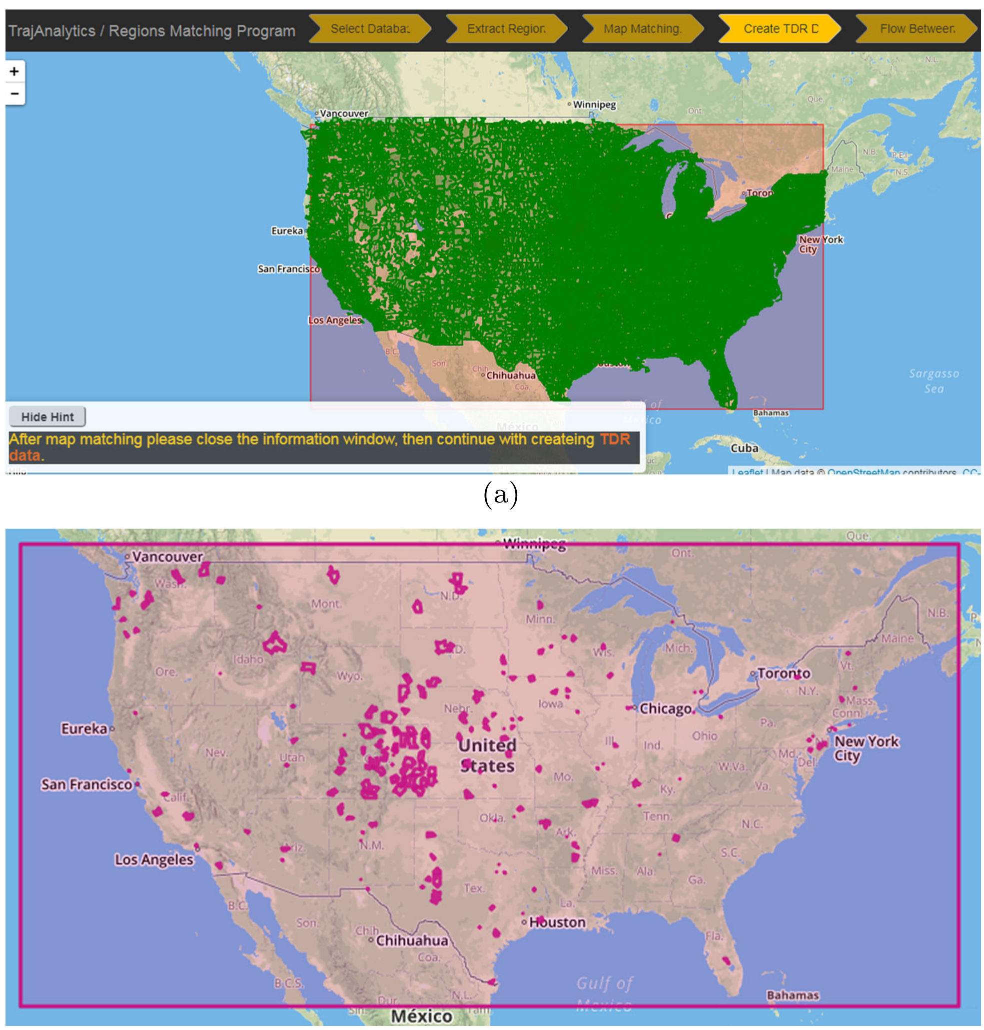
It is almost impossible for the old version of the TrajAnalytics to conduct the geocoding task with nationwide zip code regions. Because the system implemented a naive iteration matching process that needs to join every trajectory points with all zip code regions, it has a limit of 100 km^2 for the geocoding module. In our experiment with the sample migration data sets, the old system crashed after running geocoding task for 2 h. However, because the GiST index has been built over regions and trajectories in the SparseTrajAnalytics system, the geocoding process can use the GiST index to efficiently query the nationwide migration and hurricane sample data within a minute. Thus, there is no limit on the spatial coverage of a geocoding task, as shown in Fig. 2.

