Using Flow Maps to Explore Migrations Over Time

Ilya Boyandin, Enrico Bertini, Denis Lalanne

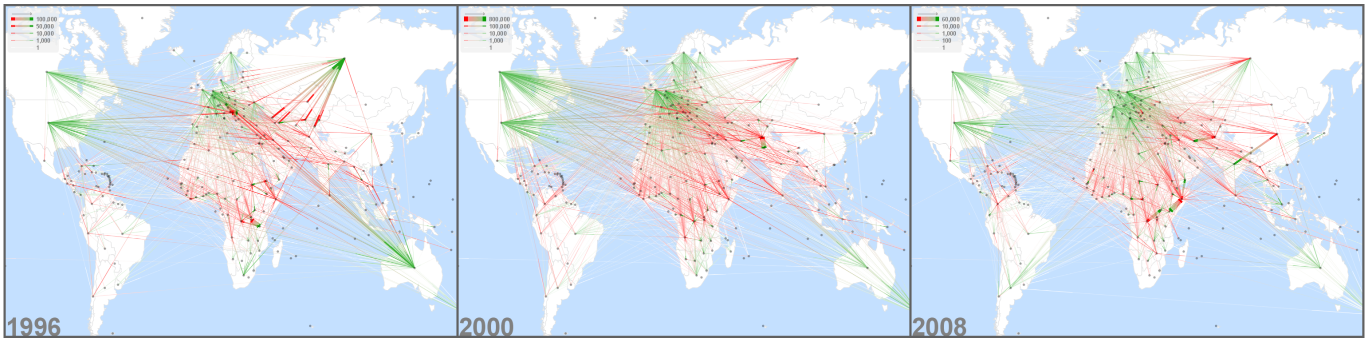

Fig. 1. Refugee Flows between the World's Countries in 1996, 2000, and 2008

Abstract— This paper presents JFlowMap, a graphical tool offering various visualization techniques for producing and analyzing flow maps, one of the most often used visualizations of migration flows. It explains how these techniques can be used in combination to create task-oriented views, i.e. views that help to solve specific user tasks. The article further demonstrates on the UNHCR Refugee Dataset how the tool can be applied to the analysis of temporal changes in migration flows, and finally, sketches the possible directions of future work.

Index Terms—Migration flows, flow maps, spatial flow visualization, thematic maps.

1 Introduction

Flow maps are visualizations that represent entities flowing between geographical locations (movement of goods and people, airline or network traffic, etc.) with lines connecting the flow sources and the destinations. They are one of the most widely used representations of migrations [17]. 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 [16]. Obviously, there is a strong need for 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. [11] a decade ago and the situation has not changed much since that time.

Therefore, we decided to create an extensible flow map visualization tool which can be used as a platform for experimenting with different migration flow visualizations and techniques and for their evaluation. In this paper we present JFlowMap, the tool which we developed, and demonstrate on the UNHCR[1] Refugee Dataset how it can be applied to the analysis of temporal changes in migration flows.

- *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]The UN Refugee Agency

2 Related Work

Probably, the first published flow map showing passenger movement by conveyance was made in 1837 by Harness [14]. 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 [18]. 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 [6, 9, 13]; using interaction [4, 20] and link-and-brushing with other representations of the data [7, 10]; employing data mining to discover the most interesting flows or to summarize the data [7, 10]; using animation [17]; or building completely new visualizations [21, 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 3.

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. [11] 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. [4] show how animation can be used to analyze 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 [3]. One of the claims that we are trying to verify with our prototype is that displaying flow maps of migrations for different time periods as "small multiples" can as well be effectively used for analyzing temporal changes in migration flows. This idea is further discussed in section 4.

3 Introducing JFlowMap

We developed JFlowMap [1], a visualization tool, with two goals in mind: to find ways of effectively representing flow maps with large numbers of flows, and to facilitate the exploration of the temporal changes.

In the next sections we will describe the most important visualization, interaction and data mining techniques supported by our prototype.

3.1 Flow Map Visualization

In the basic visualization the flows are represented by straight lines and their directions are indicated by color markers (Fig. 1). 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 [4]). The direction markers show the flow directions and at the same time help to reduce the perceived occlusion, because the opacity of the flow lines can be set to a much lower value than the opacity of the markers (the user can change the opacities of the flow lines and direction markers). Therefore despite the occlusion the underlying flows can still be seen by the user without the need of using a filter, as the flow lines are transparent. The flow lines are sorted by the flow quantities before drawing and the larger flows are drawn above the smaller ones as they are usually more important and therefore should be more apparent to the user.

