Visualizing Migration Flows and their Development in Time: Flow Maps and Beyond

Ilya Boyandin, Enrico Bertini, Denis Lalanne

Fig. 1. Refugee Flows between the World's Countries in 1996, 2000, and 2008

Abstract— Numerous migration datasets become available nowadays, but appropriate techniques and tools for their exploration still have to be developed. The most widely used representations of migrations are flow maps: they are familiar to the most people and are therefore "naturally" readable. However, they do not scale well due to the cluttering which increases with the number of represented flows. Many different techniques attempting to solve the cluttering problem have been developed, most of which involve either summarization or information extraction. Despite being effective in finding answers to certain questions concerning the data, all of these techniques have their shortcomings and there is no clear winner among them. A thorough evaluation must be performed to reveal which techniques are superior in helping the analysts with their tasks in the exploration of migration data. Another important issue, which also has to be addressed, is facilitating the exploration of temporal changes in migration flows, as many of the datasets have a temporal dimension. There are many techniques for analyzing spatio-temporal data, but only a handful of them can be applied directly to migration flows: e.g. animation or "small multiples". However, these "simple" techniques prove to be only of limited use, and therefore more appropriate solutions have to be developed. A challenging task is to find an effective representation of migrations which would show the temporal development and retain the spatial properties at the same time. In this paper we present the PhD thesis in which we plan to address these issues and develop, implement and evaluate visual analytics techniques and tools which can facilitate the exploration of migration flows and their development in time.

Index Terms—Migration flows, spatio-temporal visualization, flow maps, thematic maps.

1 Introduction

The work on this PhD thesis began as a collaboration with the UN Centre for Advanced Visual Analytics (CAVA). The goal was to build up a communication channel with the UN, find interesting datasets which require innovative visualizations and also to find potential users. We first contributed to the BirdEye[1] visualization library project which is being developed at the UN. Our contribution concerned the graph visualization part of the library in which we introduced the animated graph layouts into the new framework based on the Grammar of Graphics [18].

While working on BirdEye we decided to focus on the visualization of spatial interaction information, and especially, on migrations flows. There are many important datasets with such information which are available to the public including those collected by the UN. However, tools and techniques which allow to visualize and analyze spatial interactions, and especially, their changes over time, are not widely available and the subject matter remains underdeveloped [11].

One of the most widely used representations of migrations are flow maps [15]. They are visualizations that represent entities flowing between geographical locations (movement of goods, animals and people, airline or network traffic, etc.) with lines connecting the flow sources and the destinations. Flow maps usually do not accurately show the exact migrations paths, instead they are aimed to answer questions such as: Where on the map are the sources and the destinations of the flows? What is happening within a specific location? In which direction do the migrants go? Where are the largest and the smallest flows?

In many situations it is critical to be able to see the statistical trends and changes over time when analyzing migration data. For instance, the refugee data which is collected by the UN Refugee Agency every year is carefully studied, global trends are identified, and based on the results of this analysis various activities are organized to protect and assist the refugees [14].

There is a strong need for techniques and tools supporting the analysts in the exploration of migration flows and especially their development in time. These tools must be able to help to find answers to temporal questions, such as: How do the migration flows change over time? What was happening in a specific time range? When did a specific migration flow reach its peak? The need for such tools was stated by Marble et al. [9] a decade ago and the situation has not changed much since that time. Considering the current support for flow mapping and geodynamics in modern GIS applications Alasdair Rae concludes that the full potential of spatial interaction data remains largely unreleased [11].

This situation presents a challenge and an opportunity for research, therefore we decided to address it in our work.

- *Ilya Boyandin and Denis Lalanne are with the University of Fribourg, E-mail: {ilya.boyandin, denis.lalanne}@unifr.ch.*
- *Enrico Bertini is with the University of Konstanz, E-mail: enrico.bertini@uni-konstanz.de.*

[1] http://code.google.com/p/birdeye

2 PLAN OF RESEARCH

The following is a research roadmap for the thesis described in this paper:

- Create an extensible visualization tool which can be used as a platform for experimenting with different migration flow visualizations and techniques and for their evaluation;
- Implement several state-of-the art techniques for clutter reduction in flow map visualizations as a part of this tool;
- Develop effective techniques and visualizations facilitating the exploration of the development in time which preserve the spatial properties of the sources and destinations of the flows;
- Develop algorithms for automatic detection of important temporal changes in flows and find effective ways to present these changes to the analysts;
- Perform case studies with different spatial interaction datasets;
- Qualitatively and quantitatively evaluate the effectiveness of the implemented techniques for solving specific user tasks involving the end users (analysts).