Data Query

The data query module supports spatial-temporal queries on trajectory data stored on a cloud service.

$$Q(G, Ts, Te, E, t) \Rightarrow \psi \tag{1}$$

Equation 1 shows a typical query, where Q denotes the query; ψ is the query result; E contains all the trajectory points; G represents the query region(s); and Ts and Te are the start time and end time of the temporal constraint. Additionally, t represents spatial relationships between trajectories and a given region in the query, where four options are available, including (1) passing the region, (2) starting from the region, (3) ending at the region, and (4) completely within the region. Such kind of query Q will return a group of trajectories ψ based on their spatial-



(a)

(b)

Fig. 2 Geocoding nationwide migration data: (a) retrieving zip code regions from database; (b) matching sample data to zip code regions temporal relationships with the query geographic region(s) G and the time interval (Ts to Te), e.g., identifying all trajectories that passed through Colorado during 2010–2015.

To facilitate interactive visual examination of massive trajectory data, the data query module supports two basic types of trajectory aggregation operations: temporal and spatial aggregation. Temporal aggregation summarizes trips by different time scales (years to hours) to generate a set of attributes describing the aggregated trips. Spatial aggregation generates summarized attributes by aggregating trips over road networks, regions, or certain areas of interest.

Front-end Interactive Visualization

Taking the differences between dense trajectories and sparse trajectories into account, we designed a visualization interface that can display both types of data in a meaningful way. The front-end is built based on the open source TrajAnalytics System (Zhao et al. 2017; Zhao et al. 2016). While the basic UI layout is inherited from the previous version, new functionalities, including the self-adaptive heatmap and interactive multi-spatial-temporal-view charts, are added to the new system. The front-end is a dynamic web page constructed with HTML and JavaScript. JavaScript framework leaflet.js (Leaflet 2015) is used for the core mapping functionality. Other frameworks, such as Heatmap-js and D3.js, are used for drawing heatmap and charts, respectively. Python and PHP scripts are served as the interface between the back-end and the front-end for data exchange. The design details and demonstrations of the new functionalities are described as follows.

Self-adaptive Heatmap

A heatmap can visualize magnitudes of a phenomenon as a two-dimensional color graph (Fig. 3b). It can be used to demonstrate "where research objects are most likely to move to or leave from" in trajectory analyses. The heatmap in the previous TrajAnalytics system used fixed color ramp, which is not suitable for displaying low density data. As shown in Fig. 3a, because sample points of sparse trajectory data are distributed over a large geographic area in a very low density, the heatmap could not clearly differentiate the areas with and without trajectory points. In order to show the relative density differences clearly, we redesigned the color rendering algorithm of the heatmap.

First, we calculate the density of the trajectory data, which is defined as the number of trajectory points in a screen grid cell. The grid cell size is dynamically calculated based on specifications of user devices as shown in Eqs. 2 and 3. Because the resolution aspect ratio is 16:9 on most devices, we divide the screen into grid cells accordingly (160 × 90) to make them evenly distributed with acceptable visual effect.

$$\text{gridWidth} = \text{screenWidth}/160 \tag{2}$$

$$\text{gridHeight} = \text{screenHeight}/90 \tag{3}$$

Then we render the grid cells with the HSLA (hue, saturation, lightness, alpha) (Smith 1978) color representation system, which is commonly used for displaying web pages. The color of each grid cell is determined by the hue, saturation, lightness, and alpha (opacity). The hue is a degree on the color wheel between 0 and 360, where 0 represents red, 120 represents green, and 240 represents blue. The hue of a grid cell is calculated based on the density of trajectories as shown in Eq. 4:

$$\text{hue} = 255 - \frac{n - \min(n)}{\max(n) - \min(n)} \times 255 \tag{4}$$

where n is the number of trajectory points in a grid cell; $\min(n)$ and $\max(n)$ are the minimum and maximum number of trajectory points among all grid cells. Therefore, the grid cell with the most trajectory points is always red, and the grid cell with the least trajectory points is always blue. The saturation is a percentage value showing the brilliance and intensity of a color, where 0% represents a shade of gray, and 100% represents the full color. We use 100% saturation for all grid cells. The lightness represents the amount of white or black mixed in with the color. It is also a percentage value, where 0% is black, and 100% is white. We use 50% lightness for all grid cells. The alpha represents the transparency of a color as a number between 0.0 (fully transparent) and 1.0 (fully opaque). The alpha of a grid cell is also calculated based on the density of trajectories as shown in Eq. 5a and b:

$$alpha = \begin{cases} max\left(\frac{n - min(n)}{max(n) - min(n)}, 0.2\right) & n > 0 \quad \text{(5a)} \\ 0 & n = 0 \quad \text{(5b)} \end{cases}$$

Therefore, the grid cells with lower density of trajectories will be more transparent, and the grid cells without any trajectory points will be fully transparent. With these self-adaptive adjustments of colors, the system can guarantee that an area with relative high density of trajectory points is always clearly distinguishable in various scales (Fig. 3b).