The quantities of the flows are mapped to two visual variables: 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.

3.2 Basic Interaction

Our prototype supports flow and node highlighting, selection, and dynamic queries [15] 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 [5], which allows the users to smoothly and continuously zoom into any subregion of the map and explore it in detail. The widths of the flows stay the same even when zooming. This makes it possible to explore small regions, which could be fully or partially covered by scaled-up edges otherwise.

3.3 Node Clustering and Flow Aggregation

Our prototype can cluster nodes using a hierarchical clustering algorithm with various distance metrics. 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. 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 [8]. 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 immediately 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. For larger datasets, however, it does not work smoothly enough, because of the computational complexity of the clustering.

3.4 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 [9]. This algorithm runs a step-by-step simulation of a process in which the shape of the flows is changed by forces attracting the flows 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. Thus a bundled flow map gives a better overview than a straight-line representation. Hence it can be more effective than the basic visualization when an analyst looks at the representations of multiple time periods at the same time in a "small multiples" view.

3.5 Flow Segment Aggregation

One limitation of the bundling algorithms is that flows in the bundles which it produces are not aggregated. They are only routed to form bundles, and flows which end up in the same bundles overlap. It means that the widths of the bundles are arbitrary and thus, when flow thickness is used to represent the quantities, FDEB can produce misleading visualizations as the thickness and the colors of the bundles do not necessarily correspond to the summarized quantities of the resulting flows.

In our prototype we are experimenting with a technique which aims to solve the problem by aggregating segments of the flows which are close to each other. The aggregated segments can then be drawn in such a way that their widths and colors correspond to the summarized quantities of the flow segments of the aggregate and thus can be correctly interpreted by the analyst.

With the techniques described above the users of our prototype can analyze and explore various flow map datasets. Currently, it is only possible to interactively explore data for one particular time period, but the view settings can be propagated to the data for all the other time periods and then a static "small multiples" representation can be generated.

4 Exploring the Refugee Dataset

One of the datasets we use in our experiments is the UNHCR Refugee Dataset [2] 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 available for every year starting from 1975 and up to 2008.

To obtain flow information out of this dataset we had to calculate the differences in the refugee numbers between subsequent years for every source/destination country pair. It is impossible to obtain accurate numbers with this approach, because not all of the refugees who come to a country stay there or keep the same status in the next year, thus the actual flow quantities are higher than the calculated ones. Still the significant changes in the refugee numbers between the years, which are usually caused by preceding dramatic events, are reflected in the calculated dataset.

In the next sections we describe several user tasks which are important for studying temporal migration data and demonstrate how different views of the refugee dataset, generated with techniques available in our prototype, can support an analyst in solving these tasks.

4.1 Spotting Patterns and Temporal Changes

The "small multiples" display is one of the most often used techniques for representing temporal data. It uses multiple charts laid side-by-side and corresponding to consecutive time periods or moments in time [19, 3].

The basic "small multiples" view built with our prototype over three years of the refugee data is presented in Fig. 1. What is immediately

