



Spatial Thinking and Geographic Information Sciences

Implementation of a Flow Map Demonstrator for Analyzing Commuting and Migration Flow Statistics Data

Quan Ho^a, Hai-Phong Nguyen^b, Mikael Jern^c, a*

^{a,b,c}*National Center for Visual Analytics (NCVA), Division of Media and Information Technology (MIT),
Department of Science and Technology (ITN), Linköping University, 60174 Norrköping, Sweden*

Abstract

The interest for statistics visualization is emerging with the rise of geovisual analytics on the internet. National and sub-national statistics foundations have started to migrate from tabular representations to interactive web-enabled visualizations that can show trends. Interactive web-enabled visualizations for movement data (or flow data) such as trade flows or commuting flows, however, are still rare. This paper covers spatial interactions and animation for a wide variety of realized movements of people such as commuting and migration between an origin and a destination visually expressed by directed weighted arrows over a geographic space. The paper is built around an interactive flow map demonstrator that can interactively explore and communicate large spatio-temporal and multivariate statistical flow datasets using bidirectional flow arrows where both in-coming and out-going flows can be clearly shown coordinated and linked with a choropleth map, histogram or parallel coordinates plot. The demonstrator is implemented with low-level components from our geovisual analytics toolkit GAV Flash designed with the intention to shorten the time and effort needed to develop customized web-enabled interactive visualization applications such as the introduced flow map application.

Keywords: flow map, spatio-temporal data, multivariate, web-enabled visualization, geovisual analytics;

1. Introduction

Statistics visualization facilitating methods from the geovisual analytics research domain have recently gained interest. Global, national and even sub-national statistics foundations have started to migrate from tabular representations to interactive web-enabled visualizations. Interactive web-enabled visualizations for movement data visualizing trade, commuting or migration flows, however, are still rare. This paper covers spatial interactions for a wide variety of realized movements of people such as commuting and migration between an origin and a destination. This type of flow data can be visually expressed by directed weighted arrows over a geographic space. For a small number of adequately distributed regions directed arrow symbols can be an attractive means of visualization. Cartographic flow maps showing official statistics related to a larger number of sub-national regions (e.g. counties and municipalities) are still problematic and often skewed and detailed which leads to cluttered flows where important details are obscured. In this paper, we introduce an interactive flow map demonstrator that can effectively explore

* Corresponding author. Tel.: +46-11-363-328; fax: +46-13-149-403.
E-mail address: ho.van.quan@liu.se.

spatio-temporal and multivariate statistical flow datasets using bidirectional flow arrows where both incoming and out-going flows can be clearly shown. The choropleth map with overlaid weighted (size) flow arrows is linked to an interactive histogram that gives a detailed and ordered representation of all regional flow data, for example, migration from Oslo (Norway) to Swedish counties and vice versa (see Fig 1). The specific applied research task here is to provide tools to stakeholders to help them gain knowledge exploring possible opportunities for sustainable development in the regions along the long national borders between Sweden and Norway. The geographic area is mostly sparsely populated mountain and forest regions but in the south there are the most densely populated regions in both Sweden and in Norway which represent an important area with a tradition of cooperation and cross-border movement in both directions. The research objective here is to evaluate for the period 2001-2008 social systems and environments with a potential to identify growth in economic, social and environmental development but also to get a better understanding of how migration and commuting across the national border could have a negative influence on local taxes for local regions. Norwegians pursue recreation and leisure activities in the attractive coastal and inland areas of the nearby Swedish border regions. On the other hand, the Swedes living in the border regions mostly commute to Norway to work. Our web-enabled tools are introduced to support visualization and animation aimed at measuring economic, social and environmental developments and to engage policy makers, statisticians and also the citizens.