Interactive Multi-spatial-temporal-view Charts

Exploratory analyses of trajectory data require researchers to examine trajectory patterns at various spatial and temporal scales. The previous version of TrajAnalytics only provides summary statistics of trajectories by hour and by day. However, the temporal scales of sparse trajectory

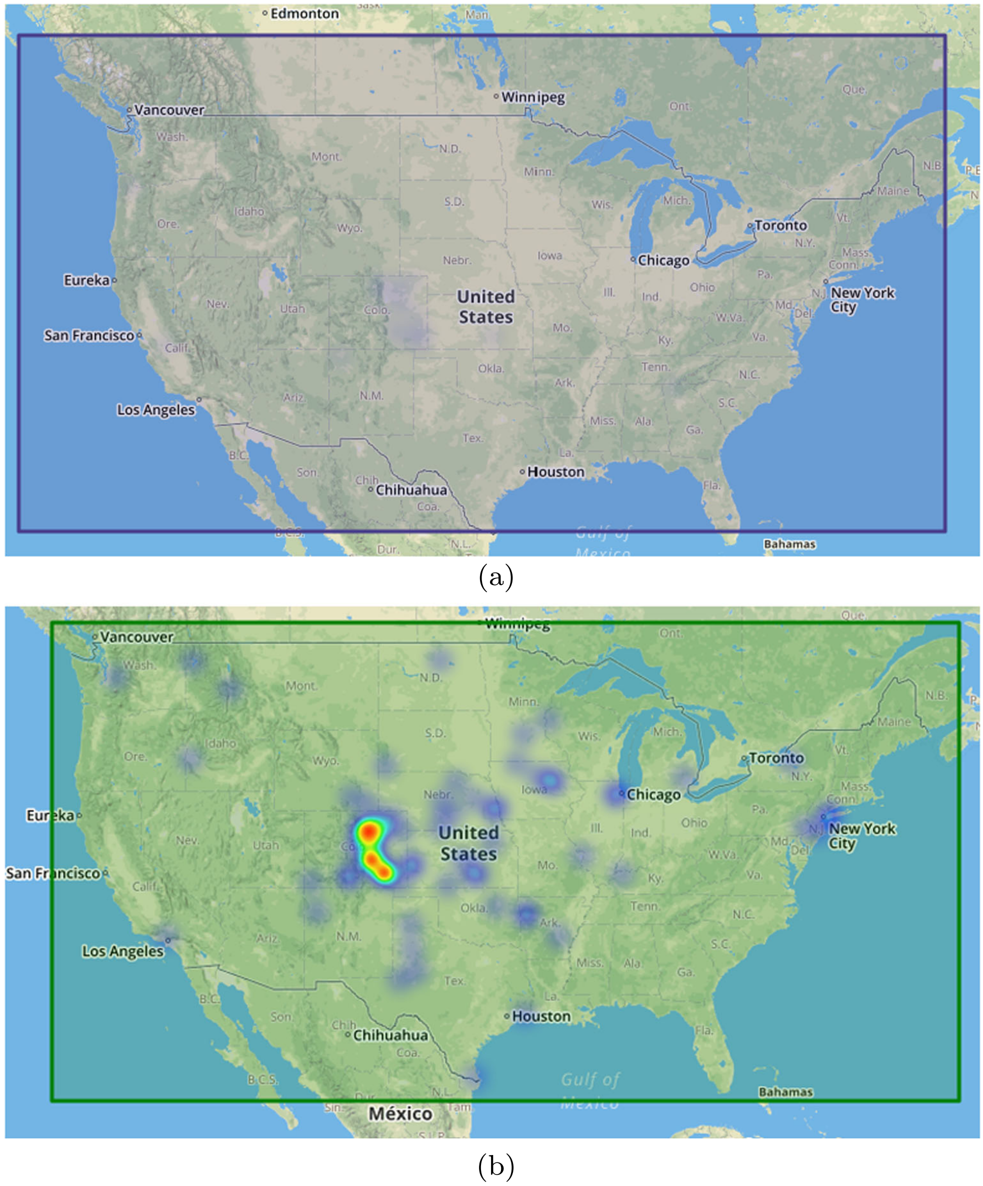


Fig. 3 Self-adaptive adjustment of the heatmap's color ramp for sparse trajectory data: (a) before the adjustment; (b) after the adjustment

