ORIGINAL ARTICLE

TrajectoryVis: a visual approach to explore movement trajectories

Samiha Fadloun[1] · **Yacine Morakeb**[1] · **Erick Cuenca**[2] · **Kheireddine Choutri**[3]

Abstract
Social networks are a dominant data source for sharing, participation, and exchanging information. For example, Twitter is a microblogging site that enables users to express opinions by transmitting brief messages (i.e., Tweets). Tweets can be used to extract information on users' movements or trajectories over time. Information visualization (InfoVis) is helpful to understand, analyze, and make decisions about these trajectories. To better understand and compare existing visual encoding methods in InfoVis, we propose *TrajectoryVis*, a generic trajectory visualization tool to represent social network datasets (e.g., Twitter). Individual and aggregated trajectories can be visualized using different visual coding approaches. Our approach is assessed using a user and a COVID-19 case study to prove its effectiveness.

Keywords Information visualization · Spatio-temporal visualization · Trajectory visualization · Social network data

1 Introduction

Nowadays, social networks (e.g., Facebook, Instagram, Twitter) play an essential role in the web landscape. For instance, during the coronavirus lockdown, people used to stay at home, and the time spent on social networks increased. Twitter is a microblogging platform that allows users to exchange short messages called Tweets. These messages relate to different events (e.g., politics, culture, sports, health, etc.) and areas. Tweets can often contain multi-dimensional details such as locations (e.g., people placements) and topics (e.g., food transportation) that move over time. These movements build paths (trajectories) or flow (set of trajectories) in a period. For example, Hu et al. (2019) proposed a solution to visualize trajectory patterns related to tourism industry on Twitter.

Several datasets contain information related to trajectories, such as human travels, vehicles, airplanes, etc. Information Visualization (InfoVis) Fekete et al. (2008) can help users to analyze, understand and extract information from such datasets. The InfoVis domain translates trajectories to visual encoding variables (e.g., colors, lines, arrows, glyphs, shapes, etc.) on geographic maps or abstract views (Schöttler et al. 2021).

Using existing visualization tools like Tableau[1] can be inaccurate for trajectory visualization. Most business tools are designed around large data types and visual encoding methods to influence more users. Also, users have limited access to some standard visual encoding methods. They have to pay to get all preferred visual coding permissions that are not customizable for a trajectory dataset.

This paper proposes *TrajectoryVis*,[2] an approach to visualize trajectories from social networks. We start with a literature review rich in the most representative visualization approaches of Spatio-temporal data. Next, we describe the proposed tool to help us achieve our goals. Finally, TrajectoryVis is assessed by possible means of a qualitative evaluation and a comprehensive case analysis.

✉ Samiha Fadloun
s_fadloun@esi.dz

Yacine Morakeb
fy_morakeb@esi.dz

Erick Cuenca
ecuenca@yachaytech.edu.ec

Kheireddine Choutri
choutri.kheireddine@univ-blida.dz

[1] Laboratoire des Méthodes de Conception de Systèmes (LMCS), Ecole nationale Supérieure d'Informatique (ESI), BP 68M -16 270, Oued Smar, Alger, Algérie

[2] Yachay Tech University, Urcuquí, Ecuador

[3] Aeronautical Sciences Laboratory, Aeronautical and Spatial Studies Institute University Blida 1, Blida, Algeria

---

[1] https://www.tableau.com/ (accessed on March 2022).

[2] https://youtu.be/aqoVOMu_Z6Q (accessed on March 2022).

2 Related work

Trajectory visualizations can handle several visual coding techniques. These techniques can consist of mapping spatial, temporal, or Spatio-temporal information onto visual representations. This section presents an overview of techniques dealing with these dimensions.

2.1 Spatio-temporal data

Spatio-temporal data refers to the analysis of the attributes of an object in a spatial and temporal context. Peuquet (1994) proposed a triadic model to describe how these three components interact: space (*where*), time (*when*), and objects (*what*). Peuquet particularly highlights the following correlations:

- **When + where → what**: expresses the information of the object(s) at a given location(s) at a given time(s).
- **When + what → where**: illustrates the location(s) of the object(s) at a given time(s).
- **Where + what → when**: defines the time(s) of the object(s) at a given location(s).

Ward et al. (2010) presented a visualization characterizing the *when* by joining the *where* and the *what* information. They highlight the spread of cholera outbreaks in London in the 19th century by statistical analysis of deaths around a water distribution. A well-known visualization including the three attributes (*where*, *when*, and *what*) was proposed by Charles Minard in 1869.[3] It shows Napoleon's Russian campaign, and it includes six data values in a single view: geography of the terrain, time of events, the evolution of the number of military troops over time, the direction of the troops, the distance traveled by the troops, and the terrain temperature.

2.1.1 Spatial trajectory visualization (where and what)

It can be divided into visualization based on points, lines, and regions. Point-based trajectory visualizations are the most straightforward visualization for presenting and analyzing geolocations. These displays place trajectory samples in spatial context as individual discrete points, each indicating a target or event, and encodes linked variables with visual channels of points. For example, in the field of traffic, Ding et al. (2015) designed a Trains of Data project which designates each train as a moving point operating on a 2D map. This type of visualization is promising, particularly in the distribution of pickup and dropout events in trajectories through transportation hubs. Line-based visualizations are represented as straight or curved lines from the point of origin to the point of destination (Crnovrsanin et al. 2009; Gupta et al. 2016) (Origin-Destination). These trajectories can also be transformed spatially via topological or geometric algorithms, then restored in other spaces.

2.1.2 Temporal trajectory visualization (when and what)

It is linked to the encoding of time in geographic space (Yang et al. 2017; Abbott 2013; Aigner et al. 2011; Miller 1999). For instance, Yang et al. (2017) proposed a representation of trajectories of Origin-Destination (OD) pairs with different colors on particular road segments over a given period. The time-distance transformation technique (Miller 1999) provided insight into individual trajectory periods with different granularities.

2.1.3 Attributes trajectory visualization (what)

It reflects changing patterns over time relative to the non-spatial information of research objects, where characteristics of specific attributes are usually encoded by visual elements (Cuenca et al. 2018; Willems et al. 2010; Ryoo et al. 2018; Scheepens et al. 2014; Guo et al. 2011; Adrienko and Adrienko 2010; Andrienko and Andrienko 2008; Hu et al. 2019; Wallner et al. 2019; Zhu and Guo 2014). For example, Scheepens et al. (2014) suggested the design of a glyph. They indicate information on the types of vessels and the respective quantities. In addition to explicit attributes, such as visual encoding and visual models, there are also implicit attributes, such as path shape as a continuous geometric attribute feature. Another example is the safety monitoring system designed by Willems et al. (2010). It displays the correlation between a pair of attributes in a trajectory contingency table. Ryoo et al. (2018) presented a visual analysis based on a pixel grid. It comprehensively displays changes over time in individual performance and characteristics of a soccer team. One of the objectives of applying rational visualization of attribute characteristics is to implement predictive analysis.

2.2 Movement and trajectories

The movement of an object represents the change in its location(s) overtime(s), while a trajectory is a sequence of its locations in a given time interval. Most of the time, datasets are made up of multiple objects that move over time. In this context, trajectory flows represent the aggregation of movements between common origin and destination points. These datasets are called Origin Destination flows (OD).

---

[3] https://www.masswerk.at/minard/ (accessed on March 2022).

Visualizing trajectories is challenging since it typically involves representing the spatial (where), temporal (when), and the context of data (what). Trajectories can be visualized using geographic or abstract spaces. The first one commonly uses a 2D map visualization to represent data focusing on its geographical location. The second one uses abstract spaces to encode the necessary data visually.

2.2.1 Trajectory flows visualization

There are some approaches to visualize trajectory flows: flow maps (Card et al. 1999; Maahs et al. 2012; Phan et al. 2005; Zhu and Guo 2014), origin-destination matrix (OD-matrix) (Guo 2007), and the origin-destination maps (OD-Maps) (Wood et al. 2010).

Flow maps use straight or curved lines to connect locations on a map, i.e., origin and destination points. Zhu and Guo (2014) proposed an aggregation method specially designed for a large dataset containing origin-destination flows. This method is based on clustering with the aim of reducing the difficulty of mapping and understanding the patterns in a large dataset due to the problem of occlusion and cluttered display, where thousands or millions of streams overlap and intersect. Phan et al. (2005) used the magnitude of the amount of migration in a single flow line with different thicknesses. This allows the user to indicate how migration is distributed geographically. In addition, merging flow lines can help to prevent crossings and therefore reduce visual overlap on the map. When properly designed, flow maps are beneficial because they allow users to detect differences in amplitude of a wide variety across space with minimal cartographic overlap.

The OD-matrix technique (Guo 2007) used a matrix as the name suggests, where the rows represent the origins and the columns represent the destinations. This technique is suitable when data increase since a matrix scales better than a flow map; however, the spatial context is lost. The number of flows is often represented using a color palette in a matrix. The major drawback of this technique is the unavailability of the geographical representation, which is why the OD-Maps appeared. OD-Maps (Wood et al. 2010) used a matrix to divide the geographic space into two integral levels. The first level shows the locations of the origins, and the second nested level shows the destinations. This approach shows the spatial dimension of the data; however, it suffers from scaling issues as the data increases.

To be able to compare approaches from previous work, a certain number of comparison criteria must be defined. We take into account the methods that interact with: spatial dimension (C1), temporal dimension (C2), origin-destination data flows (C3), and large datasets (C4). We observe that C1 and C3 are the most used features in the approaches cited in Table 1.