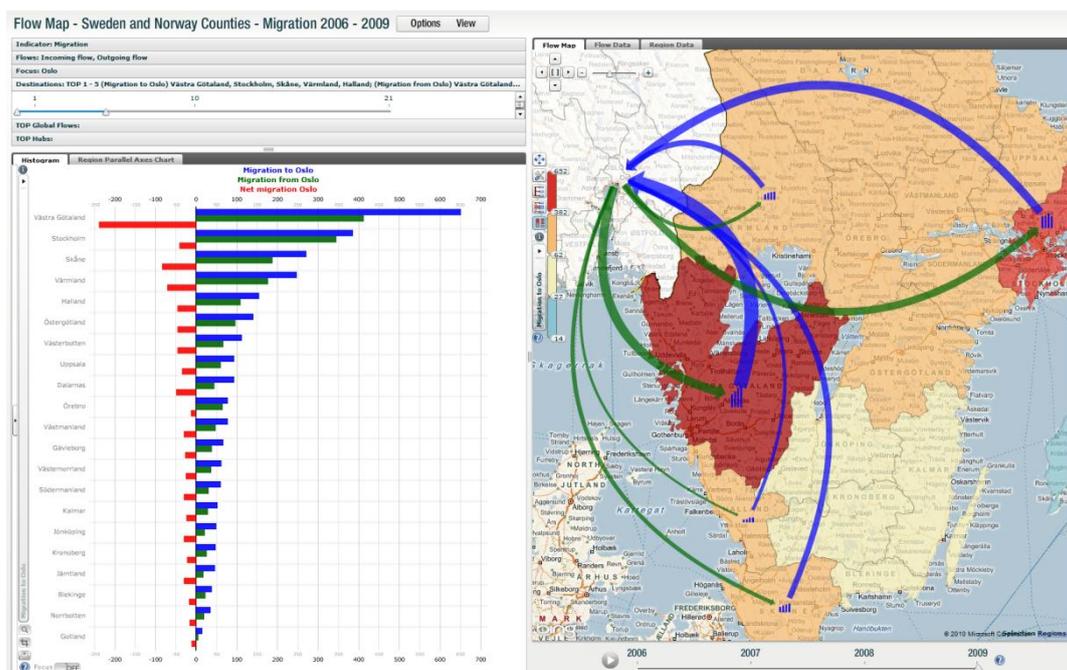


Fig. 1. Migration flows from Oslo to Swedish counties and vice versa during the period 2006-2009. Histogram shows (1) the volume of migration flows to Oslo (blue bars) to Sweden, (2) the volume of migration flows from Oslo to Sweden (green bars), and (3) the difference (or net value) between out-going flows and in-coming flows (red bars) which shows that there is a positive trend of migrating to Oslo from Swedish counties. Blue arrows (right part) show top 5 migration flows from Swedish counties to Oslo. Green arrows show inverse flows from Oslo to Swedish counties. Bar glyphs (blue ones) show time-series values of top 5 migration flows from Swedish counties to Oslo. Polygon layer (or region layer) is colored according to migration volumes from Swedish counties to Oslo.

An early result from evaluations made by analytics experts confirm a lot of interest and awareness among the local politicians and could have an effect on future regional development. The objective of our research has been to develop and evaluate a flow map demonstrator that supports:

- both exploration and presentation of flow data;
- that tools are web-enabled through Flash, a requirement from our statistics research partners;
- interaction with spatio-temporal and multivariate flow interaction table with interactive performance;
- animation of flow time series;
- different background map layers, e.g. Google map or Bing map, for identifying the name of geographic locations;
- dynamic queries and filter operations;
- time series glyphs to show changes over time;
- a dynamic histogram to show order and magnitudes of flow values;
- a parallel coordinates plot for filtering data;
- a high level of user interaction controls answering various questions about flow data such as:
 - Which are the dominant flows (or the trend of movement) in a certain year?
 - Which are the top municipalities in Norway to which people living in Swedish border municipalities tend to commute or migrate?
 - What is the net migration i.e. the difference between out-going and in-coming flows?
 - How do flows vary over time?

2. Related work

Visual exploration of spatial-temporal data has been the subject of many research papers. Andrienko et al. [1] provides several motivating approaches to this area. As a response, a generic “GeoAnalytics” visualization (GAV Flash) tool was introduced by Ho et al. [2] and similar methods are employed in this work. In this paper, we expand the toolkit with flow mapping.

Flow mapping is not new. There have been many attempts to visualize spatial interactions (flows) for various purposes (Tobler [3, 4]; Phan et al. [5]; Rae [6]; Guo [7]). A common approach to flow mapping is to use straight lines or curves (for simplicity we call them lines) to represent flows and the line thicknesses and/or line colors for the volume of the flows.