In the rest of this paper we discuss the related work and the achievements which we have made so far.

3 RELATED WORK

Probably, the first published flow map showing passenger movement by conveyance was made in 1837 by Harness [12]. In the following decades Charles Joseph Minard created numerous complex and detailed flow maps, many of which represented migration data, and popularized the technique.

Since then the general technique of the presentation of such maps has not changed a lot. The source and destination of each flow are depicted on a geographical map and connected by a line, which can be straight or routed to avoid intersections with other flows or the underlying areas of the map. The line widths are often used as a natural way to represent the flow quantities (e.g. the number of migrants).

One of the early computer systems for the generation of flow maps was developed by Tobler in the late 1980s and was called FlowMapper [16]. However, flow visualizations created using this system suffered from visual clutter. There have been many different approaches presented in the literature to solve the problem of cluttering: routing and grouping the edges [4, 7, 10]; using interaction [2, 19] and link-and-brushing with other representations of the data [5, 8]; employing data mining to discover the most interesting flows or to summarize the data [5, 8]; using animation [15]; or building completely new visualizations [21, 20, 22]. We incorporated some of the techniques and ideas described in these publications into our prototype and tried to make the most out of the visual representation so that even datasets with large numbers of flows can be effectively represented. For a detailed discussion refer to section 5.

The visualization and exploration of spatio-temporal data is also a very active research area. However, not much has been written on the exploration of temporal changes in migration flows. Marble et al. [9] note 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.

Many different techniques for analyzing spatio-temporal data have been developed so far: space-time cube, time-series graph linked to a map, change maps etc. Only few of them can, though, be applied directly to migration flows. Becker et al. [2] show how animation can be used to analyse network traffic data from many time periods. A thorough discussion of the usage of the "small multiples" display for representing and analyzing stork migration trajectories can be found in [1]. With our prototype we tried to find out whether displaying flow maps of migrations for different time periods as "small multiples" can be effectively used for analyzing temporal changes in migration flows. This idea as well as the alternatives are further discussed in section 6.

Fig. 2. Visualizing the World's Refugee Flows of the year 2008 in JFlowMap

4 ACHIEVEMENTS

Below is a summary of the work which has been accomplished in the thesis research so far:

- Acquisition and preparation of the datasets with migration flows (UNHCR refugee dataset, US migrations, Swiss and Slovenian commuters, etc);
- Creation of a platform for the visualization and exploration of spatial interaction information;
- Development of a flow map visualization and various interaction techniques for the exploration of migration flows;
- Implementation of several algorithms for reducing the cluttering in flow map visualizations;
- Development of a novel visualization for the exploration of the temporal changes in migration flows.

These achievements are discussed in the following sections in more detail.

5 INTRODUCING JFLOWMAP

JFlowMap[2] is a tool for the visualization of spatial interactions which we have been developing with two main goals in mind: to implement and evaluate different ways of representing flow maps with large numbers of flows, as well as different summarization techniques reducing clutter, and to facilitate the exploration of the temporal changes. Currently, there are two separate views in JFlowMap: a space-centric which is a flow map representing the refugee flows of one specific year; and a time-centric view which is an aggregated overview of the whole dataset represented in a form of a timeline. In the next sections we will describe these views along with the interaction and visual analytics techniques they support.

5.1 Flow Map Visualization

In the basic flow map visualization the flows are represented by straight lines and their directions are indicated by color markers (Fig. 2). The markers are proportional to the flow lengths so that by looking at one side of a flow it is apparent whether the other side is close or far away (this is similar to the line shortening technique described in [2]). The direction markers show the flow directions and at the same time help to reduce the perceived occlusion, because the

[2]http://code.google.com/p/jflowmap/

Fig. 3. Joining the nodes and edges in clusters. Only aggregated edges between clusters are shown. The light gray lines connect the nodes in a cluster with the cluster centroids.

