Flowstrates: An Approach for Visual Exploration of Temporal Origin-Destination Data

Ilya Boyandin[1], Enrico Bertini[2], Peter Bak[3] and Denis Lalanne[1]

[1]University of Fribourg, Switzerland
[2]University of Konstanz, Germany
[3]IBM Haifa Research Lab, Israel

Abstract

*Many origin-destination datasets have become available in the recent years, e.g. flows of people, animals, money, material, or network traffic between pairs of locations, but appropriate techniques for their exploration still have to be developed. Especially, supporting the analysis of datasets with a temporal dimension remains a significant challenge. Many techniques for the exploration of spatio-temporal data have been developed, but they prove to be only of limited use when applied to temporal origin-destination datasets. We present **Flowstrates**, a new interactive visualization approach in which the origins and the destinations of the flows are displayed in two separate maps, and the changes over time of the flow magnitudes are represented in a separate heatmap view in the middle. This allows the users to perform spatial visual queries, focusing on different regions of interest for the origins and destinations, and to analyze the changes over time provided with the means of flow ordering, filtering and aggregation in the heatmap. In this paper, we discuss the challenges associated with the visualization of temporal origin-destination data, introduce our solution, and present several usage scenarios showing how the tool we have developed supports them.*

Categories and Subject Descriptors (according to ACM CCS): D.2.2 [Design Tools and Techniques]: User interfaces—H.5.2 [User Interfaces]: Interaction styles—

1. Introduction

Numerous datasets representing entities moving between geographical locations are being produced nowadays. Many of these datasets are collected in the form of origin-destination data (or "spatial interactions"), meaning that only the origins and the destinations of the flows and the flow magnitudes are known, but not the exact movement routes. The analysis of geographical movement data is particularly important, since "much *change* in the world is due to geographical movement" [Tob05]. However, the full potential of such data remains largely unreleased, because tools and techniques which support their exploration, and especially the analysis of the changes over time, still have to be developed [Rae09, AAD*10, MGLS97]. Making a visualization tool which can support the analysis of temporal origin-destination data is the main focus of our work.

One of the most widely used representations of origin-destination data are flow maps [Tob87, Har99]. These are visualizations which represent entities flowing between geo-graphical locations on a map overlaid with lines connecting the flow origins with the destinations. Flow maps are aimed to support analysts in finding answers to spatial questions such as: Where are the largest and the smallest flows? Where on the map are the origins and the destinations? In which direction do the migrants go? What is happening in a specific location?

However, additional important questions and tasks which flow maps were not designed for arise when exploring datasets with a temporal dimension. These tasks are concerned with the analysis of both the spatial and temporal dimensions of the data and the relationships between them.

The requirement for the technique we are developing is to support both spatial and spatio-temporal tasks (see the full list of the tasks in section 2.1). For the spatial tasks the technique has to make use of a representation on a geographic map. Being well familiar to everybody, maps allow to reason about the geographic patterns of the movement as opposed to non-geographic representations (more details in section 3.1).

However, it is hard to augment a geographic map with temporal origin-destination data in a meaningful and readable way. There are several alternatives to embedding the temporal data directly into a flow map, e.g. small multiples, animation, or creating an abstract view. They are discussed in detail in section 2.2.

In this paper we propose Flowstrates, a "hybrid" solution which brings together a geographic and a temporal representation and overcomes some of the deficiencies of the above-mentioned approaches providing means for the analysis of spatio-temporal patterns in origin-destination datasets. Our technique is discussed in more detail in section 3.

2. Problem definition, tasks and design alternatives

The problem we address in this work is focused on the representation of a specific, though very common, type of data. Origin-destination data is a collection of flows of entities between geographic locations, where each flow is characterized by the following features:

- Origin: a geographic location (e.g. a country or a city)
- Destination: a geographic location
- Magnitude: a numerical value characterizing the number or the amount of entities flowing from a specific origin to a specific destination
- Type: a nominal value describing the type of the entities flowing (e.g. people, men or women, types of goods, etc.)
- Time: a precise date, or a time period (e.g. year) during which the flow magnitude was measured

Our goal is to find a way of representing such data in an integrated and natural fashion in order to help data analysts to explore and analyze them efficiently.

2.1. User tasks

In this section we discuss the list of user tasks which we selected for our tool to support, deriving it from our discussions with practitioners and researchers, from analytic reports [UNH10,KP99], other visual analysis task taxonomies (which are discussed below), tasks which existing related techniques (in particular, flow maps [Har99]) support, as well as from our own experience with the analysis of various datasets [BBL10]:

- T1: Finding where the largest and the smallest flows are for a specific time period
- T2: Locating the origins and the destinations of the flows and determining their directions for a specific time period
- T3: Examining possible effects on the neighborhood of what was happening in a specific location in a specific time period
- T4: Examining the "big picture" (an overview of the whole dataset) or focusing on a region or a specific location for a specific time period

- T5: Comparing flows between different origin-destination pairs or comparing flows of two locations for a specific time period
- T6: Examining changes over time for a specific location
- T7: Examining a specific time range (over several periods) in a specific location
- T8: Finding when there were peaks/significant changes in the flow magnitudes for a specific origin/destination pair
- T9: Comparing temporal changes of the flows of two locations

These tasks are compatible with the taxonomy for the analysis of spatio-temporal data proposed by Andrienko [AA06] in which our tasks can be classified as lookup (T2, T7), pattern identification (T1, T4, T6, T8), comparison (T5, T9), and relation-seeking (T3). When compared to the low-level analysis task taxonomy proposed by Amar et al in [AES], there is also a mapping between the tasks: Retrieve Value (T2), Find Extremum or Anomalies (T1, T8), Characterise Distribution (T4, T6, T7), Correlate (T3, T5, T9). The Filter, Sort and Cluster tasks are not in our list, but they are very much related to the higher-level T1, T8 and T4. We deliberately left out graph topology tasks [LPP*06], as we wanted to concentrate on the relationships between the flow origins, destinations, and the changes over time of the flow magnitudes.