When applied to large geographical datasets, flow maps tend to be cluttered. To address this problem, there are a number of different approaches. One approach is to group lines into bundles (Phan et al. [5]; Holten [8]; Cui et al. [9]; Holten and Wijk [10]). This approach can effectively reduce the cluttering in a flow map; however, it does not seem to easily support tasks such as comparison of flows back and forth between two locations or flow filtering, which are our objectives. Every time users want to focus on a subset of flows, e.g. top ten flows, or only flows originating from an origin, and filter out other flows, then the bundles should be redrawn or flow lines should be clustered again. In the latter case, the curves representing the flows can be changed. This can make it more difficult for users to observe the flows.

The second approach is to use a grid to compute the line density of each cell based on the total number of migrants moving along flow lines through each cell (Rae [6]). This approach is suitable for giving an overall pattern of interactions between places; however, this approach results in an image, therefore, it does not seem to support operations such as filtering, selection, brushing, zooming, and panning, which are our objectives.

The third approach is to use interactive queries to select and show only a small subset of data at a time (Rae [6]). This approach is more suitable to our objectives; therefore we adopt it in our work.

Although there has been a lot of research in this area, most of them do not seem to meet our objectives which does not focus on giving an overview of spatial interactions but focus more on the aspects of usability, user interactions, animation, analysis and knowledge communication.

First, except for Guo work [7] most of them do not support user interaction operations such as filtering, selection, brushing, zooming, panning, and linking which are very important to facilitate the process of exploration and analysis of users.

Second, none of them use an extra dataset containing additional location attributes (e.g. income level, employment rate, purchase power) in combination with the flow dataset to support the analysis of the flow dataset, for example, to answer one of the most basic questions: why do people move from a location to another location?

Third, they do not seem to support visualization of flow data time series which is an important aspect in most migration and commuting scenarios.

Fourth, they do not support web-enabled visualization which allows public users to access their work easily through internet; therefore, public users (statisticians, citizens, etc.) cannot benefit from their work.

Fifth, with their work users cannot communicate or share their discoveries easily which is one of important aspects of geovisual analytics systems.

3. Approach

To develop the flow map demonstrator, we have followed a human-centered approach and focused on the needs of our partners and target end users (statisticians, public users) who want to:

- find and explore the dominant movements (flows) in a certain time period;
- find hubs which have a large number of people commuting/migrating to or from them;
- select a region as the origin and explore flows originating from this region to destinations and inverse flows from the destination to the origin; find and explore the dominant flows;
- gain insight why people tend to move from a location to another location;
- explore the change of flows over time;
- communicate discoveries with other users;

From analyzing the needs of our target users we found that a tool showing an overview of static flows is less important than a tool which allows users to interact with, explore and analyze flow data and then share their discoveries and understandings with others. Therefore, in this paper, we focus more on aspects such as user interactions, visual exploration, and knowledge communication showing only a subset of flows which are of particular interest. This task is achieved by implementing a dynamic filtering mechanism that dynamically displays only flows meeting defined query conditions and removes other flows. By plotting a subset of flows of interest which is normally quite small we avoid the problem of cluttering which is normally a common problem of flow map applications.

3.1. Flow mapping

To visualize flow data, depending on the various usage scenarios we apply different approaches. The main approach is to use directed weighted arrows, where each arrow represents a movement from an origin to a destination and the thickness of the arrow represents the number of people. Arrow thickness can be scaled dynamically to make arrows more readable. Nevertheless, they always reflect the values they represent. To avoid overlapping, arrows are displayed as quadratic Bezier curves instead of straight lines (see Fig 1, right part). The curvatures of arrows can also be adjusted dynamically and edited individually to reduce clutter and be more readable.

The second approach is to use colored-shaded regions. Nevertheless, this approach can only be applied to the case when users want to select an origin and explore flows originating from that origin. In this case, each flow corresponds to a destination and therefore can be visualized by coloring the region representing that destination (see Fig 1, right part).