Table 1 Flow trajectories visualization approaches comparison

Technique	C1	C2	C3	C4
Flow Maps Card et al. (1999), Maahs et al. (2012), Phan et al. (2005), Zhu and Guo (2014)	✓	✗	✓	✗
OD-matrix Guo (2007)	✓	✗	✓	✓
OD-Maps Wood et al. (2010)	✓	✗	✓	✗

2.3 Extracting trajectories from twitter

Twitter is an important resource for several studies, especially in data mining in general (Rettore et al. 2020). For example, in the tourism industry, Hu et al. (2019) proposed a visual approach to detect tourist movement patterns and represent the Spatio-temporal trajectory of tourist movement from massive and noisy data from this social network. It consists in building the tourist graph using the DBSCAN (Ester et al. 1996) algorithm to group the tourist trajectories to identify the peaks in a chart. In recent years, visualization researchers have used Tweets as a dataset to test their proposition. For example, Chen et al. (2018) proposed an approach to data analysis in which texts are associated with spatial and temporal references. Kim et al. (2017) developed a new flow analysis technique to extract, represent and analyze non-directional Spatio-temporal data flow maps unaccompanied by trajectory information.

Commercial applications have also been developed. For example, Tweet Ping[4] is a site allowing you to view all the Tweets sent in the main regions of the world (e.g., South America, North America, Europe, Africa, and Asia) as well as the number of words and corresponding characters, the last mentions (@) or the last hashtags. Another approach is proposed by Senaratne et al. (2014). It extracts Tweets and analyzes them using filtering techniques, geotag, and clustering or grouping to find the possible routes in the data flow.

2.4 Discussion

The visualization of trajectories using Tweets is a new area of research. Most of the approaches cited below Hu et al. (2019), Chen et al. (2018), Kim et al. (2017), Senaratne et al. (2014), Krueger et al. (2016) are recent, they date back at most six years. Each of these approaches has advantages. For instance, Hu et al. (2019) offers several types of visualization for better use of data, Kim et al. (2017) better interactivity and data structuring, Chen et al. (2018) better visualization of the propagation of data. But all these approaches have limitations, such as the accuracy of the results presented, the

---

Fig. 1 *TrajectoryVis* process.
(1) Data analyzing, filtering,
and transforming. (2) Storage of
structured data in database. (3)
Visual encoding with different
views
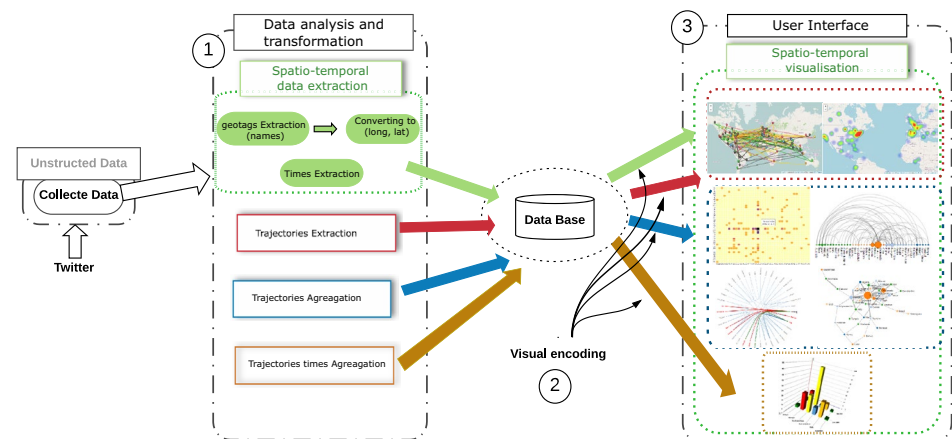
analysis of micro texts, the treatment of several languages, and the extraction of paths with these conditions.

We carefully choose Twitter as a data source in our visualization because it properly provides access to mass published data that dynamically captures the social activities of its users. Otherwise, it is utilized in many visualizations.

3 Requirement analysis

After state-of-the-art analysis in various fields. We define five requirements for proposing a general tool to visualize trajectories as follow:

- **E1: Extracting trajectories from Tweets**. Tweets can contain **where, when** and **what** information. They can also have multiple languages, spelling or abbreviation errors, messed up data, etc. Before representing data from Twitter, the processing step should extract, analyze, filter, and structure them.
- **E2: Handling spatio-temporal data**. Trajectories data contains spatial and temporal information. This information needs to be structured and represented. A visual encoding has to represent each components of **what**, **when** and **where**, and the correlation between them.
- **E3: Representing individual trajectories**. The change of Spatio-temporal information over time can create trajectories. These trajectories have to be represented. The visual encoding of these trajectories have to carry out with **what**, **when** and **where** components.
- **E4: Representing trajectories aggregation**. The number of trajectories can be large. As a result, it is necessary to find a visual encoding that aggregates these trajectories preserving information.
- **E5: Interaction**. For querying the data in the requirements **E (2, 3, 4)**, we need to handle interactions such as selecting, filtering, zooming, etc. These interactions

allow the user to quickly and easily analyze, manipulate and understand the trajectories.

4 *TrajectoryVis* design

We propose *TrajectoryVis*, a visual system that meets all of the previously stated requirements. Fig. 1 depicts the pipeline, which is followed by our visualization. We use Twitter as a source of movement data. First, we gather information from Twitter. The collected data is then analyzed and filtered (**E1**) (Fig. 1.1) to create spatio-temporal information (**E2**). These data are used to build trajectories, either individual (**E3**) or aggregated (**E4**). These types of data are saved in a database (Fig. 1.2). Flow visualizations such as flow maps, OD-matrix, and graphs are used to visually encode each data type (Fig. 1.3). In addition, histograms are used to describe temporal information. Finally, in each TrajectoryVis view, we add several interactions (**E5**).

4.1 Experiment data

To begin, we collect data using the streaming API and a hashtag filter predicate. It allows us to receive Tweets with the desired hashtags in real-time. We were able to collect over 50,000 structured Tweets in JSON format between May and June 2020, using the hashtags #Covid19 or #Corona.

Following the collection of Tweets in the preceding step, we filter, delete, clean, and convert the collected dataset as follows:

1. We sorted the Tweets by location and then removed those that lacked spatial details. We discovered issues in short text analysis such as abbreviations, language
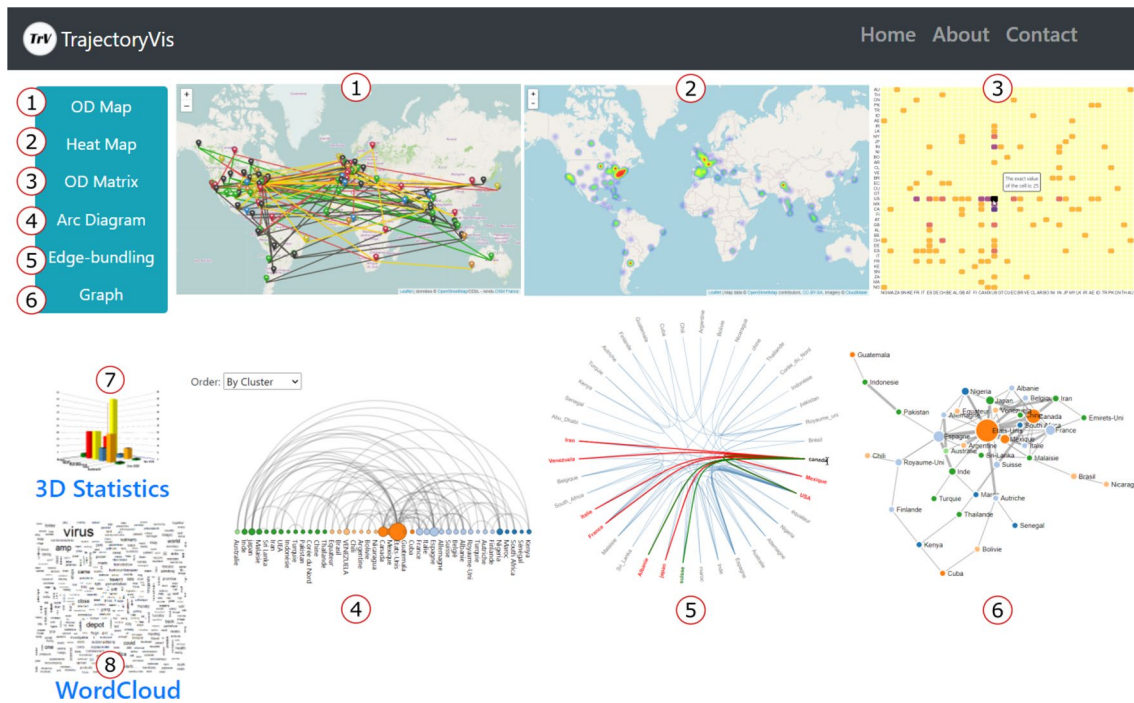
Fig. 2 *TrajectoryVis* views. (1) Origin Destination map represents the individual trajectories. (2) Heatmap encode the trajectory quantities over different locations. (3) Second representation of quantities with heat matrix. (4,5,6) Data abstraction using graphs in one, two and circle dimensions. (7) Time abstraction using 3D bar chart. (8) Additional representation of word cloud extraction from Tweets

errors (we chose English), missing information (when or where), and so on.

2. We used the Geocoding API[5] from Google to convert addresses (locations) to latitude and longitude (**E2**).
3. Then, in our dataset, each Tweet is saved as a triplet (Tweet text, location, time). % For each hashtag, we had a data set as follows:
4. We follow the hashtag spread in space and time to build the trajectories (**E3**). We classified the data based on their chronological order, such as: $t_i < t_{i+1}$
5. To reduce trajectory overlap (**E4**), we aggregated the data relative to a location (country) l_i over time t_i (we calculated the trajectory occurrence).

4.2 Visual mapping

We will describe our visual encoding for each data type in this section (Fig. 2). We begin with the spatio-temporal type, which contains longitude and latitude information encoded with a line map (Fig. 2.1).