Our tasks are less general than the above-mentioned taxonomies, but they are more focused on the problem of the analysis of one specific type of data which we address, and are therefore better suited for providing baseline requirements for our work.

2.2. Design alternatives

Considering the tasks listed in the previous section, we can conclude that the major challenge is how to bring together the spatial and temporal dimensions in a way which makes it possible to explore the relationships between these two aspects of the data. In this section, we discuss what we believe are four important design alternatives to address this problem. This is not meant to be a thorough discussion of the prior work in the domain (which you can find in section 6), but rather an overview of the approaches which can be used with the currently available techniques and which we considered for our implementation.

Small multiples. Different time periods can be represented in separate flow maps and put next to each other in a small multiples display to allow to see the changes over time. Small multiples can be immensely useful for representing changes for a relatively small number of objects or when displaying images for a small number of time periods [AA06]. One of our requirements, though, reflected in the task T4, was to provide an overview for longer time periods. Also, because of the smaller size of the individual images it is difficult to see the details and to compare flows between the
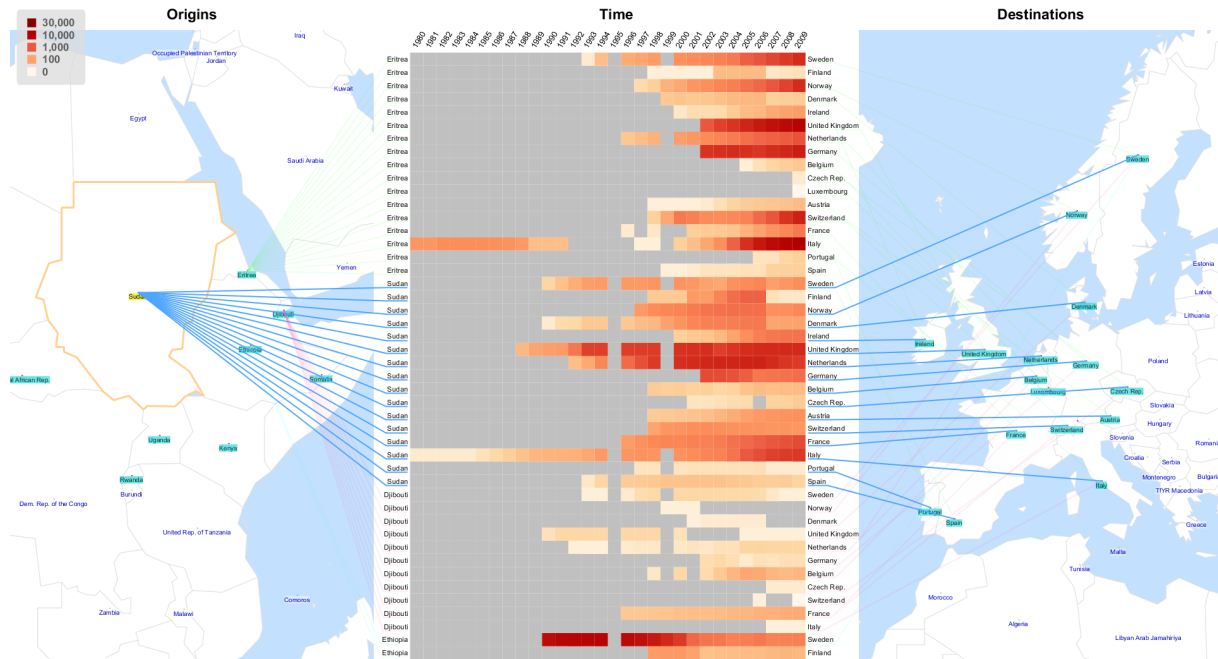
Figure 1: *Flows of refugees are shown between East Africa and Western Europe. Flows having their origin in Sudan are highlighted. The heatmap shows the flow magnitudes by year and origin-destination. By following the lines of the heatmap it is possible to see the flows' origins, destinations and the changes of the magnitudes over time. Different temporal patterns are visually salient, such as a consistently high number of refugees from Sudan to the United Kingdom and the Netherlands, a marginal decrease to Denmark, Norway and Germany, and an increase to Ireland and Italy.*

years in small multiples of flow maps [BBL10]. Thus, this solution is not scalable: the more small multiples are represented, the more difficult it is to see the details. Moreover, significant flows which are often short (e.g. for migrations) become even shorter and more difficult to see.

Animated flow maps. Animation can be used to show how flows of subsequent time periods change [BEW95]. In certain situations animations may be more effective than static graphics [APP11, HR07], but not when they are too complex to be accurately perceived [TMB02]. An animated flow map showing thousands of flow lines could hardly be accurately perceived as it would be too difficult to keep track of changes in it.

Embedding temporal data into a flow map. A direct embedding into a flow map would mean representing the temporal changes by mapping temporal data to each of the visual features of the flow lines (color, size etc). Such a solution would be feasible for datasets with a very small number of flows, but for a flow map with a substantial number of flows it would only multiply the clutter caused by the line intersections which conventional flow maps already suffer from. A more sophisticated way of embedding might be able to overcome this problem though.

Non-geographic, abstract view. A non-geographic view can present a good overview of the development in time, but questions involving the spatial arrangement can be difficult or impossible to answer. Thus, the spatial and spatio-temporal tasks listed above cannot be fully supported.