3.2. Supporting visualizations

In order to provide users with additional means to explore and gain insight to details in the flow data, we dynamically link to a histogram based on the well-known focus & context method often used in

information visualization and available as a component in GAV Flash [11]. The choropleth map is linked to the histogram and will show detailed information. This method avoids cluttering with too many arrows because the flow data is simultaneously displayed for all regions in the histogram. In a typical scenario in which users select an origin and want to explore flows of this origin, the histogram will show data of pairs of flows originating from and coming in this origin as well as the difference between two flows (back and forth) in each pair. Flows can be ordered according to their values or the name of origins or destinations. Dynamic filtering sliders are also implemented to allow users to be able to focus on a range of flows and/or a range of values. From the histogram in Fig 1 we see that there are more people migrating from Swedish counties to Oslo than migrating from Oslo to Swedish counties.

In addition to the histogram, data tables that look like excel sheets are also implemented to allow users to search for data values when necessary. All visualizations (views) are simultaneously coordinated and linked to each other to enhance the effectiveness of exploration and analysis of data as presented in section 4.5.

3.3. Visualization of flow time series

Visualisation of flow time series is another important objective of our research. To address this, we use two different approaches. The first approach is to use time chart glyphs such as bar or line charts (see Fig 1, right part). This approach can be applied to the scenario in which users want to explore flows originating from the same origin. Given a selected origin, each time chart glyph represents time series data of a flow originating from the origin and to a destination and also its inverse flow if necessary.

In addition to time chart glyphs, we also implement trend glyphs which use glyph size to represent the average amount of people moving or commuting and glyph slope to show the trend of the movement - increasing or decreasing.

The second approach for visualizing flow time series is through simultaneous animation of arrows, colored regions and histogram. The animation can be started and stopped at any time point and the animation speed can be also controlled to allow users to be able to observe changes easily.

3.4. Using regional data to support analysis of flow data

To give the analyst a better understanding and reason why people need to move, a second regional dataset is introduced. Indicators such as income level, employment rate and higher education rate could provide additional knowledge to explain, for example, an unusual high migration from/to a region.

To visualize the regional dataset we use two approaches. The first approach is to use colored regions. Regions are colored according to the values of an indicator being selected. The second approach is to use a parallel coordinates plot which represents regional data items as lines as shown in Fig 8.

By using a regional dataset we support users searching an answer to the question why people tend to move from one region to another region through finding the correlation between flow data and regional data. For example, users may find out that regions with high level of income attract more people from other regions commuting to work, or regions having many universities attract more people moving to for their education.

4. Flow map implementation

To implement the flow map demonstrator, we base on the geovisual analytics toolkit GAV Flash [2, 14] (see Fig 2) which is designed with the intention to shorten the time and effort needed to develop customized web-enabled interactive visualization applications.

First, we use the choropleth map [11] as a background to develop our flow map component. To visualize flow data as mentioned in sections 3.1 and 3.3, we add new layers such as flow arrow layer,

bar/line chart glyph layer on top of it as shown in Fig 3. By extending the choropleth map component, we inherit various features of it. For example, it allows users to use online maps such as Google map and Bing map as alternative backgrounds to identify the name of geographic locations easily. It also provides many interaction operations such as hovering, zooming, panning which support users exploring and analyzing data more effectively.