data are larger than those of dense trajectory data. To better reveal temporal patterns of sparse trajectory data, we added statistical charts that summarize the trajectories by month and by year to the new system. The multi-scale temporal charts allow users to navigate patterns across temporal scales flexibly. Additionally, the map view of sparse trajectory data also provides flexible zoom in/out functionalities for users to examine trajectory patterns at different spatial scales.

As shown in Fig. 4, the visualization interface consists of three components, including (1) query control panel (left); (2) mapping panel (center); and (3) statistical information panel (right). The three components are interrelated. The query control panel stores query histories of the users. Users can use the query control panel to filter trajectories based on given time intervals and/or geo-locations as introduced in Section "Data Query." Only filtered results will be visualized on the mapping panel and statistical information panel. Users can click a trajectory on the map panel, and the corresponding data record will be highlighted on both the mapping panel and statistical information panel. If users select a row of the table (Fig. 4a) or a bar of the chart (Fig. 4b and c) on the statistical information panel, corresponding trajectories would be highlighted on the map as well. The summary table in the statistical information panel automatically ranks the lengths of the selected trajectories (Fig. 4a). Figure 4 visualizes the hurricane trajectories that passed through Florida between 2000 to 2018 using the interactive multi-spatial-temporal-view charts. Users can interactively navigate the interrelated panels to explore hurricane trajectory patterns at different spatial-temporal scales.