A heatmap (Fig. 2.2) that encodes the granularity of trajectories in each location is also provided. Abstract types that represent locations (names), trajectories number in/ between locations, and temporal information are represented by various views such as graphs one (Fig. 2.4), two (Fig. 2.6), and circle (Fig. 2.5), heat matrix (Fig. 2.3), and 3D bar chart or histogram (Fig. 2.7). An additional word cloud is displayed to analyze the most frequently used word in the covid Tweets.

4.2.1 Line map

We use straight lines to represent a set of trajectories (**E1, E2**) on a geographic map. Each line is an aggregation of the trajectories of a hashtag (Fig. 3). We also use icons to depict the origin/destination at initial source/final target locations. The blue icons: the departures, the orange icons: the arrivals, the icons of the same color as the line: the intermediate locations. We use colors for the lines to distinguish the paths between them. It enables us to plot trajectories on a map using a simple representation.

We also provide trajectories filtering by month and week of hashtags (**E5**). Fig. 4 depicts an individual trajectory of hashtag #Covid19 from the 1st to the 15th of May. The journey begins in Africa and ends in the United States.

4.2.2 Heat map

A heat map (Liu and Heer 2018) displays aggregated trajectories (**E1, E3**) using matrix coding by colors (Fig. 5).

Fig. 3 Trajectories visualization using line coding with the hashtags #Covid19 from 1 to 15 May
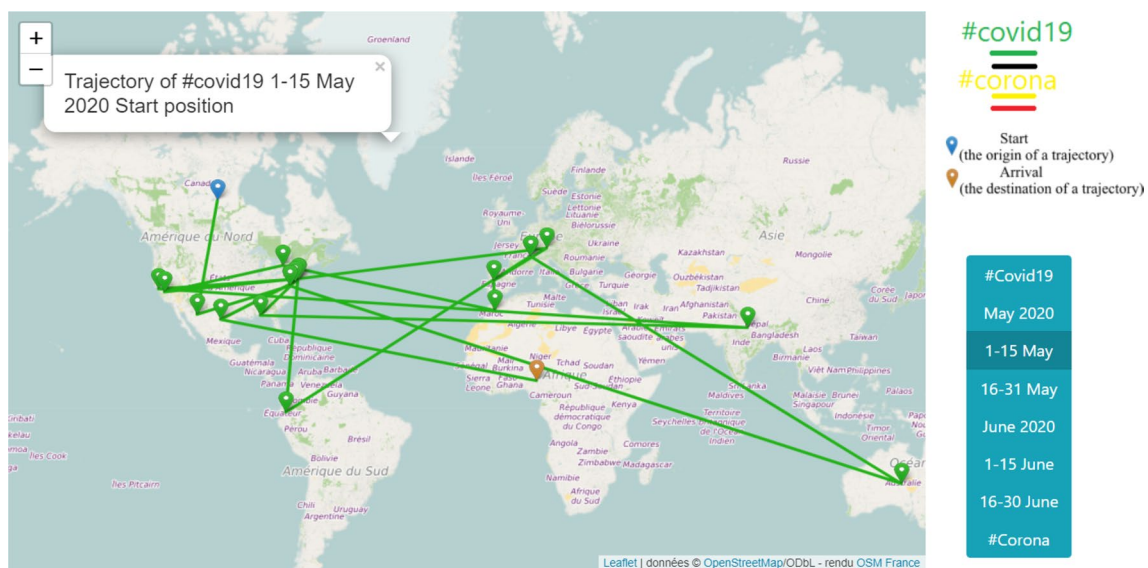

Fig. 4 Trajectory visualization of #Covid19 hashtag

Unlike a Treemap, which uses the size of the box to represent a qualitative value and a location to represent hierarchical relationships, it represents each item in the data set as an equal-sized cell. The colors of squares represent a quantitative value in comparison to other heat map cells, whereas the location may represent the sorting of another quantitative or categorical value. This allows the user to see all data items at the same time.

A heat map represents a large number of data points in a way that traditional tables or graphs would find cumbersome and difficult to interpret. In our case, the heat map is used to depict the density of trajectory flows in space. Then we employ a color grid (from blue for the least dense point to red for the most dense ones).

4.2.3 Edge bounding

The hierarchical grouping of edges (Zhou et al. 2013; Holten and Van Wijk 2009; Hu 2005; Wattenberg 2002) deals with the contiguity relations between the entities organized in hierarchy. Nodes in our case represent origins and destinations (countries), curves represent trajectories,
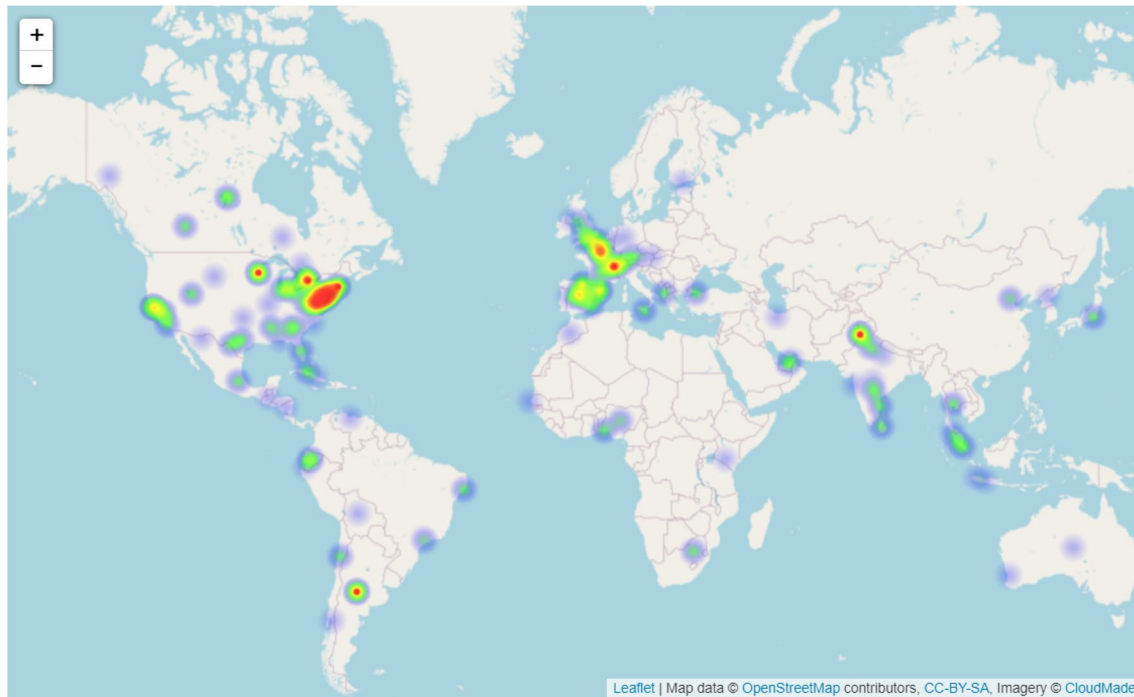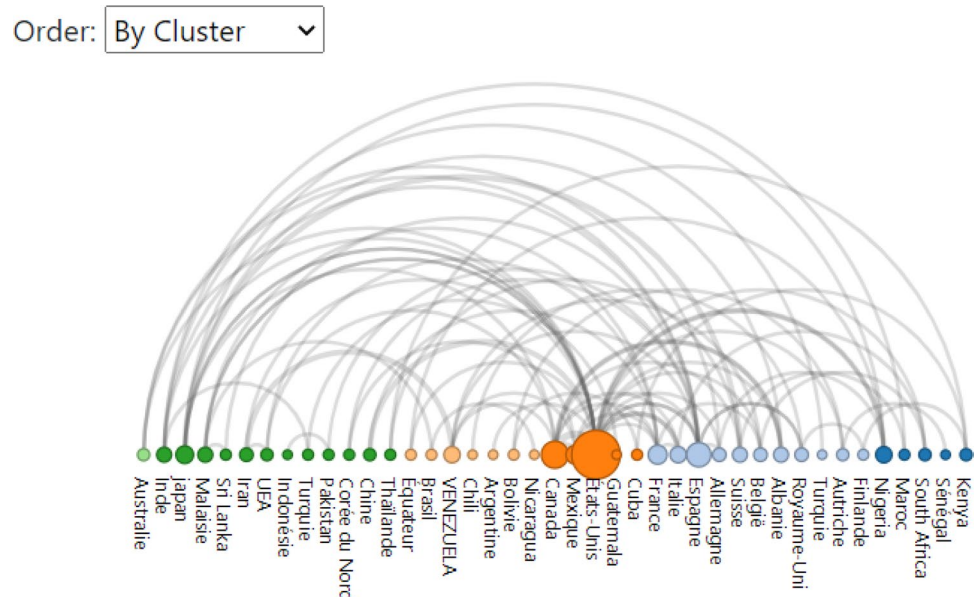
Fig. 5 Trajectories visualization of #Covid19 and #Corona hashtags using heat map

Fig. 6 Trajectories visualiza-
tion of #Covid19 and #Corona
hashtags using Arc Diagram

and colors in nodes represent continents. Finally, the number of trajectories between locations is represented by the thickness of the lines (links) (countries). The edge bounding aggregates trajectories based on the number of occurrences within and between countries (**E1, E3**). In each view, we include interactions (**E5**) such as showing in/out trajectories in the selected country.