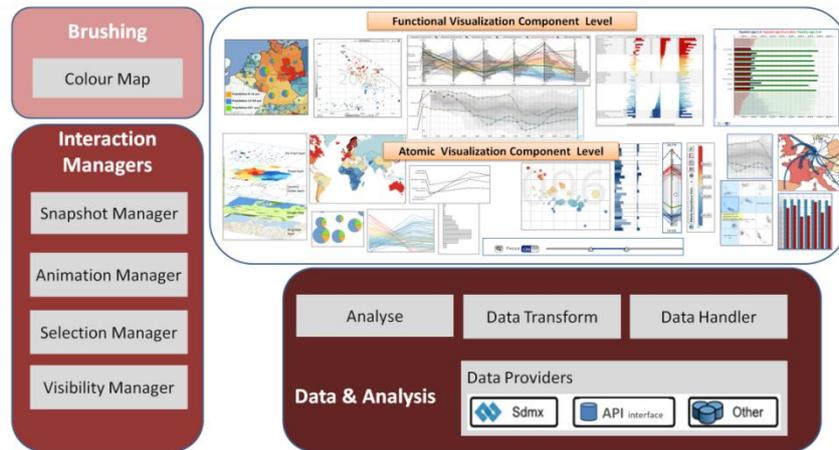


Fig. 2. GAV Flash architecture and framework

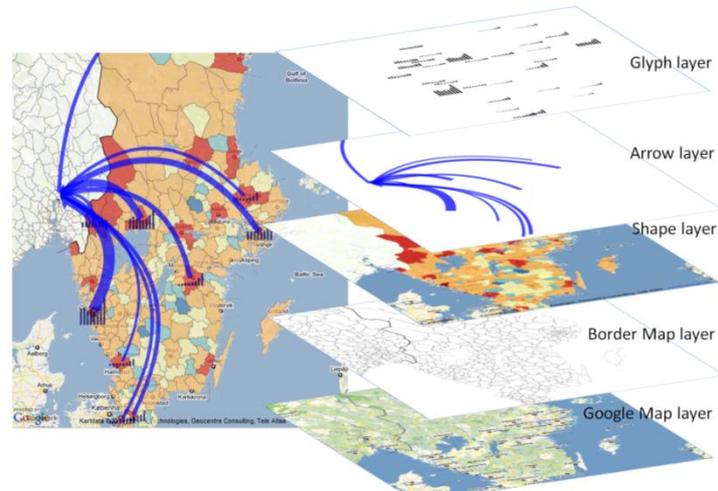


Fig. 3. Layers of the flow map component

		NO-01	NO-02	NO-03	NO-04	NO-05	NO-06	NO-07	NO-08	NO-09
		Østfold	Akershus	Oslo	Hedmark	Oppland	Buskerud	Vestfold	Telemark	Aust-Agder
SE-01	Stockholm	39	106	386	15	15	41	31	22	5
SE-03	Uppsala	15	24	95	3	3	10	4	3	6
SE-04	Södermanland	4	11	62	3	4	4	3	8	0
SE-05	Östergötland	11	28	142	0	4	12	12	5	0
SE-06	Jönköping	19	18	51	3	7	3	3	0	0
SE-07	Kronoberg	3	15	49	3	0	4	9	3	3
SE-08	Kalmar	3	10	54	0	0	4	9	3	0

Fig. 4. An example of an interaction table which represents flow data input

Second, we customize other components of GAV Flash such as the interactive histogram, data table to adapt them to our needs of visualization and exploration.

Third, basing on GAV Flash data model, we build a data provider which has an interaction table (see Fig 4) as input. It is optimized for flow data (in term of performance) and can handle large flow datasets such as trade flows among countries or commuting flows among municipalities of a country.

Finally, by using GAV Flash, we inherit various features of it. For example, we inherit linking view mechanisms supported by GAV Flash to link views in the flow map application to each other. Linking is done through various mechanisms such as selection, filtering, color mapping, and animation. Therefore, when there is some change in a view (e.g. a region is selected in the flow map view) other views will be updated to reflect this change. We also inherit the storytelling mechanism of GAV Flash to support communicating discoveries among users or publishing discoveries in blogs or web sites. An example of a published story can be found at <http://vitagate.itn.liu.se/GAV/flowmap/STGIS2011/story/> explaining about important discovered patterns for migration between Norway and Sweden in 2008.

5. Flow data analysis case studies

In this section we apply the approaches presented in section 3 to analyze two different flow datasets and find trends of movement as well as gain insight into these flow datasets. The first flow dataset contains data of commuting across the border between Norway and Sweden from year 2001 to year 2008 at municipality level. Using a dynamic filter to find dominant flows in 2008 we found out that (1) top five flows are from Göteborg, Karlstad, Stockholm, Linköping, Lund of Sweden to Oslo of Norway; (2) in top ten flows there is only one flow from Norway to Sweden: Halden (Norway) to Strömstad (Sweden) (see Fig 5); and these two municipalities are neighbors of each other. By doing this for every time periods we found out that flows from to Göteborg, Karlstad, and Stockholm of Sweden to Oslo of Norway and from Halden of Norway to Strömstad of Sweden are dominant flows.