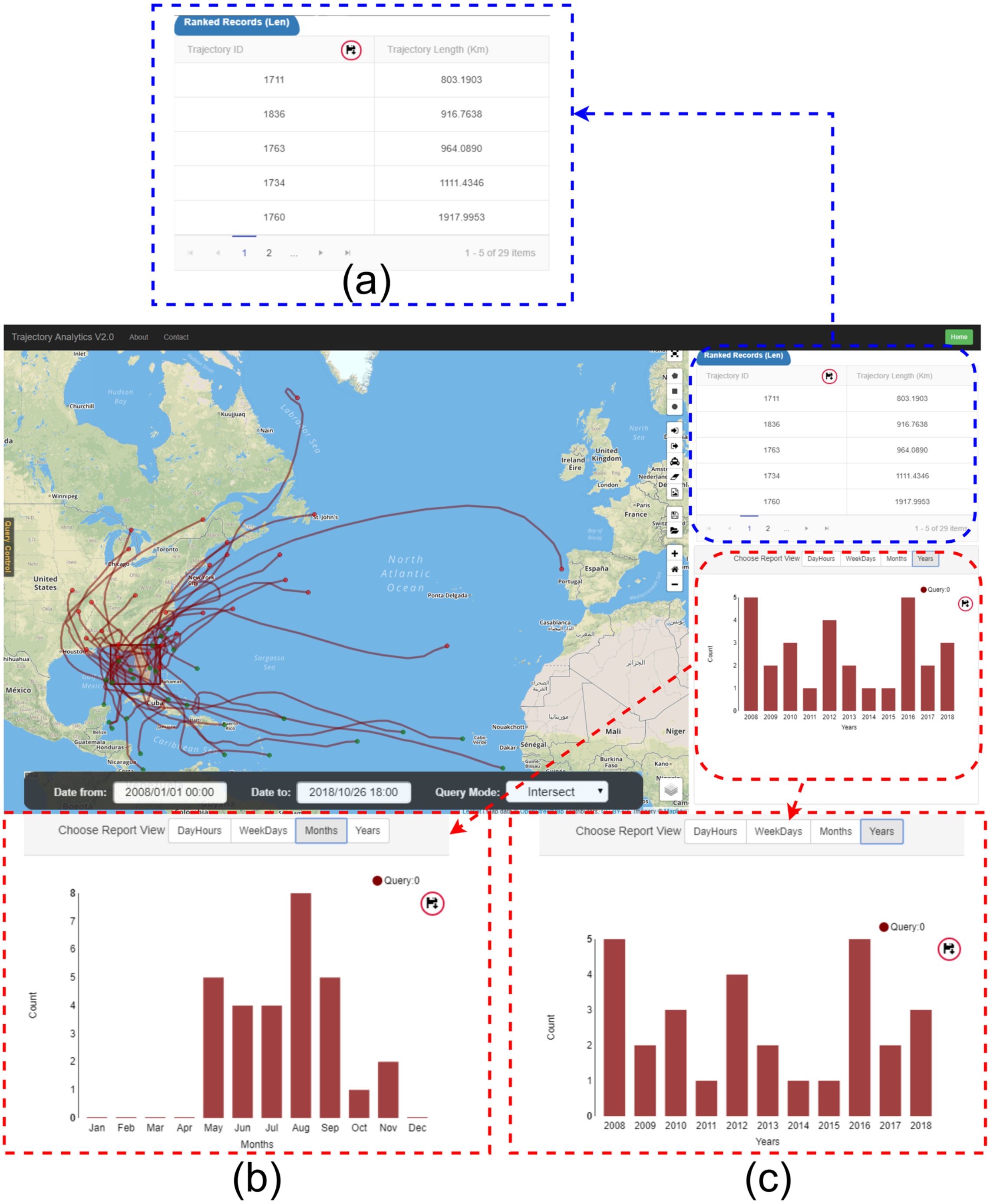


Fig. 4 Visualization of hurricane trajectory data using interactive multi-spatial-temporal-view charts: (a) a statistical table showing the trajectory information (b) a statistical chart showing trajectories counts aggregated by month; (c) a statistical chart showing trajectories counts aggregated by year

Evaluation and Discussions

Managing and visualizing spatial and temporal information of sparse trajectory data are the cornerstones for further trajectory research (Wang et al. 2020), which also cover the major concepts and features in the SparseTrajAnalytics system. Therefore, we would like to evaluate whether researchers can utilize our system to manage and visualize sparse trajectory data smoothly and efficiently.

We conducted a survey among 90 domain experts (30 in Oceanology, 30 in Transportation, and 30 in Geographic Information Science). The survey with each participant

Table 4 Questions for visualization tasks

For migration data	For hurricane trajectory data
1. Which place were people most likely to move to during 1980 1990?	1. What year/month was hurricanes most likely to happen?
2. Identify the top 3 years during which people were mostly likely to migrate.	2. If a hurricane passed through New York, where did it usually come from?
3. Where were people most likely to escape from during the study period?	3. Which year had the most hurricane strikes in America?

lasted 1 h. First, we described the functionalities and interface of the system in details to these experts through a case study about taxi trajectories. Then, each expert explored the system on a touch screen for about 20 min. After these preparations, the participants need to visualize two sparse trajectory data sets (migration data and hurricane trajectory data provided in .csv format) using the system. These two visualization tasks are specifically designed to cover most features of the system. The migration data are a nationwide data set consisting of only start and end points of trajectories recorded on an annual basis. Participants need to geocode these sample points to zip code regions. The hurricane data contain trajectory points that may be located in oceans or multiple countries, which could not be matched to administrative areas (such as zip code regions). Participants need to geocode the trajectories to self-defined grid regions or visualize directly using sample point coordinates. The two tasks also require participants to use spatial-temporal queries and all visual charts/tables and map views in the SparseTrajAnalytics to answer questions related to the data (Table 4). The time used by each participant on each task and their answers to the questions were recorded to quantitatively measure their performance. After the visualization tasks, each participant rated their experience on learning and using the SparseTrajAnalytics on a 10-point scale and voluntarily provided additional comments about the system.

All participants successfully completed the two visualization tasks and answered the questions in Table 4. They used an average of 15.5 min (standard deviation σ = 5.61 min) for the migration data, and an average of 12.77 min (σ = 4.30 min) for the hurricane trajectory data (Table 5). The relative short time spent by these first-time system users demonstrates the intuitiveness of our system. The average user experience scores are 8.0 for learning (10—very easy, 1—very difficult), 9.17 for data visualizations (10—very easy, 1—very difficult), and 7.15 for data queries (10—very functional, 1—not functional). Generally, all participants agreed that the system is intuitive and user-friendly based on their domain knowledge and experience. It matches our goal to make this system a universal tool for sparse trajectory data analysis.