Both embedding and non-geographic view have their advantages. Being able to see the flow origins and destinations on a map is important to observe spatial patterns. Using an abstract temporal view allows to better visualize changes over time without having to fit the visualization in a map. Flowstrates, the solution which we propose, takes advantage of these two alternatives bringing them together in a simple yet elegant way.

3. The Flowstrates

In Flowstrates the origins and the destinations of the flows are displayed in two separate maps. As it is not necessary to show the exact flow paths (they are usually not known in origin-destination datasets), we can reroute the flow lines in any way. So we represent the temporal information in an abstract view (a heatmap in which the columns correspond to different time periods) and draw the flow lines so that they connect the flow origins and destinations with the cor-
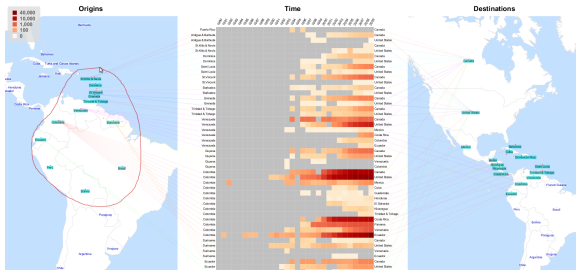
Figure 2: *Selecting origins using lasso: When a selection is made, the heatmap is updated, so that only the flows between the selected origins and destinations are displayed.*
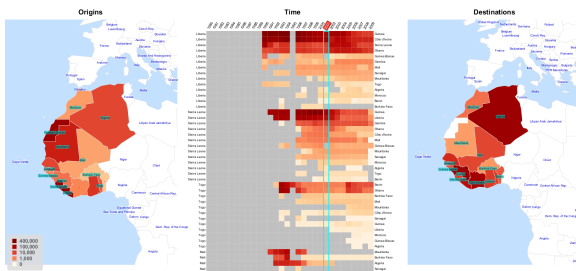

Figure 3: *Selecting a year: Here the year 2001 is selected in the heatmap header, so the countries in the geographic maps are colored according to the total magnitudes of the outgoing and incoming flows in 2001. The heatmap rows are sorted by the maximum (over time) total magnitudes for the origin countries, and by the max magnitude in each row within the same origin country.*

responding rows of the heatmap, as if the flows were going through it (see Fig. 1).

In other words, considering the problem, the tasks we want to support, and the available design alternatives, we made the following design choices:

- Represent locations on geographical maps
- Use two separate maps for origins and destinations
- Show the temporal information in a separate abstract view
- Visually link the geographic and temporal views

3.1. Design considerations

In this section, we give a detailed discussion of the reasons for making the design choices and of their implications.

Why maps? Maps are well familiar to everybody. They allow to reason about the geographic patterns of the movement as no other representation by naturally providing answers to questions such as: "What is the spatial distribution of the locations?", "How far are they from each other?",

"What are the neighbors of a location/area?", "Which areas constitute a region?" etc.

Why two separate maps? Displaying the flow origins and destinations in two different maps allows to:

- clearly show the flow directions (origin → destination)
- use any appropriate representation for the temporal data without being constrained by having to fit it into a map
- focus on different regions for the origins and destinations and perform visual queries for them in two separate maps (see 3.2 for details)
- augment the two maps with aggregated information for both origins and destinations at the same time (e.g. showing the outgoing and incoming totals by coloring the countries).

These advantages come at a price. Compared to a conventional flow map, the distances between the origins and destinations, the flow routes and orientations cannot be naturally visualized in Flowstrates (see section 7). Despite that, the two-map solution is advantageous in situations when these properties of the flows are less important for the analysis than the temporal changes of their magnitudes.

Why links? The idea to show the locations and the temporal changes of the flow magnitudes in separate views and to visually link the corresponding origins and destinations across the views was inspired by the *semantic substrates* approach for the interactive exploration of complex graphs proposed by Shneiderman and Aris in [SA] (see section 6 for a discussion of semantic substrates).

The visual linking can be very useful in some situations. For instance, in Fig. 4 without the links, only with the ability to highlight a row in the heatmap or a country in the maps, it would be only possible to see flows from one origin at a time (when a country is selected in the origins map) or the origin of one flow at a time (when a row is selected in the heatmap). With the links we can clearly see what the origins of a several hundred flows in the heatmap are. Color-coding and coloring countries in the maps and highlighting the corresponding segments in the heatmap instead of drawing lines would also be possible, but then we would be limited in the ability to use coloring to show country totals in the maps (Fig. 3).

Why an abstract temporal view? Separating the geographic and the temporal views allows to present the changes over time of the flow magnitudes in a way which is most suitable for the analysis of temporal patterns. The temporal view can be manipulated by the user, e.g. it can be filtered, reordered, aggregated. Still, the connections between the geographical locations and the rows of the temporal view representing flows are maintained, so that the analysts can track down the relationships between the spatial and temporal aspects of the data. In addition, this clear separation between the spatial and the temporal representations provides flexibility in terms of the initiation of the task. The analyst

can begin the exploration from the temporal view and then use the spatial representation to understand where the events took place. Conversely, the user can begin from a specific region of interest and then isolate the temporal patterns pertaining to the region of interest. Refer to section 4 for more details on the exploration strategies.

Why heatmap? We chose the heatmap as the temporal data representation for two main reasons. First, it can seamlessly represent the temporal changes of the flow magnitudes at different zoom levels, thus, providing support for task T4. Second, the same color scheme as in the heatmap (for the images in this article we chose ColorBrewer's OrRd and RdBu [BH09]) can be used to show the totals of the outgoing and incoming flow magnitudes in the origin and destination maps. Hence, the totals in the geographic maps can be compared to the individual values in the heatmap.