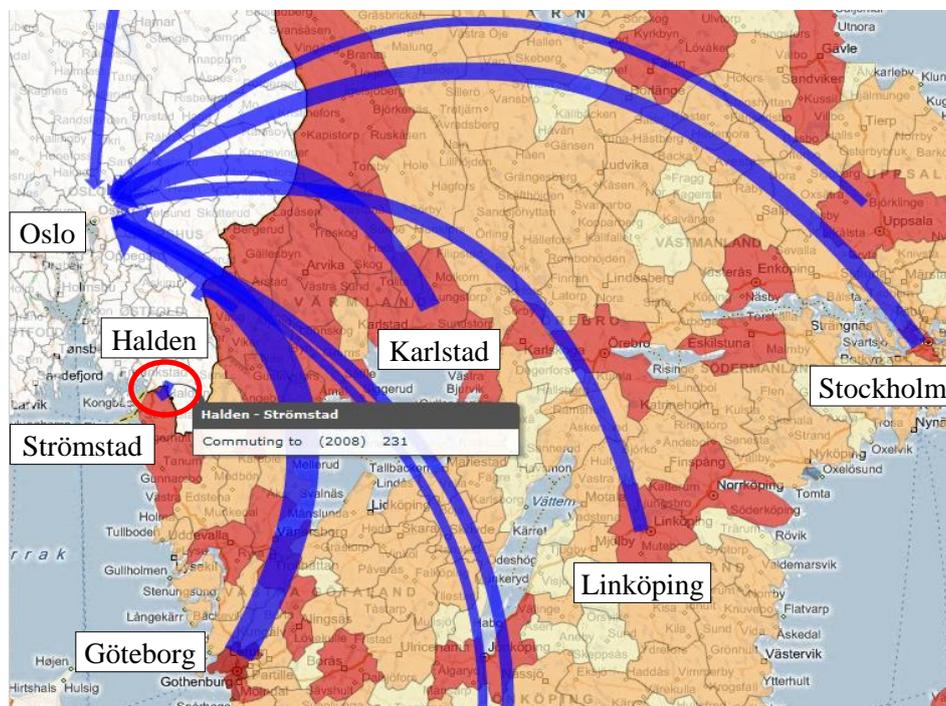


Fig. 5. Top ten global commuting flows in 2008

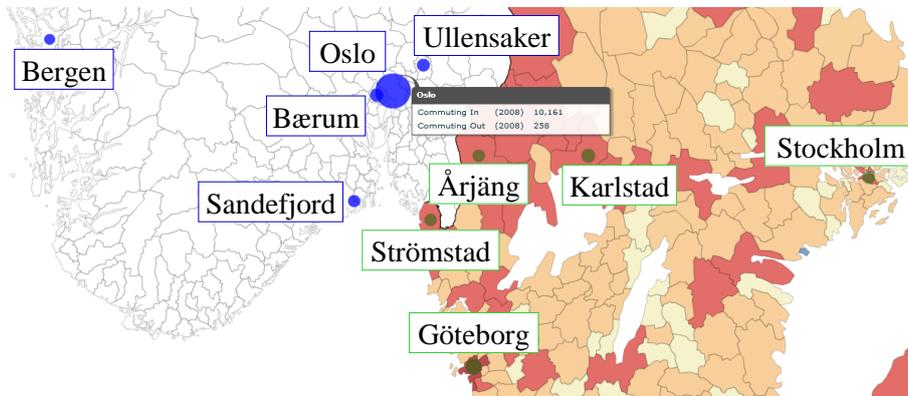


Fig. 6. Top five hubs-in (blue circles) and hubs-out (blue circle) for commuting in 2008; the circle sizes represent the volumes of commuters.