4.2.4 One dimension: arc diagram

A one-dimensional graph network is represented by an arc diagram (Wattenberg 2002). It is made up of nodes that describe entities and links that show how entities are related to one another. Nodes are showcased along a single axis in arc diagrams, and links are expressed by arcs. The frequency
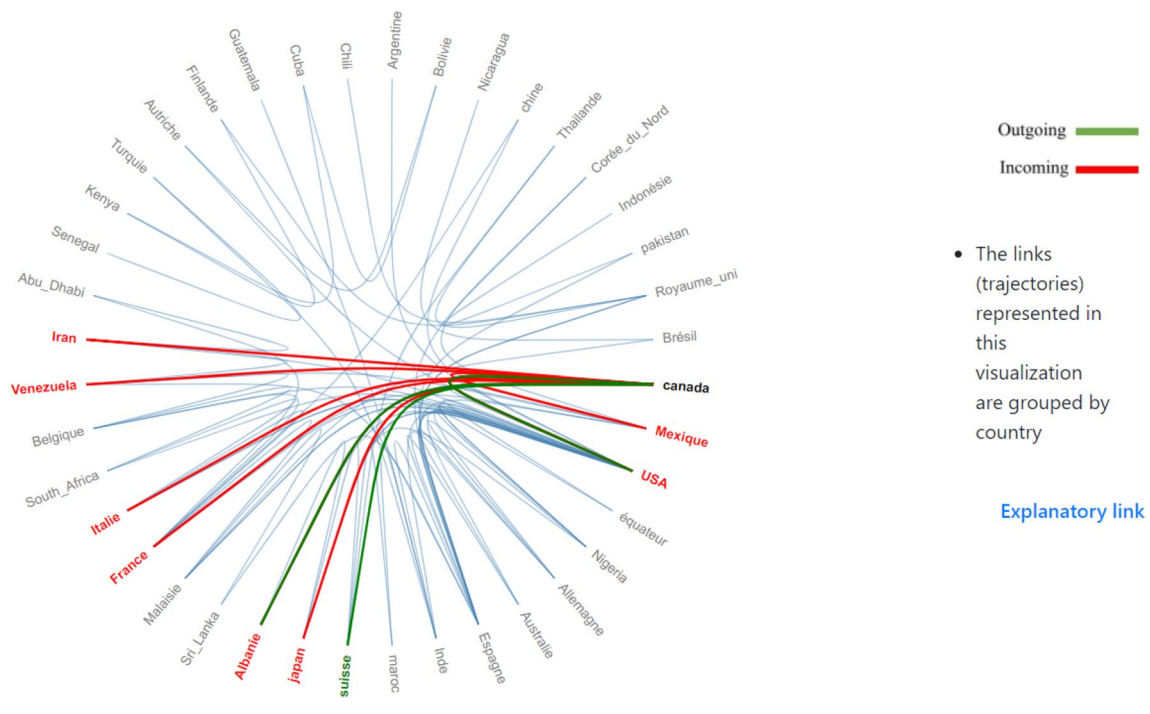
Fig. 7 Trajectories visualization of #Covid19 and #Corona hashtags using Circular Force graph

between the source and target nodes can be represented by the thickness of each arc line. It can be helpful in locating co-occurrences in data. There are three ways to arrange it: by group (continent), frequency, and name (alphabetical order) (Fig. 6).

4.2.5 Two dimension

Circular Force graph (Zhou et al. 2013) connects adjacency edges to reduce the overlap that is common in complex networks. Dependency curves are routed between the source and target nodes along the tree path. Colors are used to distinguish the incoming and outgoing flows from the trajectories (The red color for the incoming flows, and the green color for the outgoing flows) (Fig. 7).

4.2.6 Force-directed graph

The force graph algorithm (Holten and Van Wijk 2009) is an InfoVis field that is used to visually represent entities and relationships (graphs). Their goal is to place the graph nodes in two-dimensional or three-dimensional space with as many edges of equal length as possible and as few crossed edges as possible. Forces are assigned to a set of edges and a set of nodes (based on their relative positions) and then used to simulate the motion of the edges and nodes or to minimize their energy. We add the volume of the nodes to represent

the frequency of the trajectories (incoming and outgoing) (Fig. 8).

4.2.7 OD-matrix

The OD matrix (Guo 2007) can be used to describe the flows of trajectories between a finite set of places. Flows are frequently represented in a matrix by a color palette. Interactions (**E5**) observed between places are frequently relevant for designing, operating, and improving the systems that have been developed. We assign a unique value to each cell as well as a color that represents the quantity (number) of trajectories (the dark color represents a large number of trajectories)(**E1, E3**) (Fig. 9).

4.2.8 3D time statistic

Histogram Young et al. (2011) is a popular statistical chart. It is useful for comparing data categories and displaying data evolution over time. The data is represented by a histogram as bars (the most common), cylinders, cones, and even images, with the height proportional to the purpose of representation.

Histograms in 3D have an advantage because they have one more dimension, allowing us to represent three variables (X, Y, Z) instead of two. In this manner, trajectory data (Fig. 10) is presented as: X the times, Y the locations, and Z the number of trajectories in Y at the X time. Then, to
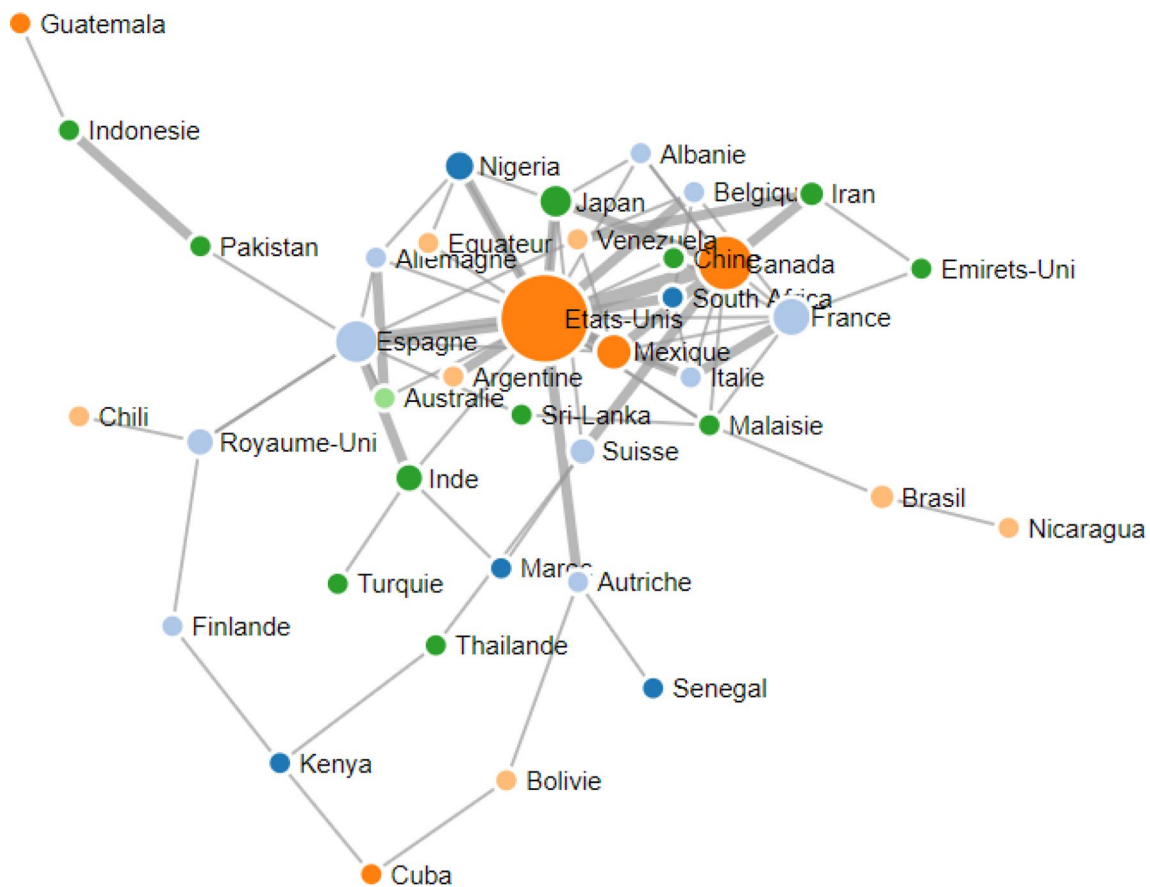
Fig. 8 Trajectories visualization of #Covid19 and #Corona hashtags using Force-directed graph
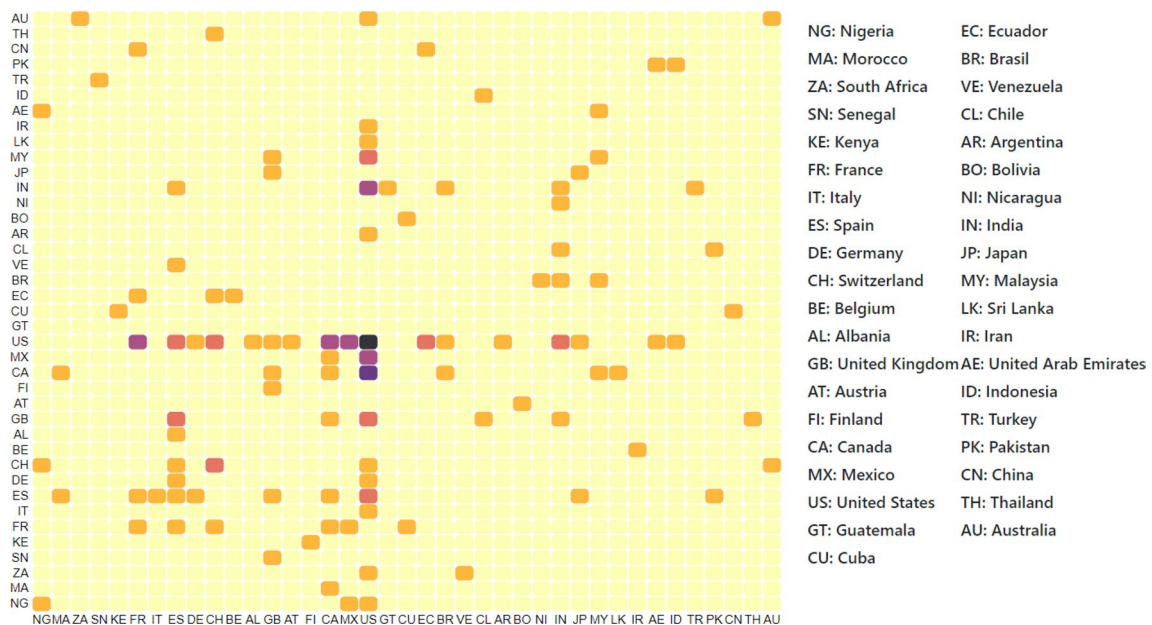