Many experts gave additional positive comments on the SparseTrajAnalytics after comparing it with other systems. One expert mentioned:

> Tableau and other graphic design tools has a bit of a learning curve, but there isn't as much of a learning curve in your system.

Another expert wrote:

> ArcGIS may have more features, but your system is free and easy to use in the internet, I would try your system first if I have sparse trajectory data.

A third expert commented on the functionalities of the system:

> I want to call it a platform instead of a tool because domain users and researchers can get many hints of how to use the sparse trajectory data along with the maps and time constraints in a more intuitive way. In other words, the current default setting can allow the user to become familiar with the data from multiple perspectives of geography such as locations with map information, temporal information in a comparative context.

Other experts' comments suggest that they like the charts and tables on the statistical information panel, where they can compare and study the sparse trajectory data in different views.

Based on the experts' feedback, our SparseTrajAnalytics system has several strengths. First, the system provides

Table 5 Results of the survey among 90 domain experts

Tasks and user experience	Mean (μ)	Standard deviation (σ)
Time used to analyze the migration data (minutes)	15.5	5.61
Time used to analyze the hurricane trajectory data (minutes)	12.77	4.30
Learning (10—very easy, 1—very difficult)	8.0	0.96
Data visualizations (10—very easy, 1—very difficult)	9.17	0.77
Data queries (10—very functional, 1—not functional)	7.15	0.90

an efficient platform for users to manage and analyze the sparse trajectory data intuitively. Second, the system can be hosted as either public or private cloud services, so no extra downloading/installation is needed for end users. Third, multiple indexes are integrated into the data prepossessing to improve the performance of the system. Finally, users can explore spatial-temporal patterns of sparse trajectory data visually and interactively with the specifically designed self-adaptive heatmap and interactive multi-spatial-temporal-view charts.

Meanwhile, the experts also gave us valuable suggestions to improve the system. Some experts suggested that an interactive tutorial of the system for first-time users and more descriptions to the icons in the top panel are needed. Based on their suggestions, we modified some icons in the system and added a video tutorial to improve the user experience. Some experts would like to add a panel for real-time traffic and traffic predictions in a city if data is available, which matches our future plan to combine the system with commercial map service APIs. Some tasks like retrieving region information would take 5 to 10 s, but the system might freeze during this process. Some experts mentioned that users might feel anxious and wonder if the system still runs normally. This problem is caused by the limited computing resources we have now. We plan to enhance the system performance using high performance computing in the future. Besides the geocoding process (which takes up to 5 min each in our migration and hurricane sample data sets), other operations can all have immediate responses in our system. This is because both the back-end services and the front-end visualizations are very efficient. The back-end utilizes the PostGIS (Obe and Hsu 2011) , which is a powerful spatial database extension that can perform spatial operations efficiently. The front-end is created using leaflet.js (Leaflet 2015), which can easily visualize over 100,000 trajectory points without delay. One possible bottleneck in the system might be the network speed between the back-end and the front-end, but it is hard to test for the complex situations in the real world. There was no network delay reported in our system evaluations. However, to deal with the possible network speed problem, the source codes are also provided to allow local deployment of the system, where no data transfer over Internet will be needed.

Several additional future works can also be done to improve the system, such as (1) integrating more offline geospatial data to facilitate aggregation at different spatial scales; (2) supporting more data formats and conversions among these formats; (3) enabling users to share their open data on the platform more conveniently; (4) adding more map visualization types, such as migration map, flow map, and dynamic movement map; and (5) adding more spatial analysis functions to the interactive multi-spatial-temporal-view charts.

