A Low-Threshold Geovisualization of Statistical Data on Commuter Relationships

Sebastian Specht, Christian Hanewinkel
Leibniz-Institute for Regional Geography, Leipzig, Germany
{s_specht | c_hanewinkel}@ifl-leipzig.de

***Abstract*—This work presents the ongoing efforts made in creating a web application for the visualization of statistical data on commuter relationships. The results of the cooperation efforts between the Leibniz-Institute for Regional Geography (IFL) and the Statistics Department of the German Federal Employment Agency (BA) are targeted toward a broad audience of users such as politicians, journalists, planners, and the general public. The presentation will elaborate on the challenges such a geovisualization offers such as the right choice of symbolization, the proper reduction of data complexity, the integration of detail and overview through methods of both, interaction and visualization. Included are demonstrations and an invitation to discuss proposed solutions.**

I. Introduction

Statistical data on socio-economic flows are crucial for the assessment and planning of regional and urban infrastructures. Statistical indicators on magnitude and number of relations of migrants and commuters give insight about the level of centralization of a region, the inter-connectedness of its entities and they allow an estimation of development potentials.

The Federal Employment Agency (BA) generates statistical data on commuters when processing social insurance / social security registrations of employed people. The data have been made available as files containing matrices on the administrative levels of states, counties, and municipalities (Länder, Kreise, Gemeinden). These files are usually employed by local government, local politicians, planners and journalists under a local or regional perspective. It can be assumed that only few of those users have access to a geovisualization of the data. The Leibniz-Institute for Regional Geography Leipzig (IFL) is in the process of creating a web application for the visualization of these commuter relationships.

The development of this application benefits greatly from previously gained experiences involving geovisualizations of commuter and migration data [8] [2].

II. Scope of the Application

Our application will visualize the relevant in-commuter and out-commuter relationships for each of the German municipalities on each of the administrative levels. The application will run within a web browser and will be available to the general public, although particularly sensitive data will not be available in selected areas.

The purpose of the application is to provide a visual analysis of *local and regional commuter patterns* for users with a regional interest. Under a data model perspective these views correspond to a local subset of the commuter matrix.

Users with a more general interest in commuter data will acquire insight from *overview visualizations* depicting aspects of the whole of the nation-wide commuter matrix. These views will offer diverse perspectives upon the total of the commuter matrix.

The primary focus is not to provide tools that satisfy the requirements of experts. Planners, for example, use a number of complex indicators that are derived from commuter relationships (e.g. commuter ratios, work place density, the Independence-Index). Synthetic indicators like these, although helpful to the expert [7], will not play a prominent role. The shift away from synthetic indicators is due to the difficulties of communicating their complex meaning to a larger audience. A *low-threshold access* to visualizations of the total of the German commuter matrix will benefit experts and non-experts alike in making huge amounts of data available for interpretation.

The *core variables* of interest are the number of in- and out-commuters, the balance between source and destination, the distance and angle between source and destination. Out of these core variables a number of interactive images will be created. The purpose of these maps is to visualize the zones of influence of a city or a system of cities and possibly reveal patterns of spatial distribution.

The *time dimension* of these variables will contain yearly values and will be used to display the development of relations over time and to depict a change in the zones of influence.

III. Previous Work

A. Symbolization of Relationships

Linear signatures like arrows, bands and vectors appear as a logical choice to associatively connect commuter data with terms like movement, dynamics, and action. In fact, the pre-attentive perception of the visual system processes the weight/strength and length of linear symbols in the very early stages of perception [9]. This makes both of these properties strong cartographic variables for lines. However, Rae [6] makes clear that a map with thousands of linear thematic map symbols, as it would be often the case with our data, is incomprehensible and imperceptible, even with the most

careful symbolization and on the largest computer screens. It appears to be a symbolization option "for a small number of adequately distributed regions" [3] only.

Also, the length of the symbols is dominating the map and is interfering with the intended visualization of the quantity of the relationship. We proposed earlier [2] to make a move towards a point symbolization and use area diagrams. Whether this is a satisfying solution in all circumstances remains a point of discussion.

B. Reduction of Complexity

One way to handle complexity is to use a filtering approach and decide to visualize the most relevant relations only. Both Pütz [5] and Rae [6] use this technique and decide for relatively high filter threshold values. This results in a grave loss of information, sometimes as high as 99%. We proposed [2] to investigate aggregation of data and entities as another way to handle complexity.