NG: Nigeria	EC: Ecuador
MA: Morocco	BR: Brasil
ZA: South Africa	VE: Venezuela
SN: Senegal	CL: Chile
KE: Kenya	AR: Argentina
FR: France	BO: Bolivia
IT: Italy	NI: Nicaragua
ES: Spain	IN: India
DE: Germany	JP: Japan
CH: Switzerland	MY: Malaysia
BE: Belgium	LK: Sri Lanka
AL: Albania	IR: Iran
GB: United Kingdom	AE: United Arab Emirates
AT: Austria	ID: Indonesia
FI: Finland	TR: Turkey
CA: Canada	PK: Pakistan
MX: Mexico	CN: China
US: United States	TH: Thailand
GT: Guatemala	AU: Australia
CU: Cuba	

Fig. 9 Trajectories visualization of #Covid19 and #Corona hashtags using OD Matrix

Fig. 10 Global distribution of hashtag #Covid19 and #Corona over the word
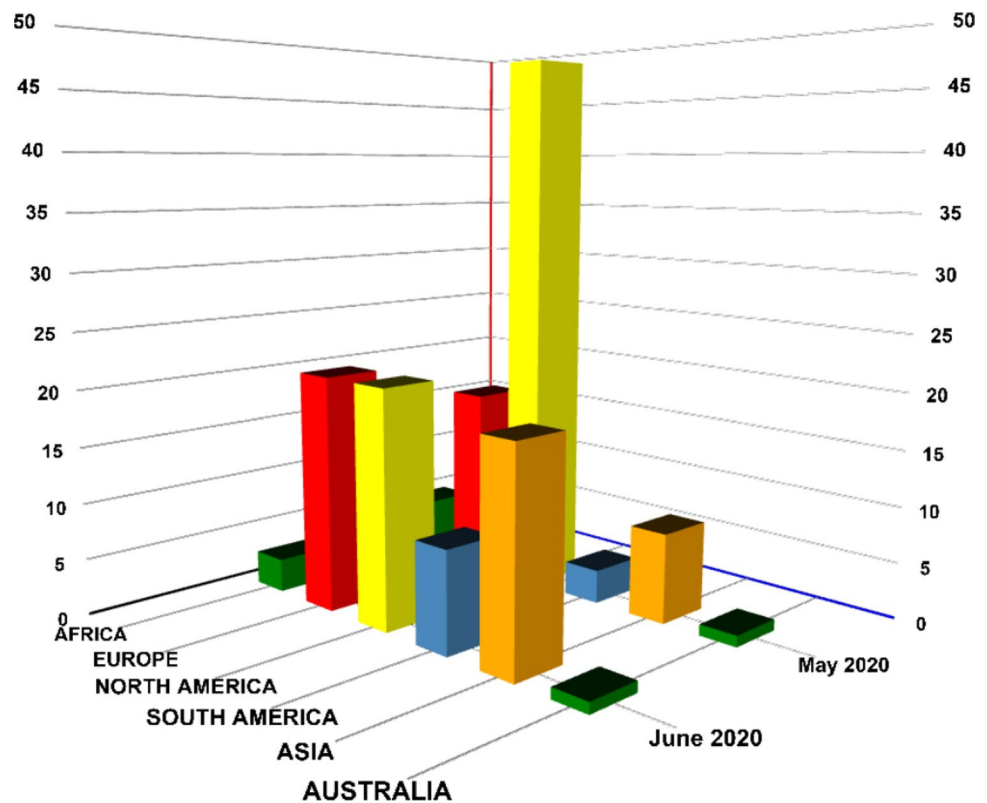
make the representation easier to understand, we assigned a different color to each location in the histogram.(**E1, E3**).

4.2.9 Technical considerations

TrajectoryVis is a representative visualization, where the user interface is implemented using bootstrap.[6] *Trajectory-Vis* views are implemented using d3.js library Bostock et al. (2011). We also used some examples from d3.js and 3D[7] library for represent the bar diagram. We also used Python[8] for Tweet extraction and analytic, and Leaflet[9] for map encoding and draw layouts such as lines and heath maps. The input data are format with JSON, and the user interface is connected to the server using the JQuery library.[10] *Trajec-toryVis* code is available in github link.[11]

---

[6] https://getbootstrap.com/ (accessed on March 2022).

[7] https://github.com/jamesleesaunders/d3-x3D (accessed on March 2022).

[8] https://www.python.org/ (accessed on March 2022).

[9] https://leafletjs.com/ (accessed on March 2022).

[10] https://www.jquery.com (accessed on March 2022).

[11] https://github.com/samfad91/TrajectoryVis_V0 (accessed on March 2022).

5 Evaluation

TrajectoryVis is assessed by a user study and a case study. The data used in it are the hashtags of #Covid19 and #Corona. In this section, we will describe in detail the two evaluations.

5.1 User study

This section describes the assessment protocol and provides information on the testers. We present both the individual and comparative evaluations of each view in *TrajectoryVis*.

5.1.1 Assessment protocol

We have several views in our visualization. Each view serves a distinct purpose by utilizing the same data (Tweet paths) (**E2,3,4**). In this section, we established an evalua-tion protocol based on InfoVis evaluation approaches Lam et al. (2011), Qu and Hullman (2017). The protocol should include the five steps listed below:

1. **Understanding the working environment**: In this step, we must prepare the room in which the users will conduct the assessment. Because *TrajectoryVis* is a web application, all users need is a web browser to access it.

Table 2 Individual assessment of *TrajectoryVis* views

TrajectoryVis views	Usefulness (out of 5)	Usability (out of 5)
OD map	4.14	3.73
Edge bounding	3.99	3.67
OD matrix	3.68	3.27
Graph	3.68	2.91
Arc diagram	3.77	3.59
Heat map	4.05	3.64

Table 3 Comparative evaluation between *TrajectoryVis* Views

TrajectoryVis views	Usefulness (%)	Usability (%)
OD map	68	63
Circle graph	18	32
OD matrix	9	27
2D graph	36	23
Arc diagram	18	23
Heat map	27	45

Finally, we configure users' computers and assign them the test time with the data.

2. **Explanation of data and visualizations**: Following the installation of the room, we should explain the data used in *TrajectoryVis* and their various views (note that we have put links in *TrajectoryVis* to explain views).

3. **Specify the evaluation constraints**: In this section, we created a questionnaire for each view and another to compare them all.

4. **Presentation**: In this part, the users will learn about the test time, and will start the evaluation and answer the questions (normally this part must be filmed).

5. **Interview**: After testing *TrajectoryVis*, we conducted a visualization interview with users to learn about their challenges and suggestions.

We created an online form,[12] and the subsections that follow explain the steps in detail.

5.1.2 Users information

Our solution was tested by 11 men and 11 women ranging in age from 18 to 35 years. The majority are students (17 students, 3 teachers, and 2 IT consultants) from various countries including Algeria, France, and Belgium. The vast majority of users have already used visualization tools. They have a basic understanding of *InfoVis* and are beginners in visualization of trajectories.

They were asked to rate the usefulness and usability of each functionality on a scale of 1 to 5, with 1 being "not useful" / "hard to use," and 5 being "very useful" / "very easy to use." The questionnaire included tasks related to all of the functionalities of all perspectives. We've received some informal comments.

Table 2 shows these results. We discovered that the OD map was the most useful and understandable map when compared to other views. Because of overlapping circles

and the number of lines between circles, graphs are the least popular among users. Although the force graph reduces overlap, it remains a problem for users.

5.1.3 Comparative evaluation between *TrajectoryVis* Views

To compare the proposed visualizations, we asked the testers to answer a series of questions and choose which one is the most useful, the easiest to understand, the easiest to use, and which one represents better trajectories. Table 3 summarizes the results of users ranging from 0 to 100%. Users (who have no experience with *InfoVis*) choose OD maps for individual trajectories and heatmaps or graphs for aggregated ones, as expected.

5.1.4 Discussion

Based on the evaluation results, we can conclude that there is no better visualization of all criteria; each one has strengths and weaknesses. For example, while visualization bylines is the most user-friendly, it does not provide more information. Edge Bundling is the easiest to understand, but it is not the easiest to use. The OD matrix is the best in terms of information density, but it is not the best in terms of trajectory representation, etc.