The design of Flowstrates can, however, accommodate a number of alternative temporal views, e.g. multiple time series. Lam et al [LMK07] compared the effectiveness of using multiple line graphs and heatmaps for analyzing overviews over large datasets and found that heatmap was more efficient for finding the maximum values and comparison, but less efficient for finding the graph with the maximum number of peaks. Horizon graphs [HKA09], which are more space-efficient than time series, could be also used in place of the heatmap. It is not clear, though, how well they would support changing the zoom level. Another alternative would be to plot time series of the changing flow magnitudes in a single row as in TimeSearcher [KHS02]. The temporal view would then require much less space vertically, but it would not give a good overview and would make linking it to the geographic maps much more difficult.

3.2. Interaction techniques

Flowstrates are meant for interactive exploration. Unlike OD-matrices which represent exactly one flow in each heatmap cell, in Flowstrates every flow takes the whole row of the heatmap. Thus, much more screen real estate is used to represent the same number of flows. Hence, for many datasets it is impossible to display all the flows simultaneously without filtering or aggregating them. If we want the analysts to still be able to explore the data in every bit of detail, then we need to provide means of interaction for controlling filtering, zooming and aggregation. Currently, our implementation supports the following techniques:

Visual querying and filtering. A subset of locations can be selected in the origin and destination maps (either filtering by name or using the lasso tool, as shown in Fig. 2). When a selection is made, the heatmap is updated, so that only the flows between the selected locations are displayed. Had we used only one map, making a separate selection of origins or destinations directly on the map would probably be more difficult for the user. Due to the separation of the origins and

destinations in Flowstrates, we can provide support for such queries in a straightforward way.

The user can also select a time period. In this case, the outgoing and incoming totals of the regions for this time period are displayed in the maps (Fig. 3).

Zooming and panning. All the three views (the origin map, the heatmap and the destination map) can be zoomed and panned independently. Hence, the user can focus on different regions for the origins and the destinations and select the most relevant part of the heatmap.

Heatmap row ordering. Different ordering strategies are supported: by the maximum/average flow magnitudes, by similarity to a selected heatmap row, or by the geographic positions of the flow origins and destinations. When using the latter ordering method the flows sharing the same origins or destinations are grouped together in the heatmap, so the colored flow lines form "bundles" which are very easy to follow (Fig 4).

Heatmap row aggregation. The flows represented in the heatmap can be aggregated using different grouping functions. They can be grouped, for example, by their origins (so that each heatmap row represents the total magnitudes of the outgoing flows of each of the origin), by destinations, by the geographical regions of the origins or the destinations, or by any other flow attribute. This way we can analyze the data on different aggregation levels, or in other words, change the spatial resolution.

A single static view can rarely give the whole detailed information of the data being analyzed. Most temporal origin-destination datasets are no exception. Hence, providing the users with the appropriate interactive exploration techniques, which allow to analyze the data in every detail by focusing on specific regions of interest, or by performing an automated summarization producing an overview, is very important. This is what we tried to achieve with the interaction techniques which we incorporated in our Flowstrates implementation.

4. Exploration strategies

Flowstrates supports three basic exploration strategies which address the user tasks described in section 2.1. The first two strategies are both concerned with the observation of the patterns in the heatmap and differ in the initiation of the task: from location to time, or from time to location. The last one is about the comparison of either locations or time periods.

S1: Location ⟶ Spatial or temporal pattern. Select a location or a region in the origins map, then find out what is going on in the heatmap or in the destinations map. This strategy supports tasks T6, T7 and T8 (described in 2.1).