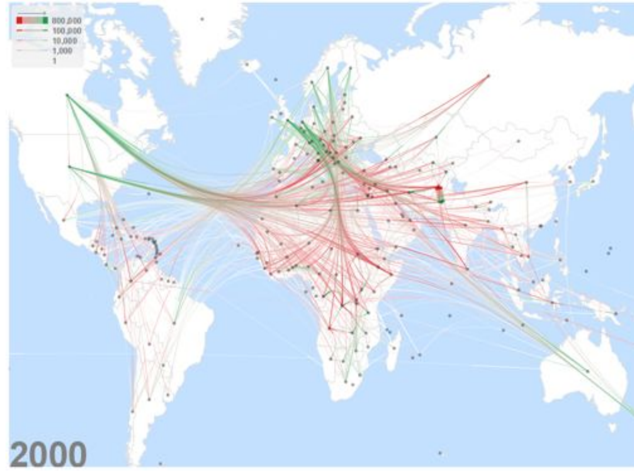

Fig. 4. Flow Bundling. The shape of the flows is changed by forces attracting them to each other.

opacity of the flow lines can be set to a much lower value than the opacity of the markers.The quantities of the flows are mapped to two visual variables: the sizes (the widths) and the color saturations of the flow lines. The user can select the maximum width of a flow line which corresponds to the flow of the largest quantity.

The flow map view supports flow and node highlighting, selection, and dynamic queries [13] for filtering out flows by their quantities or their lengths. With the filter support it is easy to find the largest or the smallest flows, the longest or the shortest ones, or flows with quantities or lengths in a specific interval.

Besides, the prototype has a zoomable user interface [3], which allows the users to smoothly and continuously zoom into any subregion of the map and explore it in detail.

5.2 Node Clustering and Flow Aggregation

Our prototype can cluster nodes using a hierarchical clustering algorithm with various distance metrics. The metrics estimate the similarities between nodes comparing their incoming and outgoing flows and/or the Euclidean distance between them.

After obtaining the clusters, the nodes inside each cluster can be merged so that only aggregated flows between the merged nodes are displayed in the cluster centroids (Fig. 3). The idea behind that is that reducing the number of displayed flows, and thus reducing the line intersections and the occlusion of the flows, can make the visualization more comprehensible [6]. However, it comes at the price of reducing the amount of visible details to the user at one time.

The user can change the maximum size of the clusters and see the updated results of the flow aggregation. This way it is possible to see a summarized overview of the data and dynamically adjust the level of detail.

5.3 Flow Bundling

Another technique attempting to reduce visual clutter in flow (and general graph) visualizations and to make them more readable is bundling. In our prototype we implemented the Force-Directed Edge Bundling (FDEB) algorithm proposed by Holten [7]. This algorithm runs a step-by-step simulation of a process in which the shape of the flows is changed by forces attracting them to each other. The forces depend on the relative positions and orientations of the flows so that only flows which are close to each other and have similar orientations attract each other. As a result of the process, the flows are visually bundled along their joint paths, similar to electrical wires or network cables, which are often strapped together in bundles. The resulting visualizations are less cluttered and can reveal some high-level patterns, like the main "traffic roads" or highly connected regions (Fig. 4).

6 Exploring the Refugee Dataset

One of the datasets we use in our experiments is the UNHCR Refugee Dataset which represents estimated numbers of refugees for every country in which they reside by the countries of their origin. The data has been collected since 1975 and is freely available on the Internet for every year up to 2008.

In the next sections we describe how we applied our tool and the views we have developed to the exploration of the refugee dataset.

6.1 Visualizing the Development in Time

One of the most often used techniques for representing temporal data it the "small multiples" display. It uses multiple charts laid side-by-side and corresponding to consecutive time periods or moments in time [17]. The basic "small multiples" view built with our prototype over three years of the refugee data is presented in Fig. 1. What is immediately apparent when looking at this figure is which countries have more in- or out-flows of refugees: the Western and "developed" countries are completely green, the others are mostly red. This general pattern stays unchanged over time. The largest flows of refugees are caused by wars or military actions: in 1996 these are the flows from Bosnia to Serbia and from Burundi to Tanzania, in 2000 from Afganistan to Pakistan and Iran, in 2007 from Iraq to Syria, in 2008 from Somalia to Kenya and from Congo to Uganda.