5.1.5 Testers improvement notes

Our work is only the beginning. It can take several twists and turns based on feedback from testers and our own ideas; above are a few examples:

- Automation and integration of the Tweet collection and processing process.
- Test other more efficient approaches to visualize trajectories on Twitter.
- Test the solution on other social networks, such as Facebook or Uber, or use multiple data sources concurrently.
- Generalization of the tool (*TrajectoryVis*), so that it can visualize the trajectories of hashtags from different social networks.
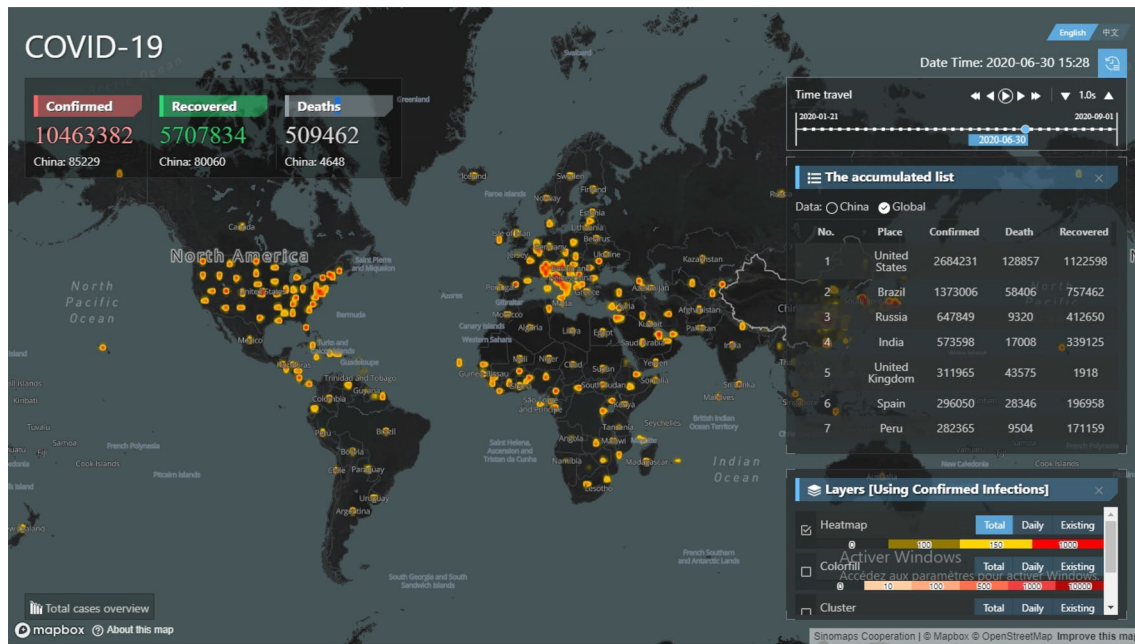
Fig. 11 Corona virus propagation using *coronavirusmap*

5.2 Case study

In this section, we will see a detailed study case named: "Coronas virus". First, we will discuss the study's objective and the tasks that must be completed. Second, we will describe and detail each task that must be completed.

5.2.1 Corona virus hashtag trajectory

The purpose of the case study is to determine whether *TrajectoryVis* provides genuine assistance to people working in the field, such as those at WHO. We examine the WHO results for the months of May and June 2020. The results of these analyses of the WHO visualization are then compared to the analyses of the trajectories in *TrajectoryVis*. We examine the data in three contexts: spatial, temporal, and aggregated spatial data.

5.2.2 Task 1: spatial information

The goal is to see the spread of hashtags (#Covid19, #Corona) in this period using spatial information from the OD map and circular edge bundling views. The spread of hashtags is then compared to the spread of the virus.

Fig. 11 depicts a visual representation of the spread of the Coronavirus. We can see that the virus spread to all continents of the world between May and June 2020. We can see from the "*coronavirusmap*"[13] that the density was higher in the continents of North America and Europe than in the other continents. On the other hand, by analyzing the *TrajectoryVis* (Fig. 12) visualizations OD map, circular edge bundling, we can achieve the same results as the previous visualization, i.e., the propagation of trajectories hashtags reached all continents of the world.

Discussion: The density of trajectories was higher in North America and Europe than in the other continents. As a result, we can conclude that the spread of the hashtags #Covid19 and #Corona is very similar to the spread of the Coronavirus. As a result, the spread of the Coronavirus has an impact on the spread of hashtags (#Covid19 and #Corona).

5.2.3 Task 2: temporal information

We analyzed data in each period of time using temporal information (Histogram). The hashtag trajectories help us determine which countries are the most frequented. Then we compare them to the virus's most infected countries during the same time period.

Fig. 13.1 illustrates the number of people infected with the Corona virus in each country during the period May 2020. We can deduce the following from this graph: The following countries have the highest number of infected
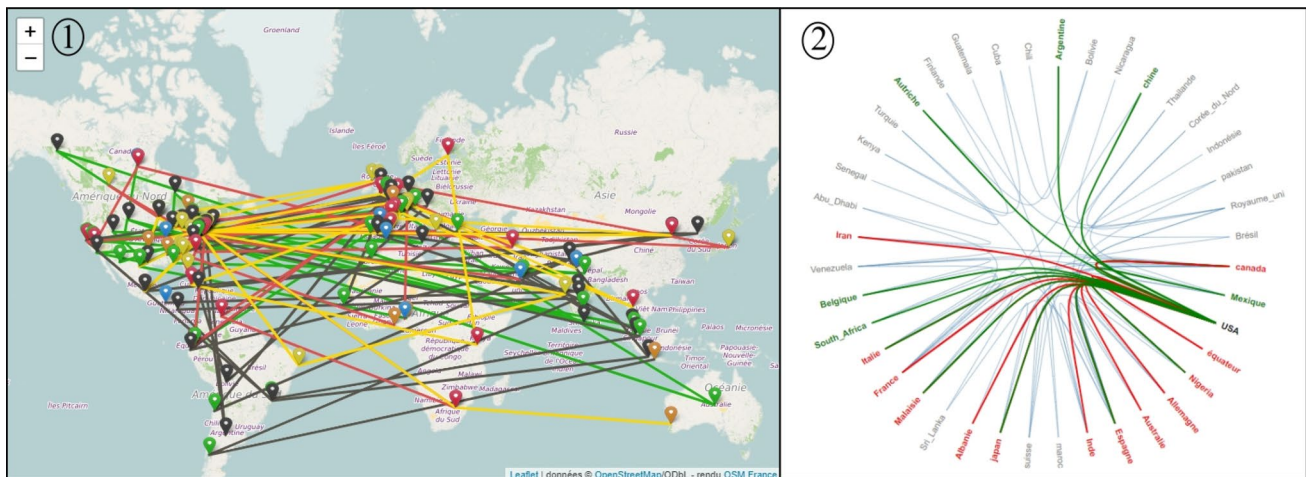

Fig. 12 Individual trajectories visualization using (1) OD map and (2) circular Edge Bundling
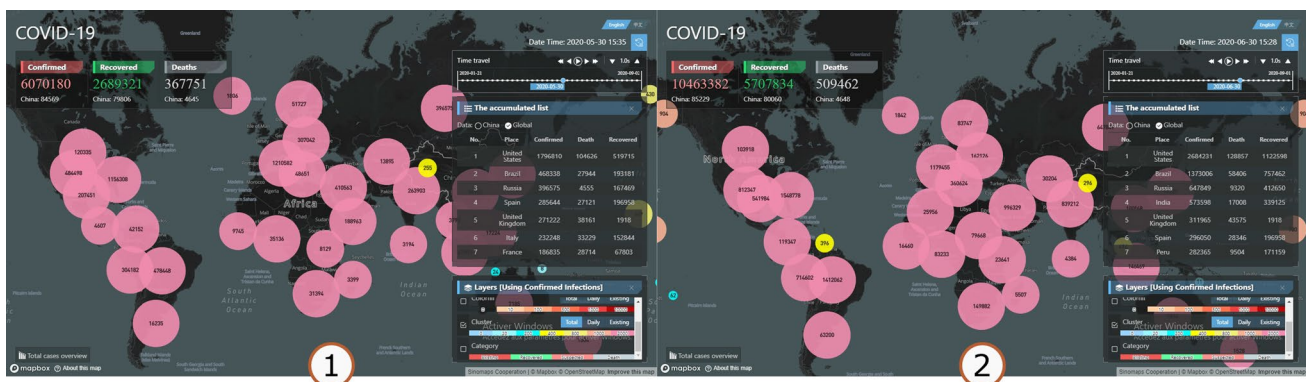

Fig. 13 Coronavirus propagation from May (1) to June (2)

people: the United States, Canada (in North America), Brazil, Ecuador, Venezuela, and Colombia (in South America), Spain, France, Italy, and the United Kingdom (in Europe), Russia, India, and Iran (in Asia), Egypt, and South Africa (in Africa). Fig. 13.2 depicts the number of people infected with the Coronavirus in each country during the period June 2020. As a result, the following countries have the highest number of infected people: the United States, Mexico (in North America), Brazil, Chile (in South America), Spain, France, Germany, the United Kingdom, Belgium (in Europe), Russia, India, Pakistan (in Asia), Egypt, South Africa, Nigeria, and Togo (in Africa).

The histograms in Fig. 14 show that the countries most frequented by the trajectories during the period May 2020 are: France, Spain, Switzerland (in Europe), the United States, Canada (in North America), Ecuador, Venezuela (in South America), Japan, Malaysia (in Asia), and Nigeria (in Africa). During the period June 2020, the busiest countries are: the United Kingdom, Spain, France (in Europe), the United States, Mexico (in North America), Brazil,

Chile (in South America), India, Pakistan, and Indonesia (in Asia).

Discussion: We can conclude that the results of the statistical analysis presented in *TrajectoryVis* are similar. As a result, the countries most affected by the Coronavirus are concentrated in North America and Europe. The lack of similarity in Africa is due to a lack of data (Tweets) available in this continent due to its limited use of Twitter.

5.2.4 Task 3: spatial aggregations

We use aggregations to find the countries with the most frequented trajectories and compare them to the other countries with the highest coronavirus infection rates.