S2: Temporal pattern ⟶ Location. Find something interesting in the heatmap, select the time period and then
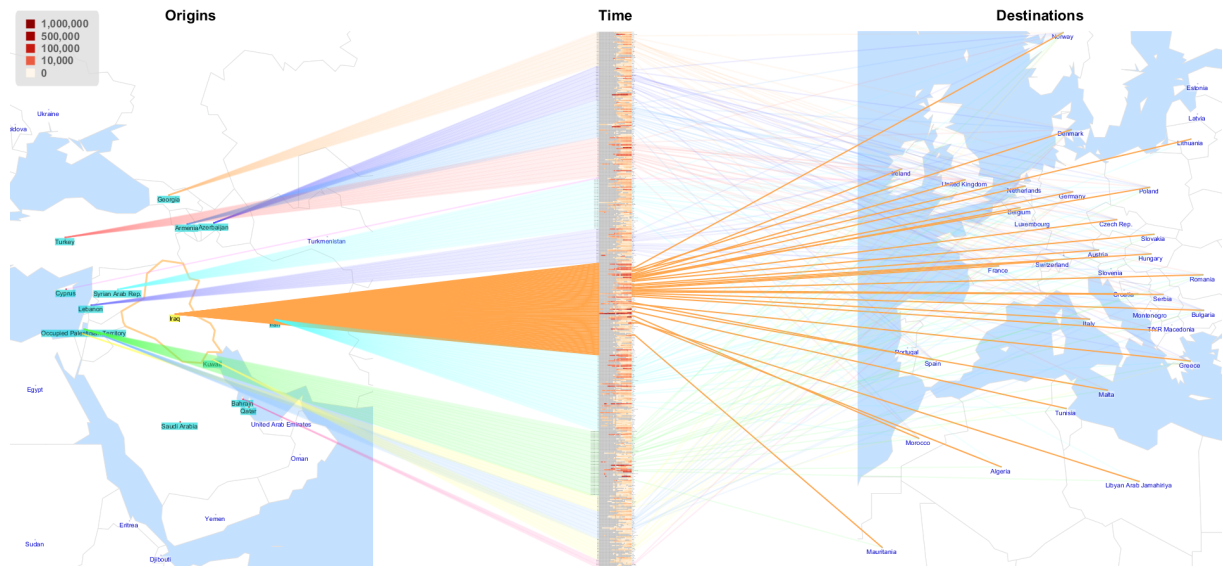
Figure 4: *Flow line coloring: The flow lines of a selection of countries are colored by the flow origins, using a qualitative color map. The heatmap rows are sorted by the vertical positions of the origins, so that flows from the same origins are grouped together. This makes it easy to see the parts of the heatmap which represent the flows originated in the selected locations. Iran is selected in the origins map, therefore the lines from Iran are highlighted and are more opaque than the others.*
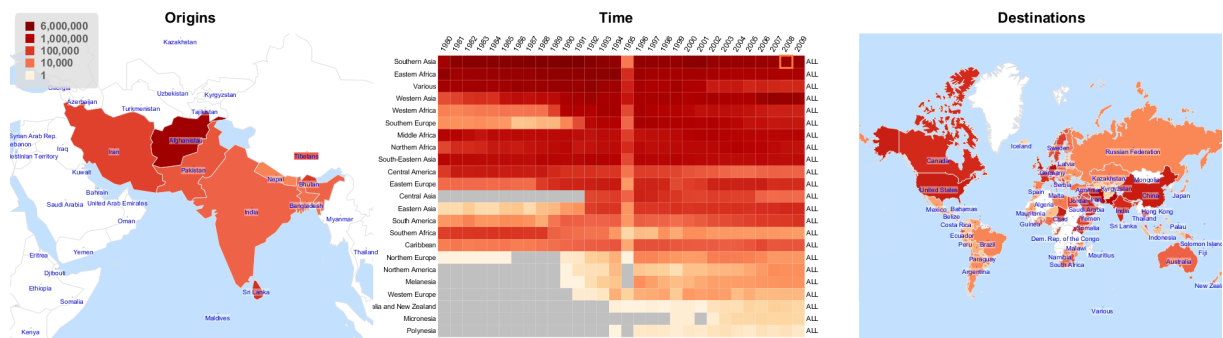

Figure 5: *Flow aggregation: Here all the individual flows between the world's countries were aggregated by the geographic region of the origin country, so that we could see the totals of the magnitudes of the flows originated in each region. We selected Southern Asia in 2008 in the heatmap, thus the maps are colorized showing the outgoing totals for the countries of Southern Asia and the incoming totals for the countries the flows from Southern Asia went to in 2008. Here we also sorted the heatmap rows by the average magnitude in each row. In 1995, there was apparently a problem with the data acquisition, because flows for many countries are missing.*

find out where it was happening in the two geographic maps. This strategy supports tasks T1, T2, T3, T4 and T5.

S3: Comparison of two locations. Select two locations or regions in the origins map and compare the temporal changes of the flows of these locations in the heatmap and their respective destinations in the destinations map. This strategy supports task T9.

Of course, these strategies are stereotyped on purpose. In a real-life scenario analysts would generally combine them in order to solve their exploration tasks and gain new knowledge about the data. We will show how these three strategies can be applied in the case studies in the next section.
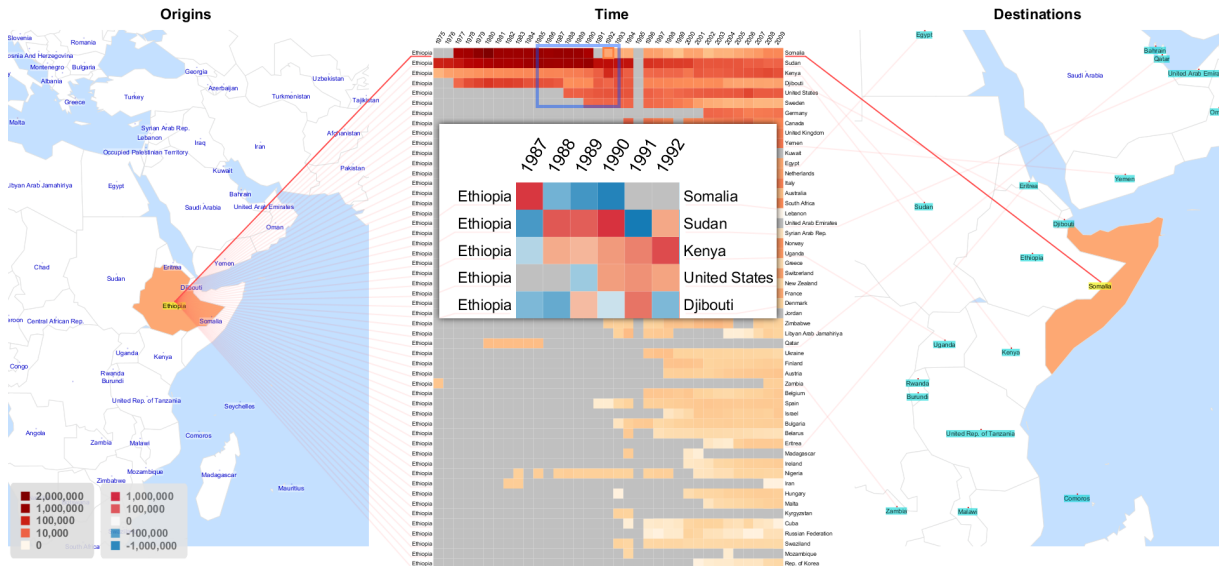
Figure 6: *Refugees from Ethiopia: The enlarged heatmap shows the differences between the pairs of subsequent years (red shows an increase and blue a decrease in the number of refugees). The rows are sorted by the average number of refugees (over the whole range of years). Ethiopia is selected so that we only see flows from this country. Somalia is highlighted, therefore the flow Ethiopia→Somalia is more opaque than the others. The blue rectangle drawn over the heatmap highlights an interesting "staircase" pattern. Here we first used the geographical maps making visual queries to select countries. Then we used the heatmap to find the curious temporal pattern, referred back to the geographical maps to see where the other relevant locations are and found out that most of them were neighboring countries.*

5. Case studies

In this section we illustrate how Flowstrates can be used to analyze real-world temporal origin-destination datasets applying the exploration strategies presented in section 4.