Conclusion

This study developed a visual analytics system for sparse trajectory data called SparseTrajAnalytics. The system extends the TrajAnalytics, which was designed for dense trajectories, to better support sparse trajectory data, such as human migration and hurricane trajectories. The system allows users to store and manage trajectory data on a cloud service and to query trajectory data efficiently with keywords, spatial locations, and/or time constraints. The system also integrates a self-adaptive heatmap and interactive multi-spatial-temporal-view charts, which are specifically designed for visually analyzing sparse trajectory data. The self-adaptive heatmap can better visualize the various densities of sparse movement behaviors. The interactive multi-spatial-temporal-view charts provide an interactive interface for users to examine trajectory patterns across various spatial and temporal scales. The survey among domain experts shows that the system is intuitive and user-friendly for researchers to manage and visualize sparse trajectory data.

Acknowledgments We greatly appreciate the valuable comments and suggestions from the editors and anonymous reviewers. This material is partially based upon work supported by the National Science Foundation under Grant Nos. 1739491 and 1937908. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

References

AL-Dohuki S, Kamw F, Zhao Y, Ye X, Yang J (2019) An open source trajanalytics software for modeling, transformation and visualization of urban trajectory data. In: The 22nd IEEE Intelligent transportation systems conference. IEEE, pp 150–155. https://doi.org/10.1109/ITSC.2019.8917058

Basiri A, Jackson M, Amirian P, Pourabdollah A, Sester M, Winstanley A, Moore T, Zhang L (2016) Quality assessment of OpenStreetMap data using trajectory mining. Geo-Spatial Information Science 19(1):56–68. ISSN 10095020. https://doi.org/10.1080/10095020.2016.1151213

Bell M, Charles-Edwards E, Kupiszewska D, Kupiszewski M, Stillwell J, Zhu Y (2015) Internal migration data around the world: assessing contemporary practice. Population, Space and Place 21(1):1–17