Fig. 15 depicts the evolution of Corona virus-infected cases from May to June 2020 using a line diagram. We can see that the country with the most cases during this time period is the United States. We can observe that the most frequented country by hashtag trajectories (#Covid19, #Corona) is the United States by ordering the Arc diagram

Fig. 14 *TrajectoryVis* Histogram of all continent in the period from May (left) to June (right): (1) Global trajectories over the world. (2) South America. (3) Asia. (4) Europe. (5) Africa. (6) North America. (7) Australia
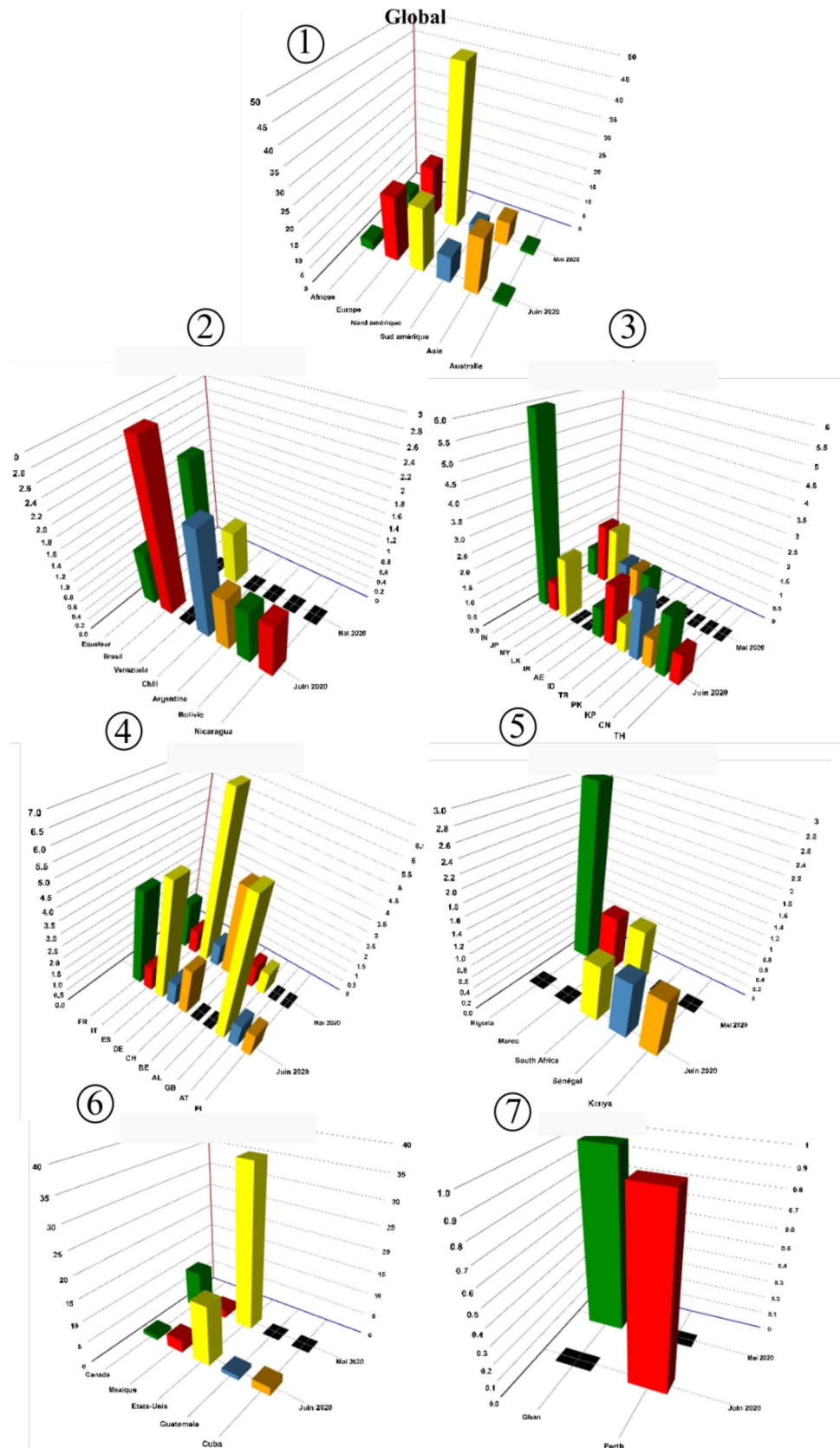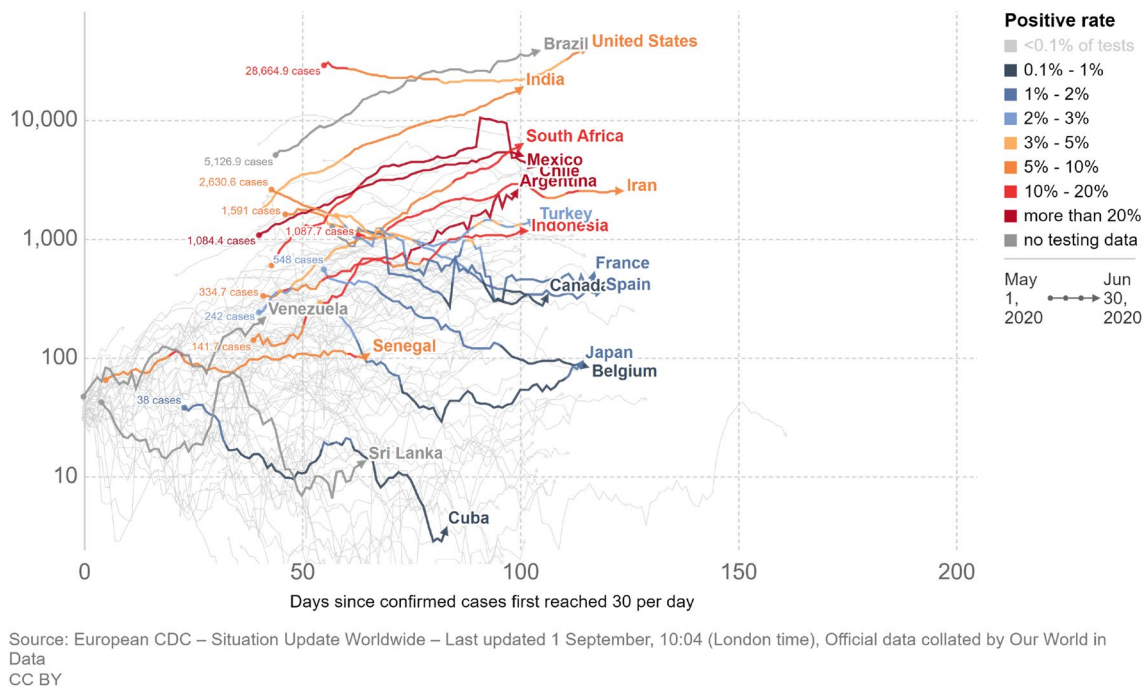
Fig. 15 Evolution of Corona virus cases by country

visualization of *TrajectoryVis* by frequency (Fig. 16.3) and analyzing the Heat map, Graph visualizations of *TrajectoryVis* (Fig. 16.1, 2).

Discussion: The country most frequented by the hashtag trajectories #Covid19 and #Corona is also the country most infected by the Coronavirus. As a result, we can identify that there is a strong resemblance between the real world and the virtual world (social networks). As a result, *TrajectoryVis* will serve as a good tool for the WHO (World Health Organization). *TrajectoryVis* can help WHO analyze the Coronavirus situation. It also assists them in determining which countries are the most affected by this virus.

5.3 General discussion

Russia does not appear in the trajectories of *TrajectoryVis* because the Russian government has launched administrative procedures[14] against social networks Twitter and Facebook, accusing them of failing to comply with Russian law, which requires that the data of Russian users be stored on national territory. The companies that operate the social networks

Facebook and Twitter have provided formal responses to their requests to confirm the location of personal data of Russian users in Russia. This is why Russia does not appear on our maps. The lack of data (Tweets) in Africa is due to the continent's limited use of Twitter (and lack of location information). This explains why its countries have no hashtag trajectories.

6 Conclusion

In this paper, we propose a state of the art of the spatio-temporal and trajectory data. We also compare various visual techniques for depicting individual and aggregated trajectories. Furthermore, we present TrajectoryVis, a visual approach that allows users to explore trajectories using various visual techniques. Finally, we evaluated this tool through a user and case study.

7 Future work

TrajectoryVis can handle large datasets, but the data collected from processed Tweets is not one of them. *TrajectoryVis* is capable of representing a large dataset using