5.1. UNHCR Refugee Dataset

The refugee dataset which has been collected by the UN Refugee Agency contains numbers of people who find asylum in one of the world's countries after leaving their country of origin [UNH10]. The dataset contains data for the last 35 years, and for each of these years there are several thousands of triples: *origin, destination, number of refugees*. These data can tell stories about millions of "forcibly displaced persons": people who had to leave their homes and find refuge in a foreign country. With Flowstrates it is possible to see the dynamics of the changes in the refugee flows.

Suppose, we want to analyze the largest flows from Ethiopia. Let us begin the exploration with strategy S1 (in section 4). We zoom in to Ethiopia in the origins map, select the country (or simply use the name filter), then sort the heatmap rows by the flow magnitudes (Fig. 6). Now we look at the heatmap to find something curious that was happening and then where it was happening, thus switching to

strategy S2. In 1987, most refugees from Ethiopia were in neighboring Somalia, but between 1988 and 1992 the number of Ethiopian refugees in Somalia drastically decreased. We can better see it in the heatmap showing the differences between the years. From 1988 to 1990 it was increasing in Sudan (second row). In 1991, it decreased in Sudan and increased in Kenya, another neighboring country. This is the reason why we can see the "staircase". Apparently, many refugees from Ethiopia were leaving Somalia from 1988 to 1990 and most of them were going to Sudan. In 1991, many were forced to flee again, this time from Sudan (probably, because of the severe drought and food shortage), and went to Kenya.

5.2. Commuters in Slovenia

This dataset contains the numbers of people who commute to work between the towns and villages in Slovenia. There are about 17 thousand flows for the years from 2000 to 2008, so we can explore the temporal development of these flows with Flowstrates.

In Fig. 7 you can see the result of the comparison of the flows to Ljubljana and Maribor, the two largest cities in Slovenia. Here, we use strategy S3 (see section 4). First, we select Ljubljana and Maribor in the map of destinations so
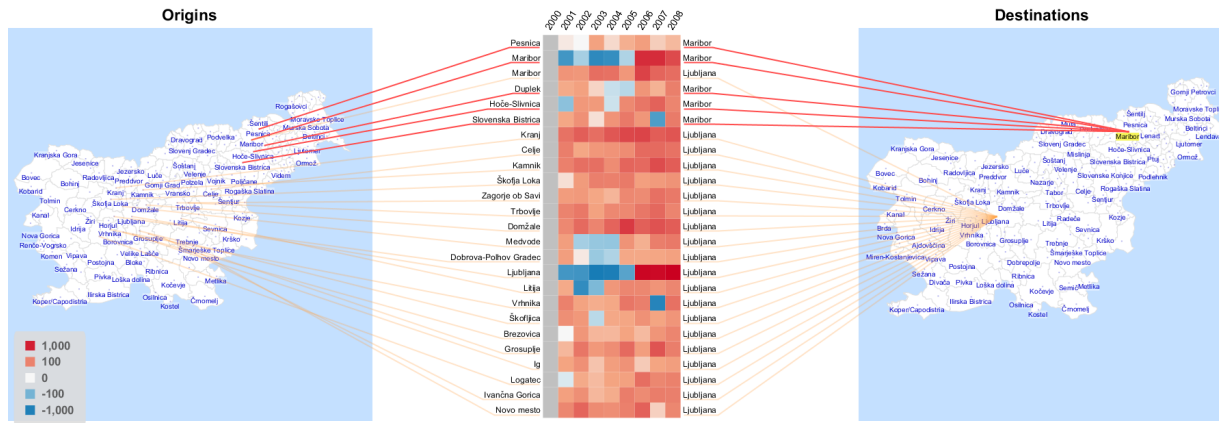
Figure 7: *Commuters in Slovenia: to Ljubljana and Maribor, the two largest cities. Here we see the differences between the number of commuters in each pair of subsequent years (hence, there are no values for 2000). Red corresponds to an increase and blue to a decrease in the number of commuters. Only the most significant flows are shown here (filtered by the average magnitude). The numbers of non-commuters (the half-blue rows of the heatmap labeled as Maribor to Maribor, and Ljubljana to Ljubljana) were decreasing from 2001 to 2005 and increasing significantly from 2006 to 2008, whereas the number of commuters from almost all the other places were steadily increasing all the time. The flows to Ljubljana come from more distant locations than the flows to Maribor, which all come from nearby towns and villages.*

that only the flows to these two cities are shown. Then we use a flow magnitude filter to show only the flows having the largest average magnitudes (over time). In the heatmap we chose to show the differences of the magnitudes between each pair of subsequent years and not the actual numbers of commuters, because we want to better see where they were increasing and decreasing.

What we can see in the heatmap in Fig. 7 is that for most places the numbers of people commuting from them to both Ljubljana and Maribor were steadily growing. However, the numbers of non-commuters (rows of the heatmap labeled as Maribor to Maribor, and Ljubljana to Ljubljana) were decreasing from 2001 to 2005 and increasing significantly from 2006 to 2008.