Although, the "small multiples" display can be effectively used for the analysis of the development in time in many situations [1], it has the disadvantage that the individual maps must be smaller, thus less detail can be visible to the spectators. The more images are presented to the analyst in a series of "small multiples" at the same time, the smaller the details are, and the more difficult it is to recognize the changes. In the refugee dataset the most significant flows are spatially very local. As a result, the "small multiples" of the flow maps for different years look very similar and it is hard to see the biggest flows most of which are very short.

In order to overcome this problem and to simplify the analysis of the development in time, we decided to create an abstract time-centric view in which the summarized numbers of refugees are represented in a timeline (Fig. 5). In this view the total numbers of the incoming refugees are represented by the left half-circles and of the outgoing refugees by the right ones. The lighter inner half-circles represent the summarized numbers of refugees who found an asylum in a country of the same region. Therefore, the smaller the darker part of a half-circle is, the higher is the "locality" of the flows. The view is hierarchical: the countries are aggregated into regions and the user can expand a region by clicking on it to see the values for the individual countries.

In the time-centric view even some less obvious patterns become apparent which can be more difficult to recognize in the "small mul-

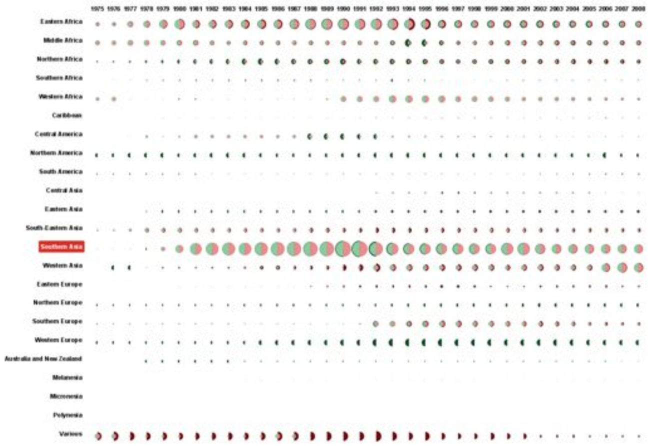

Fig. 5. Timeline of the Numbers of Refugees for the World's Regions. The left half-circles correspond to the numbers of incoming refugees, the right ones to the outgoing. The inner (lighter) half-circles represent the numbers of refugees who found asylum in a country within the same region: i.e. "intraregional" flows; the outer (darker) half-circles show the total number of refugees. Clicking on a region will expand it, showing the individual countries of the region.

tiples" view of the flow maps. For instance, it is apparent from the Fig. 5 that the incoming and outgoing flows of Southern Asia and Eastern Africa are spatially mostly local (the lighter inner half-circles are almost as big as the outer darker ones). Whereas refugees who find asylum in the regions of the developed countries like Northern America and Northern or Western Europe are mostly coming from other regions.

Still, the representation on a map should not be abandoned as it has its important advantages. Most importantly, it shows where the flow sources and destinations are geographically located. We are currently experimenting on linking the flow map view to the time-centric view with the goal to make the transition between the two representations as fluent as possible for the user, so that it feels like one view which combines the best of the two worlds: using maps for representing spatial information and timelines for temporal data.

Another question an analyst may be interested in is: "Which countries or regions are similar in terms of their incoming or outgoing flows?" We are currently experimenting in our prototype with employing clustering to find groups of nodes with flows going to or coming from the same or spatially close locations.

7 Conclusion and Future Work

In this paper we have presented the ongoing work on the PhD thesis with the goal of the developing and evaluating techniques for visualization and analysis of migration flows and, in particular, of their development in time.

In general, when the time dimension is added to spatial data, it becomes much more difficult to visualize them on a map in a natural yet effective way. Therefore, it appears to be easier to develop more abstract and specialized views for the analysis of temporal changes in flows and then link them to a map representation. In this paper we presented one such visualization of the refugee dataset (the timeline view). As a downside, such abstract views do not retain the spatial arrangement of the migration sources and destinations, and therefore, make impossible to see the spatial patterns in the data without consulting the representation on a map.