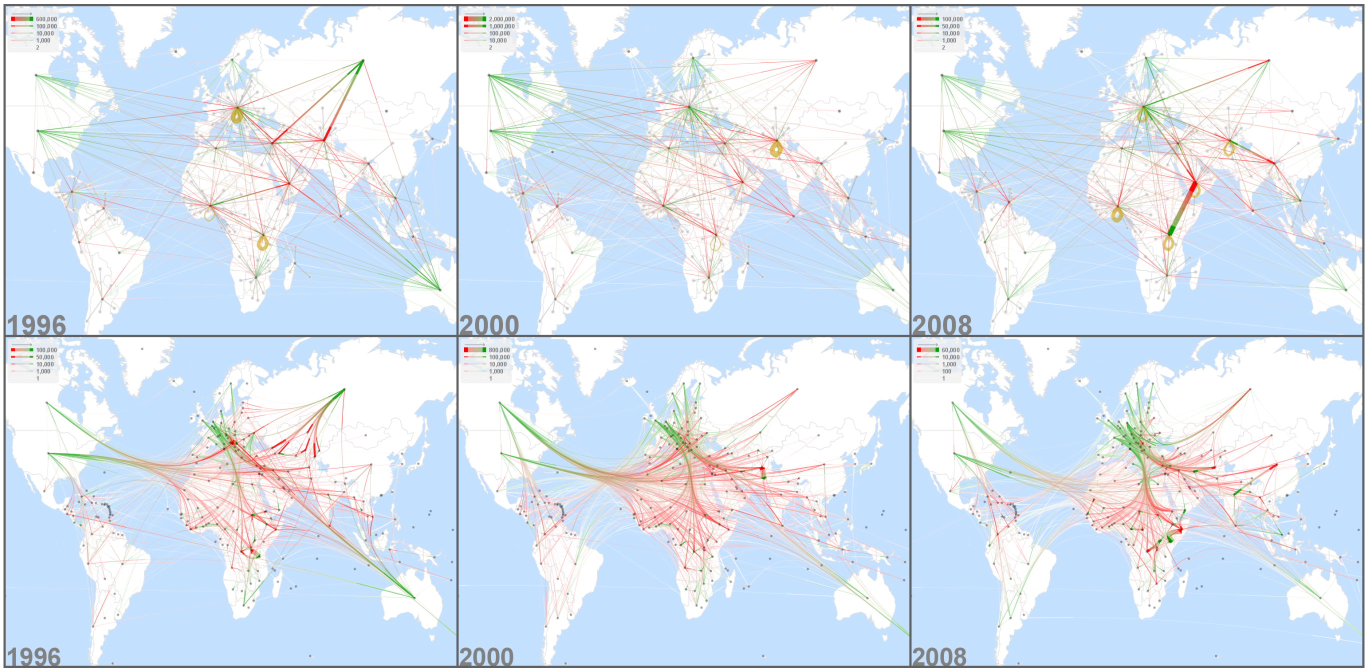

Fig. 2. Refugee Flows between the World's Countries in 1996, 2000, and 2008: Clustered Version (top); and Bundled Version (bottom)

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.

Countries can "change their color" over time: in 1996, five years after the collapse of the Soviet Union, many refugees from the ex-Soviet republics moved to Russia, making it a major refugee-"receiver" for a short period of time, i.e. mostly green. In the subsequent years though Russia has had more outgoing flows than incoming ones, because of the political situation, and hence most of its flows are red.

One decision that must be made when generating "small multiples" is either to use a single global mapping between the quantities and visual parameters (e.g. flow thickness and color) for all the time periods or a separate mapping for each period. One could argue that it depends on what the user task is whether a single mapping or separate mappings should be used: If the user needs to compare the actual numbers between the periods, then a global mapping is needed, so that the flows of different time periods can be compared. If, however, the task is to understand what happens in every particular image, then it's better to use per-image mappings, so that every single image exploits the full visual expressiveness and makes the patterns of the time period it represents more apparent. Therefore for this article we generated small multiples with separate mappings for every year showing a legend in every image.

4.2 Visualizing the "Big Picture"

The "small multiples" display has the disadvantage that the individual maps must be smaller, thus less detail is visible to the spectators. The readability can be improved if the information in the individual maps is presented in a summarized way, which makes high-level patterns more apparent (e.g. in a clustered or bundled form).

A clustered version of the refugee data for the same three years as in the basic view is shown in Fig. 2 (top). The gray lines go from each cluster centroid to the countries which are a part of the cluster. The yellow loops are self-loop flows, appearing because there are flows between some countries belonging to the same cluster. As a result of the clustering the number of nodes, flows and the cluttering are reduced, but the flows between the nodes inside of each cluster are not visible anymore.

A bundled version of the refugee data produced with FDEB (see 3.4) for the same three years is presented in Fig. 2 (bottom). On these images it is easier to see the major "highways" formed by the refugee flows: from Eastern Europe and Africa to the US and Canada, and from Africa to Western Europe. This makes the whole graph more readable and easier to comprehend at a glance. Both the bundled and the clustered version show a "summarization" of the flows, but the bundled version has an advantage: the level of detail is not sacrificed, and therefore it may be preferred. However, a user evaluation must be performed in order to find out which of these representations is quantitatively more effective.

4.3 Focusing on a Region or a Subset

There are situations in which the task of the analyst is to explore what is happening in a specific region of the map or a subset of nodes. In this case it makes sense to only display incoming or outgoing flows in this region, so that the flows of interest are not obscured by the irrelevant ones and the full expressiveness of the visual variables can be used to present them. Fig. 3 illustrates how this technique is applied to analyze a subset of the world's refugee dataset: the flows of the EU countries. It is now easier to compare these flows than in Fig. 1 and some patterns that were not visible become apparent.