In Fig. 7 we also highlighted the flows to Maribor in order to see in the map of origins where the largest flows come from geographically (strategy S2). The largest flows to the capital Ljubljana come from more distant locations than the largest flows to Maribor which all come from nearby towns and villages.

6. Related work

The complexity and multidimensionality of spatio-temporal data pose significant challenges, including the scalability to the number of moving objects, locations and resolutions of time and assessment of complex research questions involving space, time and the characteristics of the moving objects

at the same time. Research has been addressing these challenges from many different perspectives.

Geographic flows. One of the most often used representations of entities moving between geographical locations are flow maps [Tob87, Tob05, PXY*05, Guo09] in which the flows are represented as lines or arrows with their widths proportional to the flow magnitudes. Flow maps only show cumulative movements which took place at a particular time period, but can be extended to animated flow maps which show the changes over time [BEW95]. An alternative non-geographic visualization of origin-destination data is OD-matrix [Voo55] in which the flow magnitudes are represented by the cell colors in a heatmap with the rows corresponding to the origins and columns to the destinations. OD-maps is a promising geographic approach proposed by Wood et al in [WDS10] in which the map is divided into a regular grid in which each cell is used as a small replica of the whole map. The aggregated incoming flows from the corresponding grid cells of the large map are shown as colored squares in each of the small maps. This approach allows to represent datasets with enormous numbers of flows, avoiding the clutter caused by the flow line intersections which conventional flow maps suffer from. However, as the other approaches described in this section, OD-maps were not designed for the analysis of origin-destination data with a temporal dimension.

Temporal changes of geographic flows. Many different techniques for the analysis of spatio-temporal data have been developed so far, e.g. space-time cube [Kra03], time-series

graph linked to a map, change maps. Only few of them can, though, be applied directly to origin-destination data. Becker et al [BEW95] show how animation can be used to analyze network traffic data from many time periods. A thorough discussion of the usage of a small multiples display for representing and analyzing stork migration trajectories can be found in [AA06]. Weaver et al propose an alternative representation for the exploration of spatio-temporal patterns with the use of calendar layouts [WFR*06]. Mosaic diagrams were introduced in [AA08] for displaying spatio-temporal patterns in traffic situations. A set of multiple small calendar-like views representing temporal data are displayed in a regular rectangular grid on top of a geographic map.

In general, not much has been written on the exploration of temporal changes in origin-destination data. Marble et al [MGLS97] noted in 1997 that the limitations of the data and the empirical difficulties encountered in their analysis have restricted researchers to the examination of flows within a single time period. This situation has not changed much since then, and there is still a strong need for techniques capturing the spatio-temporal aspects simultaneously. This is the particular problem we addressed in our work.

Visually linked views. The idea of drawing visual links across different views of related data is thoroughly discussed in VisLink [CC07] and utilized in several other systems, e.g. Caleydo [WPL*10].

Before coming up with the idea of Flowstrates, we were very much inspired by Aris and Shneiderman's work on *semantic substrates* [SA, AS07] for supporting the interactive exploration of complex graphs. Substrates are rectangular regions in which the nodes are placed according to a selection of their attributes. Two or more substrates can then be displayed simultaneously on the screen allowing to analyze different properties of the graph. Then, for a selection of nodes in one substrate their connections to the nodes in another can be shown across the substrates, thus visualizing the relationships between the nodes and different attributes.

Flowstrates borrow from semantic substrates the general idea of representing different aspects of the same data in separate views and showing visual links across the views. But in our work we address a more specific problem and propose a technique which is less general, but better suited for temporal origin-destination data. As far as we can say, the idea of representing the flow origins in one geographical map and the destination in another is new.

7. Limitations

One limitation of Flowstrates arises when too many lines are shown between the geographic maps and the heatmap. Too many intersecting lines become tedious to follow. This happens unless only a few locations are selected from either side of the heatmap. A partial solution is to use the heatmap row ordering by vertical positions of the origins (as in Fig. 4).

But this only works on one side, either for the origins or for the destinations. A more sophisticated line crossings minimization algorithm could give better results, but would still require a specific ordering of the rows of the heatmap, thus limiting the possibilities for reordering. A radical solution is to only display the flow lines for a few selected or highlighted nodes.

Not being able to see the flows on *one* map as with conventional flow maps (i.e. a single map with the flow lines drawn between each origin and destination) can also be considered a limitation. With Flowstrates the orientation of the flows is not realistic and the distances between the origins and destinations cannot be estimated from the lengths of the flow lines as in a conventional flow map. To partially compensate this limitation, a conventional flow map could be displayed on demand showing the flows for a specific time period selected by the user. Alternatively, an additional column could be added to the heatmap showing the length of the flows.

8. Conclusion

In this article we presented Flowstrates, a technique for the visualization and exploration of temporal origin-destination data. The design of this technique is based on the task analysis and on a study of design alternatives. Flowstrates does not only represent spatial and temporal aspects of the data, it also highlights the relationships between these two dimensions and facilitates the interactive exploration by querying, filtering, various ordering and grouping techniques. In the article we illustrated how Flowstrates can be used to analyze real-world datasets.

In the nearest future, we plan to collect more user feedback and run a user evaluation. We are currently in contact with several geographers who are very enthusiastic about using Flowstrates for the analysis of their data after trying out the preliminary visualizations which we produced with our current implementation. As soon as our tool is ready for it, we are going to run a think-aloud evaluation (with the tasks described in section 2.1) to gather qualitative feedback from the users, to find out which types of the tasks are better supported by our tool, and to elicit the types of patterns users can discover and the process they follow using Flowstrates.