In the future, we will continue working on the timeline view and, especially, on its interactive linking to the flow map with the goal of bringing together the advantages of both representations. At the same time, we will try to develop an effective visualization representing the temporal changes which will retain the spatial properties (at least some) of the migration data, yet will be effective for the exploration of the development in time. Then, we plan to run a user evaluation to assess the effectiveness of the implemented techniques and visualizations and find out which of them are the most effective for specific tasks. Lastly, we are going to try to develop algorithms for automatic detection of important temporal changes in flows and find effective ways to present these changes to the analysts so that they are not overwhelmed with huge amounts of information.

8 Acknowledgements

The authors wish to thank our colleagues Peter Bak, Jean-Luc Bloechle, Micheal Baechler and Florian Evequoz for their insights and fruitful discussions. This work is supported by the Swiss National Science Foundation, Grant No. 200012_122159.

References

[1] N. Andrienko and G. Andrienko. *Exploratory analysis of spatial and temporal data : a systematic approach*. Springer, 2006.

[2] R. Becker, S. Eick, and A. Wilks. Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics*, 1(1):16–28, 1995.

[3] B. Bederson, J. Grosjean, and J. Meyer. Toolkit design for interactive structured graphics. *IEEE Transactions on Software Engineering*, 30(8):535–546, 2004.

[4] W. Cui, H. Zhou, H. Qu, P. C. Wong, and X. Li. Geometry-Based edge clustering for graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1277–1284, 2008.

[5] D. Guo. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1041–1048, 2009.

[6] I. Herman, G. Melancon, and M. Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):24–43, Mar. 2000.

[7] D. Holten and J. J. van Wijk. Force-Directed edge bundling for graph visualization. *Computer Graphics Forum*, 28(3):983–990, 2009.

[8] L. Liu. PPFLOW - An interactive visualization system for the exploratory analysis of migration flows. *Annals of GIS*, 1(2):118–123, 1995.

[9] D. F. Marble, Z. Gou, L. Liu, and J. Saunders. Recent advances in the exploratory analysis of interregional flows. *Innovations in GIS 4*, pages 75–88, 1997.

[10] D. Phan, L. Xiao, R. Yeh, P. Hanrahan, and T. Winograd. Flow map layout. In *Proceedings of the 2005 IEEE Symposium on Information Visualization (INFOVIS'05)*, pages 29–29, Minneapolis, MN, USA, 2005.

[11] A. Rae. 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):161–178, 2009.

[12] A. Robinson. *Early thematic mapping in the history of cartography*. University of Chicago Press, Chicago, 1982.

[13] B. Shneiderman. Dynamic queries for visual information seeking. *IEEE Software*, 11(6):70–77, Nov. 1994.

[14] The UN Refugee Agency. 2008 global trends: Refugees, asylum-seekers, returnees, internally displaced and stateless persons, June 2009.

[15] W. Thompson and S. Lavin. Automatic generation of animated migration maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 33(2):17–28, 1996.

[16] W. R. Tobler. Experiments in migration mapping by computer. *Cartography and Geographic Information Science*, 14(2):155–163, 1987.

[17] E. Tufte. *Envisioning information*. Graphics Press, Cheshire Conn., 5th printing, august 1995. edition, 1995.

[18] L. Wilkinson. *The grammar of graphics*. Springer, New York, 1999.

[19] N. Wong, S. Carpendale, and S. Greenberg. Edgelens: an interactive method for managing edge congestion in graphs. In *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No.03TH8714)*, pages 51–58, Seattle, WA, USA, 2003.

[20] J. Wood, J. Dykes, and A. Slingsby. Visualisation of origins, destinations and flows with OD maps. *Cartographic Journal, The*, 47(2):117–129, 2010.

[21] J. Wood, J. Dykes, A. Slingsby, and R. Radburn. Flow trees for exploring spatial trajectories. In *Proceedings of the GIS Research UK 17th Annual Conference*, pages 229–234, Durham, UK, 2009. University of Durham.

[22] N. Xiao and Y. Chun. Visualizing migration flows using kriskograms. *Cartography and Geographic Information Science*, 36(2):183–191, 2009.