4.4 Looking for Specific Features

Finding "the most and the least" is one of the most common geographic analysis tasks people perform every day in their jobs [12]. The ability to filter out nodes by their features can be extremely useful for getting answers to questions like: "How many flows have a certain characteristic"? and "Where are they located"? JFlowMap provides a number of different filters to support the analysts in answering these questions. It is also possible to propagate the filter settings to every single view of the "small multiples" display so that the views representing different time periods are synchronized.

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

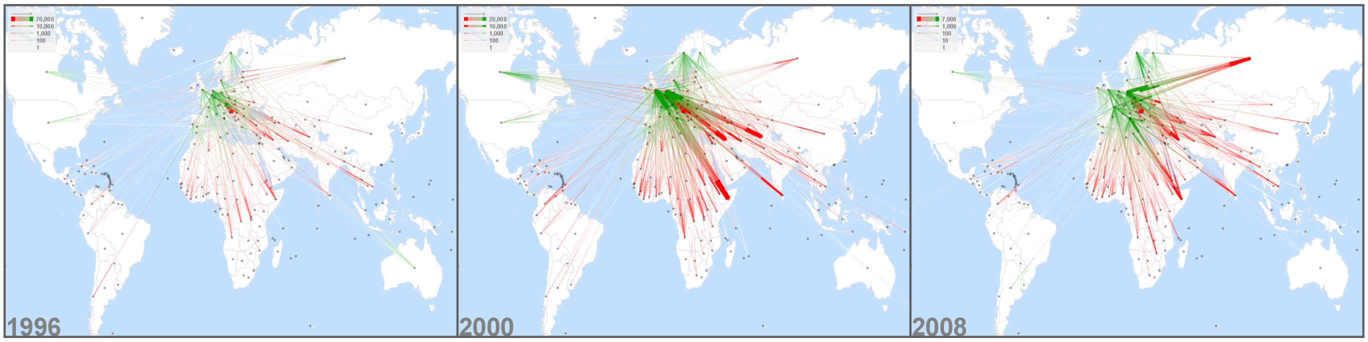

Fig. 3. Focusing: Refugee Flows of the EU countries in 1996, 2000, and 2008

from the same or spatially close locations.

5 Conclusion

In this paper we have presented JFlowMap, a tool offering several techniques for visualization and analysis of migration flows and, in particular, of their development in time. We discussed how these techniques can be combined to produce visualizations of migration data and to support their interactive exploration in order to help the analysts to find answers to specific questions and demonstrated it on the UNHCR Refugee Dataset.

In future work,we plan to run a user evaluation with our prototype to assess the effectiveness of the presented techniques and visualizations and find out which of the techniques are the most effective for specific tasks.

Although we tried to show that the good old "small multiples" display can be quite effectively used for the analysis of temporal changes in migration flows when the number of the considered time periods is relatively small, we believe that it is possible to develop even more effective visualizations tailored to answer specific questions about the changes in flows. In general, when the time dimension is added to spatial data, in particular, to migration flows, it becomes much more difficult to visualize it on a map in a natural yet effective way. Therefore, we believe that it might be advantageous to try to develop more abstract and specialized views for the analysis of temporal changes in flows and not to stick with the map representations only.

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.

Acknowledgments

The authors wish to thank the UN Centre for Advanced Visual Analytics for their collaboration and our colleagues from the DIVA group of the University of Fribourg, especially Jean-Luc Bloechle and Micheal Baechler, for their insights and fruitful discussions and Tanja Boyandin and Florian Evéquoz for the suggestions for improving this article. This work is supported by the Swiss National Science Foundation, Grant No. 200012_122159

References

[1] JFlowMap - Flow map visualization tool: Project Web page on Google Code. http://code.google.com/p/jflowmap/.

[2] UNdata. http://data.un.org/.

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

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

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

[6] W. Cui and H. Zhou. Geometry-based edge clustering for graph visualization. 2008.

[7] D. Guo. Flow mapping and multivariate visualization of large spatial interaction data. 2009.

[8] I. Herman. Graph visualization and navigation in information visualization: A survey. 2000.

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

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

[11] 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.

[12] A. Mitchell and E. S. R. I. (Redlands, Calif.). *The ESRI guide to GIS analysis*. ESRI, Redlands Calif., 1st ed. edition, 1999.

[13] D. Phan, L. Xiao, R. Yeh, P. Hanrahan, and T. Winograd. Flow map layout. 2005.

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

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

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

[17] W. Thompson and S. Lavin. Automatic generation of animated migration maps. 33(2), 1996.

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

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

[20] 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.

[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.