---

[14] https://www.24heures.ch/monde/russie-s-attaque-facebook-twitter/story/21257698 (accessed on March 2022).
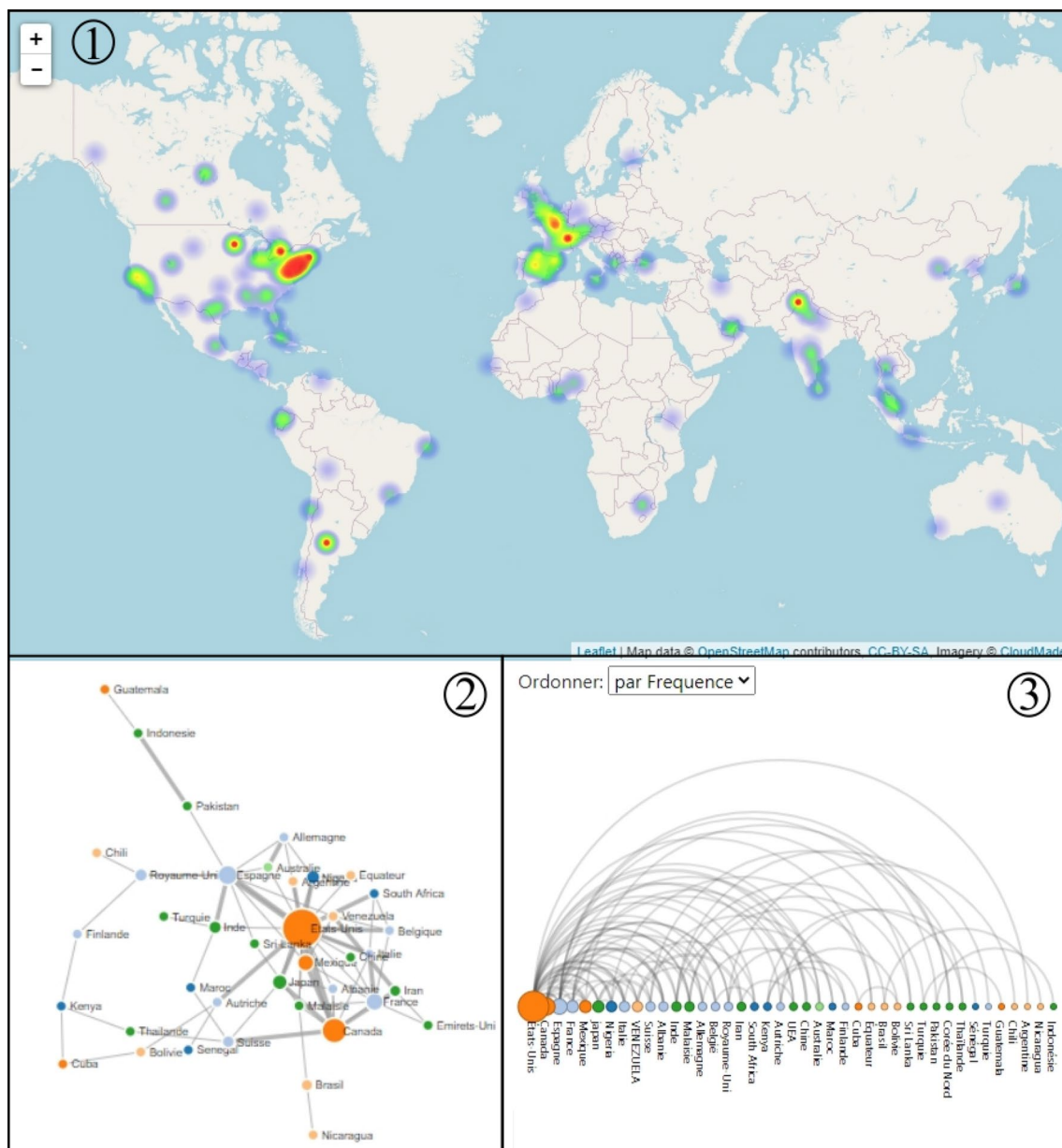
Fig. 16 Comparison of different aggregation trajectories using heat map (1) and one/two D graphs (2, 3)

aggregated or individual visual encoding views. Individual visual encodings, such as line maps, can have line overlapping. Increasing the dataset size in graphs produces a surcharging representation with links and nodes, similar to graphs such as arc diagram, edge bounding, and force graph. We intend to introduce and apply algorithms for reducing overlapping (nodes and links) in graphs and line maps in the future. In addition to using data from various sources such as Facebook, reedit...etc. We aim to highlight our tool as a generic one that represents trajectories for free.

References

Abbott M (2013) Visualising a temporal cartography of travel. In: Geospatial visualisation. Springer, pp 3–17

Adrienko N, Adrienko G (2010) Spatial generalization and aggregation of massive movement data. Trans Visualization Comput Graphics 17(2):205–219

Aigner W, Miksch S, Schumann H, Tominski C (2011) Visualization of time-oriented data. Springer Science & Business Media

Andrienko G, Andrienko N (2008) Spatio-temporal aggregation for visual analysis of movements. In: Symposium on visual analytics science and technology, 51–58. IEEE

Bostock M, Ogievetsky V, Heer J (2011) D3: Data-driven documents. Trans Vis Comput Graphics 17(12):2301–2309

Card SK, Mackinlay JD, Shneiderman B (1999) Information visualization. Morgan Kaufmann Publishers Inc., pp 1–34

Crnovrsanin T, Muelder C, Correa C, Ma KL (2009) Proximity-based visualization of movement trace data. In: Symposium on visual analytics science and technology, pp 11–18. IEEE

Cuenca E, Sallaberry A, Wang FY, Poncelet P (2018) MultiStream: a multiresolution streamgraph approach to explore hierarchical time series. Trans Vis Comput Graphics 24(12):3160–3173

Ding L, Yang J, Meng L (2015) Visual analytics for understanding traffic flows of transport hubs from movement data. In: Proceedings of the international cartographic conference, pp 19–21

Ester M, Kriegel HP, Sander J, Xu X et al (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the international conference on knowledge discovery and data mining 96:226–231

Fekete J, van Wijk JJ, Stasko JT, North C (2008) The value of information visualization. In: Information visualization - human-centered issues and perspectives, 1–18. Springer

Guo D (2007) Visual analytics of spatial interaction patterns for pandemic decision support. Int J Geogr Inf Sci 21(8):859–877

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: Proceedings of the pacific visualization symposium, pp 163–170. IEEE

Gupta S, Dumas M, McGuffin MJ, Kapler T (2016) Movementslicer: better gantt charts for visualizing behaviors and meetings in movement data. In: Proceedings of the pacific visualization symposium, pp 168–175. IEEE

Holten D, Van Wijk JJ (2009) Force-directed edge bundling for graph visualization. In: Computer graphics forum, 28, 983–990. Wiley Online Library

Hu F, Li Z, Yang C, Jiang Y (2019) A graph-based approach to detecting tourist movement patterns using social media data. Cartogr Geogr Inf Sci 46(4):368–382

Hu Y (2005) Efficient, high-quality force-directed graph drawing. Math J 10(1):37–71

Kim S, Jeong S, Woo I, Jang Y, Maciejewski R, Ebert DS (2017) Data flow analysis and visualization for spatiotemporal statistical data without trajectory information. Trans Vis Comput Graphics 24(3):1287–1300

Krueger R, Sun G, Beck F, Liang R, Ertl T (2016) Traveldiff: visual comparison analytics for massive movement patterns derived from twitter. In: Proceedings of the pacific visualization symposium, pp 176–183. IEEE

Lam H, Bertini E, Isenberg P, Plaisant C, Carpendale S (2011) Empirical studies in information visualization: seven scenarios. Trans Vis Comput Graphics 18(9):1520–1536

Li J, Chen S, Chen W, Andrienko G, Andrienko N (2018) Semantics-space-time cube. a conceptual framework for systematic analysis of texts in space and time. Transactions on Visualization and Computer Graphics

Liu Y, Heer J (2018) Somewhere over the rainbow: an empirical assessment of quantitative colormaps. In: Proceedings of the conference on human factors in computing systems, pp 598:1–598:12

Maahs TD, Ewers RC, Chang AT, Rothe C, Chen EC, Elmer MC, Madrid G, Gambhir K (2012) Methods and apparaus for off-axis visualization. US Patent 8,277,373

Miller HJ (1999) Potential contributions of spatial analysis to geographic information systems for transportation (gis-t). Geogr Anal 31(4):373–399

Peuquet D (1994) It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems. Annals Am Assoc Geogr 84(3):441–461

Phan D, Xiao L, Yeh R, Hanrahan P (2005) Flow map layout. In: Proceedings of the symposium on information visualization, pp. 219–224. IEEE

Qu Z, Hullman J (2017) Keeping multiple views consistent: constraints, validations, and exceptions in visualization authoring. Trans Vis Comput Graphics 24(1):468–477

Rettore PH, Santos BP, Lopes RRF, Maia G, Villas LA, Loureiro AA (2020) Road data enrichment framework based on heterogeneous data fusion for its. IEEE Trans Intell Trans Syst 21(4):1751–1766

Ryoo M, Kim N, Park K (2018) Visual analysis of soccer players and a team. Multimed Tools Appl 77(12):15603–15623

Scheepens R, van de Wetering H, van Wijk JJ (2014) Non-overlapping aggregated multivariate glyphs for moving objects. In: Proceedings of the pacific visualization symposium, pp. 17–24. IEEE

Schöttler S, Yang Y, Pfister H, Bach B (2021) Visualizing and interacting with geospatial networks: a survey and design space. Comput Graphics Forum 40(6):5–33

Senaratne H, Bröring A, Schreck T, Lehle D (2014) Moving on twitter: using episodic hotspot and drift analysis to detect and characterise spatial trajectories. In: A. Pozdnoukhov, S. Xu (eds.) Proceedings of the international workshop on location-based social networks, pp. 23–30. ACM

Wallner G, Halabi N, Mirza-Babaei P (2019) Aggregated visualization of playtesting data. In: Proceedings of the conference on human factors in computing systems, pp. 1–12

Ward MO, Grinstein G, Keim D (2010) Interactive data visualization: foundations, techniques, and applications. A K Peters

Wattenberg M (2002) Arc diagrams: visualizing structure in strings. In: Proceedings of the symposium on information visualization, pp. 110–116

Willems N, van Hage WR, de Vries G, Janssens JH, Malaisé V (2010) An integrated approach for visual analysis of a multisource moving objects knowledge base. Int J Geogr Inf Sci 24(10):1543–1558

Wood J, Dykes J, Slingsby A (2010) Visualisation of origins, destinations and flows with od maps. Cartogr J 47(2):117–129

Yang L, Kwan MP, Pan X, Wan B, Zhou S (2017) Scalable space-time trajectory cube for path-finding: a study using big taxi trajectory data. Trans Res Part B: Methodol 101:1–27

Young FW, Valero-Mora PM, Friendly M (2011) Visual statistics: seeing data with dynamic interactive graphics, vol 914. John Wiley & Sons

Zhou H, Xu P, Yuan X, Qu H (2013) Edge bundling in information visualization. Tsinghua Sci Technol 18(2):145–156

Zhu X, Guo D (2014) Mapping large spatial flow data with hierarchical clustering. Trans GIS 18(3):421–435