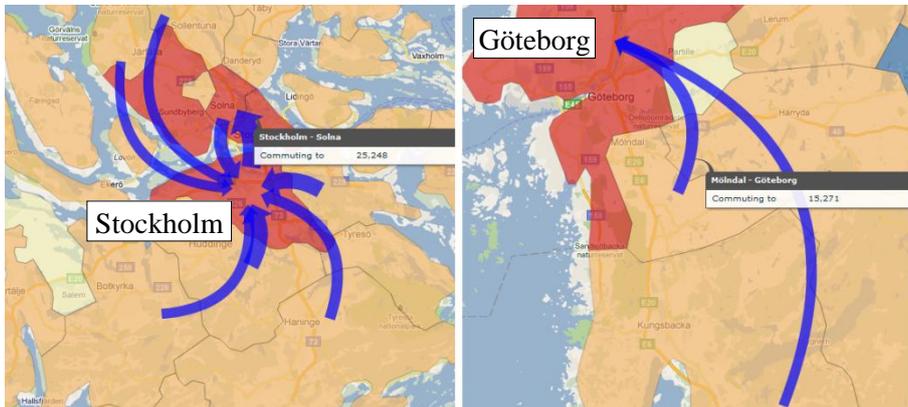


Fig. 7. Top ten commuting flows in 2008; regions are colored according to higher education rate

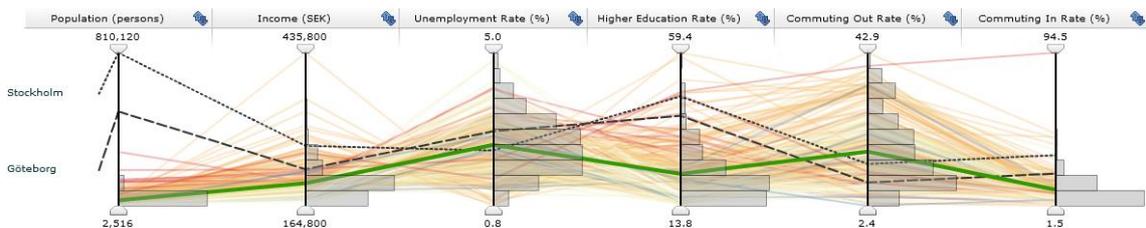


Fig. 8. Stockholm and Göteborg have high income and higher education rate and attract a large number of commuters

In a similar manner, by filtering on the total number of people commuting to and from each municipality we found that (1) Göteborg, Filipstad, Karlstad, Strömstad, Stockholm and Torsby are the Swedish municipalities which have most people commuting to Norway; and (2) Oslo is the Norwegian hub which attracts the most people commuting from Sweden (see Fig 6).

The second flow dataset contains data of commuting among 290 municipalities in Sweden in 2008. In addition, a regional dataset containing indicators such as population, average income, high education level, unemployment rate is also used to support analysis of the flow data. From Fig 7 and Fig 8 we see

that top ten flows involve two municipalities, Stockholm and Göteborg which have high higher education rate as shown in the colored region layer. The parallel coordinates plot in Fig 8 also shows that Stockholm and Göteborg have high income levels which may explain for this fact.

6. User feedback

The major part of national statistics bureaus is fulfilling the role as the nation's infrastructure for official statistics: providing high quality data, objectively and impartially, for everyone to use as a basis for decision making and information. Many of the statistics produced have geography as an essential dimension, and it is important to allow users to understand the impact of geographical differences and structures. Until recently the dissemination of such data has been rather dull and less attractive to many categories of users. For this reason they are convinced that their statistics are grossly under-utilised, as compared to their great potential.

The flow map demonstrator was used by our case study partners (Statistics Sweden, Region Västra Götaland, and the Norwegian county Ostfold Fylke) for communicating the essentials of official migration and commuting statistics to a broad range of users via their web site. The tool has been used by statistics experts to analyze in depth the geographic structures and correlations resulting in good stories which have been presented on the web site. This has proven to be an excellent way of catching the attention of many users, including the media.

Statisticians from both Swedish and Norwegian authorities summarized in their evaluation reports that gained insight and knowledge based on this interactive flow map case study identifies more efficiently commuting statistics across the border and the potential political and economic consequences compared to experience from previous used more static flow map applications. The gained knowledge from the interactive flow animations during 2001-2008 and attached time glyphs clearly indicates an increasing trend that significant more Swedes prefer to communicate and work in Norway then the reverse scenario. An important conclusion that was of particular importance related to negative tax affect for Swedish municipalities along the border. The combination of multiple linked "heat maps" with overlaid weighted arrows and a dynamic histogram results in a comprehensive insight into number of people commuting from a large number of Swedish counties (see Fig 1) was another positive comment. Consensus was that the introduced methods and demonstrator could help advance usage of interactive flow map visualization for a better understanding of both commuting and migration between sub-national regions and across national borders.