References

[AA06] ANDRIENKO N., ANDRIENKO G.: *Exploratory analysis of spatial and temporal data : a systematic approach*. Springer, 2006. 2, 9

[AA08] ANDRIENKO G., ANDRIENKO N.: Spatio-temporal aggregation for visual analysis of movements. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on* (2008), IEEE, pp. 51–58. 9

[AAD*10] ANDRIENKO G., ANDRIENKO N., DEMSAR U., DRANSCH D., DYKES J., FABRIKANT S. I., JERN M., KRAAK M., SCHUMANN H., TOMINSKI C.: Space, time and visual analytics. *International Journal of Geographical Information Science 24*, 10 (2010), 1577–1600. 1

[AES] AMAR R., EAGAN J., STASKO J.: Low-Level components of analytic activity in information visualization. In *Proceedings of the 2005 IEEE Symposium on Information Visualization (IN-FOVIS'05)* (Minneapolis, MN, USA), pp. 15–15. 2

[APP11] ARCHAMBAULT D., PURCHASE H., PINAUD B.: Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics 17*, 4 (2011), 539–552. 3

[AS07] ARIS A., SHNEIDERMAN B.: Designing semantic substrates for visual network exploration. *Information Visualization 6* (December 2007), 281–300. 9

[BBL10] BOYANDIN I., BERTINI E., LALANNE D.: Using flow maps to explore migrations over time. *Geospatial Visual Analytics Workshop in conjunction with The 13th AGILE International Conference on Geographic Information Science* (2010). 2, 3

[BEW95] BECKER R., EICK S., WILKS A.: Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics 1*, 1 (1995), 16–28. 3, 8, 9

[BH09] BREWER C., HARROWER M.: Colorbrewer: Color advice for maps. http://colorbrewer2.org, 2009. 5

[CC07] COLLINS C., CARPENDALE S.: VisLink: revealing relationships amongst visualizations. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (Dec. 2007), 1192–1199. 9

[Guo09] GUO D.: Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (October 2009), 1041–1048. 8

[Har99] HARRIS R.: *Information graphics : a comprehensive illustrated reference*. Oxford University Press, New York, 1999. 1, 2

[HKA09] HEER J., KONG N., AGRAWALA M.: Sizing the horizon. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI '09* (Boston, MA, USA, 2009), p. 1303. 5

[HR07] HEER J., ROBERTSON G.: Animated transitions in statistical data graphics. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1240–1247. 3

[KHS02] KEOGH E. J., HOCHHEISER H., SHNEIDERMAN B.: An augmented visual query mechanism for finding patterns in time series data. In *Proceedings of the 5th International Conference on Flexible Query Answering Systems* (London, UK, UK, 2002), FQAS '02, Springer-Verlag, pp. 240–250. 5

[KP99] KREMPEL L., PLÜMPER T.: International division of labor and global economic processes: an analysis of the international trade in automobiles. *Journal of World-Systems Research V*, 3 (1999), 487–498. 2

[Kra03] KRAAK M.: The Space-Time cube revisited from a geovisualization perspective. In *Proceedings of the 21st International Cartographic Conference (ICC)* (2003). 8

[LMK07] LAM H., MUNZNER T., KINCAID R.: Overview use in multiple visual information resolution interfaces. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1278–1285. 5

[LPP*06] LEE B., PLAISANT C., PARR C. S., FEKETE J., HENRY N.: Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization - BELIV '06* (Venice, Italy, 2006), p. 1. 2

[MGLS97] MARBLE D. F., GOU Z., LIU L., SAUNDERS J.: Recent advances in the exploratory analysis of interregional flows. *Innovations in GIS 4* (1997), 75–88. 1, 9

[PXY*05] PHAN D., XIAO L., YEH R., HANRAHAN P., WINOGRAD T.: Flow map layout. In *Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS'05)* (2005), pp. 29–29. 8

[Rae09] RAE A.: From spatial interaction data to spatial interaction information: Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems 33*, 3 (2009), 161–178. 1

[SA] SHNEIDERMAN B., ARIS A.: Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics 12*. 4, 9

[TMB02] TVERSKY B., MORRISON J., BETRANCOURT M.: Animation: can it facilitate? *International Journal of Human-Computer Studies 57*, 4 (2002), 247–262. 3

[Tob87] TOBLER W.: Experiments in migration mapping by computer. *The American Cartographer 14*, 2 (1987), 155–163. 1, 8

[Tob05] TOBLER W.: Display and analysis of migration tables. http://www.geog.ucsb.edu/~tobler/presentations/shows/A_Flow_talk.htm, 2005. 1, 8

[UNH10] UNHCR, THE UN REFUGEE AGENCY, DIVISION OF PROGRAMME SUPPORT AND MANAGEMENT: 2009 global trends: Refugees, asylum-seekers, returnees, internally displaced and stateless persons, 2010. 2, 7

[Voo55] VOORHEES A.: A general theory of traffic movement. *Institute of Traffic Engineers* (1955), 46–56. 8

[WDS10] WOOD J., DYKES J., SLINGSBY A.: Visualisation of origins, destinations and flows with OD maps. *Cartographic Journal, The 47*, 2 (2010), 117–129. 8

[WFR*06] WEAVER C., FYFE D., ROBINSON A., HOLDSWORTH D., PEUQUET D., MACEACHREN A.: Visual analysis of historic hotel visitation patterns. In *2006 IEEE Symposium On Visual Analytics Science And Technology* (2006), IEEE, pp. 35–42. 9

[WPL*10] WALDNER M., PUFF W., LEX A., STREIT M., SCHMALSTIEG D.: Visual links across applications. *Proceedings of Graphics Interface 2010* (June 2010). 9