Bostock M, Ogievetsky V, Heer J (2011) D3 data-driven documents. IEEE Trans Vis Comput Graphics 17(12):2301–2309. ISSN 10772626. https://doi.org/10.1109/TVCG.2011.185
Boyandin I, Bertini E, Bak P, Lalanne D (2011) Flowstrates: an approach for visual exploration of temporal origin-destination data. Computer Graphics Forum 30(3):971–980. ISSN 14678659. https://doi.org/10.1111/j.1467-8659.2011.01946.x
Cadenas C (2014) Geovisualization: integration and visualization of multiple datasets using mapbox. Computer Engineering. https://digitalcommons.calpoly.edu/cpesp/137
Coscia M, Giannotti F, Pedreschi D (2012) Interactive visual querying and analysis for urban trajectory data. PhD thesis, IT University of Copenhagen. arXiv:1206.3552
Du Z, Zhang X, Li W, Zhang F, Liu R (2020) A multi-modal transportation data-driven approach to identify urban functional zones: an exploration based on Hangzhou City, China. Trans GIS 24(1):123–141
Ferreira N, Poco J, Vo HT, Freire J, Silva CT (2013) Visual exploration of big spatio-temporal urban data: a study of New York city taxi trips. IEEE Trans Vis Comput Graphics 19(12):2149–2158. ISSN 10772626. https://doi.org/10.1109/TVCG.2013.226
Guo H, Wang Z, Yu B, Zhao H, Yuan X (2011) TripVista: Triple Perspective Visual Trajectory Analytics and its application on microscopic traffic data at a road intersection. In: IEEE Pacific Visualization Symposium 2011, PacificVis 2011 - Proceedings, pp 163–170. ISBN 9781612849324. https://doi.org/10.1109/PACIFICVIS.2011.5742386
Hurter C, Tissoires B, Conversy S (2009) FromDaDy: spreading aircraft trajectories across views to support iterative queries. In: IEEE Transactions on Visualization and Computer Graphics, vol 15, pp 1017–1024. https://doi.org/10.1109/TVCG.2009.145
Kapler T, Wright W (1997) GeoTime information visualization. Inf Vis 4(2):136–146. ISSN 14738724. https://doi.org/10.1057/palgrave.ivs.9500097
Landsea CW, Cangialosi JP (2018) Have we reached the limits of predictability for tropical cyclone track forecasting? Bull. Am. Meteorol. Soc. 99(11):2237–2243
Landsea CW, Franklin JL (2013) Atlantic hurricane database uncertainty and presentation of a new database format. Monthly Weather Review. ISSN 0027-0644. https://doi.org/10.1175/mwr-d-12-00254.1
Leaflet JS (2015) Open-source javascript library for mobile-friendly interactive maps
Li W, Wang S, Bhatia V (2016) Polarhub: a large-scale web crawling engine for ogc service discovery in cyberinfrastructure. Comput Environ Urban Syst 59:195–207
Liu H, Gao Y, Lu L, Liu S, Qu H, Ni LM (2011) Visual analysis of route diversity. In: VAST 2011 - IEEE Conference on visual analytics science and technology 2011, Proceedings, pp 171–180. ISBN 9781467300131. https://doi.org/10.1109/VAST.2011.6102455
Mather M, Cortes R (2016) Demographic data: censuses, registers, surveys. In: The Blackwell Encyclopedia of Sociology, pp 1–4. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781405165518.wbeosd018.pub2
Meng C, Xiao H, Su L, Cheng Y (2016) Tackling the redundancy and sparsity in crowd sensing applications. In: Proceedings of the 14th ACM Conference on embedded network sensor systems CD-ROM, pp 150–163
Obe R, Hsu L (2011) Postgis in action. GEOInformatics 14(8):30
Phan D, Xiao L, Yeh R, Hanrahan P, Winograd T (2005) Flow map layout. In: Proceedings - IEEE symposium on information visualization, INFO VIS, pp 219–224. ISBN 078039464X. https://doi.org/10.1109/INFVIS.2005.1532150
Skeldon R (2012) 11 migration and its measurement: Towards a more robust map of bilateral flows. Handbook of Research Methods in Migration, p 229
Smith AR (1978) Color gamut transform pairs. ACM Siggraph Computer Graphics 12(3):12–19
Von Landesberger T, Bremm S, Andrienko N, Andrienko G, Tekušová M (2012) Visual analytics methods for categoric spatio-temporal data. In: IEEE conference on visual analytics science and technology 2012, VAST 2012 - Proceedings, pp 183–192. ISBN 9781467347532. https://doi.org/10.1109/VAST.2012.6400553
Wang H, Zeng H, Li Y, Zhang P, Jin D (2020) Human mobility prediction using sparse trajectory data. IEEE Transactions on Vehicular Technology
Weinkle J, Landsea C, Collins D, Musulin R, Crompton RP, Klotzbach PJ, Pielke R (2018) Normalized hurricane damage in the continental United States 1900–2017. Nature Sustainability 1(12):808–813
Willekens F, Massey D, Raymer J, Beauchemin C (2016) International migration under the microscope. ISSN 10959203
Willems N, Van De Wetering H, Van Wijk JJ (2009) Visualization of vessel movements. Computer Graphics Forum 28(3):959–966. ISSN 14678659. https://doi.org/10.1111/j.1467-j.1467-8659.2009.01440.x
Wood J, Dykes J, Slingsby A (2010) Visualisation of origins, destinations and flows with OD Maps. The Cartographic Journal 47(2):117–129. ISSN 0008-7041. https://doi.org/10.1179/10.1179/000870410x12658023467367
Xia H, Riche NH, Chevalier F, De Araujo B, Wigdor D (2018) Dataink: direct and creative data-oriented drawing. In: Conference on human factors in computing systems - Proceedings, volume 2018-April. Association for computing machinery. ISBN 9781450356206. https://doi.org/10.1145/3173574.3173797
Yu H, Shaw S-L (2013) A custom extension of extended time-geographic framework tools in arcgis. http://web.utk.edu/sshaw/NSF-Project-Website/download/User_Guide_Time_Geographic_Framework_Extension.pdf
Zhao Y, Eynon T, Kamw F (2016) David Sheets, [Workshop] TrajAnalytics : a web-based visual analytics software of urban trajectory data. In: IEEE Workshop on visualization in practice : Open source visualization and visual analytics software. http://geoanalytics.net/
Zhao Y, Al-Dohuki S, Eynon T, Kamw F, Sheets D, Ma C, Hu Y, Ye X, Yang J (2017) Trajanalytics: a web-based visual analytics software of urban trajectory data. In: Proceeding of IEEE Visualization Conference 2016. https://doi.org/10.6084/m9.figshare.4573057.v1
Zheng Y (2015) Trajectory data mining. ACM Trans Intell Sys Technol 6(3):1–41. ISSN 21576904. https://doi.org/10.1145/2743025

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.