7. Conclusion and future work

This paper presents a web-enabled demonstrator for the visual exploratory analysis, collaboration and dissemination of flow time series data. The modularity provided by the GAV Flash component architecture makes it easier for a developer to implement and add new visualization components/layers such as the "weighted arrow" layer and the glyph layers that are integrated with the already existing choropleth map component. Extending the existing map component with a special feature, such as, a time glyph was a simple undertaking for the developer. Reviews and evaluations performed by the statistical partners highlight the following features:

- The flow map demonstrator can be customized by a statistics organization - requires only regional boundaries (shape file) and associated regional and flow data;
- The storytelling approach is regarded as a painless and more attractive alternative for the general public;
- Interactive visualization and storytelling encourages collaboration between statistics analysts and users of statistics;

- Possibility to capture, save and open discoveries (snapshots) with attached analytics reasoning metadata, e.g. storytelling;
- It is easy to import external statistical data;
- Ability to have dynamic time-link views and see the multi-dimensionality of regional development;
- Increased expectations in terms of user experience;
- It encourages more educational use of official statistics.

Finally an appreciated experience learned from this project includes the formation of a multi-disciplinary team and the involvement of statisticians from the beginning of the project. Latest flow data was applied to guarantee involvement of end users such as politicians from both countries Sweden and Norway. A number of different versions of the demonstrator for different flow datasets can be found at <http://vitagate.itn.liu.se/GAV/flowmap/STGIS2011/>.

Our next steps are to (1) apply it to massive world trading data from OECD comprising trading flow data (export and import) of thousands of commodities such as “food” “live animals”, “beverages”, “tobacco”, “crude materials”, “machinery and transport equipment”; (2) improve the graphical user interface of the application to make it easier to use; (3) provide users with more interactive features such as supporting picking of multiple origins and/or destinations, floating information panels; and (4) perform a thorough evaluation.

The latest versions of the flow map application can be found at <http://vitagate.itn.liu.se/GAV/flowmap/LatestVersions/>.

References

- [1] Andrienko G, Andrienko N, Demsar U, Dransch D, Dykes J, Fabrikant IS, Jern M, Kraak MJ, Schumann H, Tominski C. Space, time and visual analytics. *International Journal of Geographical Information Science* 2010; 24(10): 1577-1600.
- [2] Ho Q, Lundblad P, Åström T, Jern M. A Web-Enabled Visualization Toolkit for Geovisual Analytics Visualization and Data Analysis. *Proceedings of SPIE* 2011; 7868: 78680R-78680R-12.
- [3] Tobler W. Experiments in Migration Mapping by Computer, *American Cartographer* 1987; 14(2): 155-163.
- [4] Tobler W. Flow Mapper Tutorial. <http://www.csiss.org/clearinghouse/FlowMapper/FlowTutorial.pdf/> [Accessed 3 November 2010].
- [5] Phan D, Xiao L, Yeh R, Hanrahan P. Flow Map Layout. *Proceedings of IEEE Symposium on Information Visualization* 2005; 219-224.
- [6] 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* 2009; 33(3): 161-178.
- [7] Guo D. Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data. *IEEE Transactions on Visualization and Computer Graphics* 2009; 15(6): 1041-1048.
- [8] Holten D. Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. *IEEE Transactions on Visualization and Computer Graphics* 2006; 12(5): 741-748.
- [9] Cui W, Zhou H, Qu H, Wong PC, Li X. Geometry-Based Edge Clustering for Graph Visualization. *IEEE Transactions on Visualization and Computer Graphics* 2008; 14(6): 1277-1284.
- [10] Holten D, Wijk JJv. Force-Directed Edge Bundling for Graph Visualization. *Computer Graphics Forum (Eurographics/IEEE-VGTC Symposium on Visualization)* 2009; 28(3): 983-990.
- [11] GAV Flash User Guide. <http://www.ncomva.se/guide/?chapter=Visualizations> [Accessed 30 May 2011].