C. Detail, Overview and Context

One specific concern is the visualization of commuter relations that go beyond the map view, specifically when a map is zoomed-in. Baudisch and Rosenholtz [1] propose a technique that provides distance cues for off-screen objects in a context of mobile map use. Can this method be adapted to show quantities as well?

IV. Design Challenges

A serious obstacle to comparison of statistics over time is the "*Modifiable Areal Unit Problem*" [4]. This issue is resolved by the data provider on the level of the statistical data. This allows us to display all data on the most current version of the administrative boundaries. Apart from this, a number of challenges remain.

The *symbolization* of relations will involve a mixture of linear and point symbolization. The choice of the symbolization, the transition between symbolization variants and the role of interaction is subject to investigation. Also, intuitive mnemonic hints need to be given to communicate the movement aspect of the data and to code the direction of commuter balances. The use of animation is one direction to be investigated. Another interesting question is the symbolization of commuter relations that are off-screen while the map is zoomed-in. We take inspiration from the work mentioned above and will continue to experiment with a technique that provides (quantitative) hints to off-screen objects. (We call this technique "orbit".)

Closely interrelated to the symbolization are the questions of *complexity* and the use of aggregation and filtering. One specific direction we will explore is the mixture between filtering and an aggregated display of data suppressed by the filter. In the process of aggregation the hierarchical nature of the statistical data (states – counties – municipalities) will be exploited.

The *temporal aspect* of the data will be visualized through proven methods of (area) diagrams. However, the question remains whether the time aspect can find a reflection in the linear symbolization of relations as well. Also, the development of relations over time and a change in the zones of influence may be quite subtle visually. We will investigate ways to make these changes more prominent visually, for example through temporal aggregation of years.

V. Research Questions

Two obvious yet complex questions will demand attention once the application is ready for dissemination: Is the application easy to use, even without training (low-threshold)? Does it provide effective insight into local as well as regional and national phenomena of commuting?

From a more general research perspective visual aggregation techniques and the visualization of off-screen quantities ("orbit") raise a lot of questions. Does aggregation of entities communicate the complexity of the underlying quantities well? Is the contextualization of map views through a visualization of off-screen quantities helpful?

Acknowledgment

We thank the people at the Statistics Department of the Federal Employment Agency for their interest and their cooperation.

References

[1] P. Baudisch and R. Rosenholtz, "Halo: a technique for visualizing off-screen objects.", *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '03)*. ACM, New York, NY, USA: 2003, p. 481-488.

[2] C. Hanewinkel and S. Specht, "Visualizing Large Sets of Socio-Economic Flow Data", *Proceedings of the 25th International Cartographic Conference*, Paris: 2011.

[3] Q. Ho, P. H. Nguyen, T. Åström, M. Jern, "Implementation of a Flow Map Demonstrator for Analyzing Commuting and Migration Flow Statistics Data", *Procedia - Social and Behavioral Sciences*, Volume 21, 2011, p. 157-166.

[4] M. Madelin, C. Grasland, H. Mathian, L. Sanders, J.-M. Vincent, "Das MAUP: Modifiable Areal Unit – Problem oder Fortschritt?", *Kartographische Visualisierung in der Raumplanung, Informationen zur Raumentwicklung*, Issue 10/11.2009, Bundesamt für Bauwesen und Raumordnung (ed.): 2009, p. 645-660.

[5] T. Pütz, "Siedlungsverflechtungen im Großraum Leipzig.", *Materialien zu Veröffentlichungen des BBSR,* Bundesamt für Bauwesen und Raumordnung (ed.): 2007, http://www.bbsr.bund.de/..., accessed 25.01.2010.

[6] 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, 2009, p. 161–178.

[7] S. Siedentop, "Auswirkungen der Beschäftigungssuburbanisierung auf den Berufsverkehr", *Siedlungsstruktur und Berufsverkehr, Informationen zur Raumentwicklung*, Issue 2/3.2007, Bundesamt für Bauwesen und Raumordnung (ed.): 2007, p. 105-124.

[8] S. Specht, "Hin und weg – Herausforderungen bei der Visualisierung komplexer Wanderungsprozesse.", *Geokommunikation im Umfeld der Geographie; Wiener Schriften zur Geographie und Kartographie*, Volume 19, Vienna: 2009, p.41-46.

[9] C. Ware, "Information Visualization. Perception for Design.", Morgan Kaufmann Publishers, San Francisco: 2004