CO-242

VISUALIZING LARGE SETS OF SOCIO-ECONOMIC FLOW DATA

SPECHT S., HANEWINKEL C.
Institute of Regional Geography Leipzig, LEIPZIG, GERMANY

CHARACTERISTICS OF MIGRATION AND COMMUTER STATISTICS

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.

From a cartographic point of view these data comprise an interesting challenge. On one hand, the immense amount of statistical relationship data for larger spaces is craving for efficient visualizations in support of experts in spatial planning. On the other hand, the pure amount of spatial data promises compelling, highly detailed map images. Additionally, the integration of these data sets into an interactive cartographic information system on the web, which can be regarded as the ultimate medium for such purposes, is an interesting challenge from a software development perspective. In this paper we will present our efforts in prototyping such a system regarding the visualization of commuter data, covering Germany on the level of municipalities.

The Federal Employment Agency (BA) generates statistical data on commuters in the process of the social insurance registration of employed people. Thus, these data do not cover people outside the public social insurance system, like freelancers. The data have been made available as matrices on the administrative levels of states, counties and municipalities (Länder, Kreise, Gemeinden). For our prototype the relevant in-commuter and out-commuter relationships for each of the German municipalities have been used.

For the analysis of the processes involved in commuting, a number of supporting indicators are being used in the relevant areas of research (Siedentop 2007). For some of them, commuter data are being linked to fundamental statistic measures of population and economy. A selection is presented in Table 1.

Table 1: Common statistical indicators on commuter data. (A: out-commuters, E: in-commuters, Ew: population, B: persons employed, AO: at place of work, WO: at place of residence)

<u>I. Entity Indicators</u>		
commuter ratios	in-commuter ratio	(E/B(AO)*100)
	out-commuter ratio	(A/B(WO)*100)
	non-commuter ratio	(B(WO=AO)/B(AO)*100)
work place indicators	work place density	(B(AO)/Ew*1000)
	work place employment ratio	(B(AO)/B(WO)
employment indicators	employment intensity	(B(WO)/(Ew/1000))
	employment ratio	(B(WO)/Ew(15-64J)*100)
	Independence-Index	(B(WO=AO)/(E+A))
<u>II. Relationship Indicators</u>		
commuter count	in-commuter	
	out-commuter	
	commuter balance	(E-A)
	commuter volume	(E+A)

As properties of one spatial entity, the first group of indicators can be visualized through the simple and efficient methods of choropleth or area diagram mapping. A special case, however, are the relationship indicators (group II). Their nature as streams, as properties of a relationship between two spatial entities, implies two geometric references. These streams link two administrative units in such a way, that the out-commuters of one place are the in-commuters of another place. The efficient visualization of these relationships is the ultimate goal of the project presented in the course of this paper.

A Visualization Semantics Considering Varying Target Audiences

Statistical data regarding the relationship of spatial entities are defined by carrying two spatial references. Even though a rare case in thematic mapping of statistical data, linear signatures like arrows, bands and vectors appear as a logical choice to associatively connect the data with terms like movement, dynamics, action etc.

Most maps created for a heuristic analysis of commuter streams use indicators that filter the data either for the source or the target of a movement, thus "fixing" one side of the relationship data. Speaking in map terms, one side of the relationship becomes part of the map title. The "lose" side is in the center of visual interest, its distribution is dominating the map image. One looks at a map of in-commuters of a specific city, because one is interested in the sources of those commuters. Knowing about the target, the idea of a linear motion has been traded for the position of its origins. Figure 1 demonstrates this and shows identical statistical data with a shift in the semantic focus. What differentiates these symbolization alternatives?

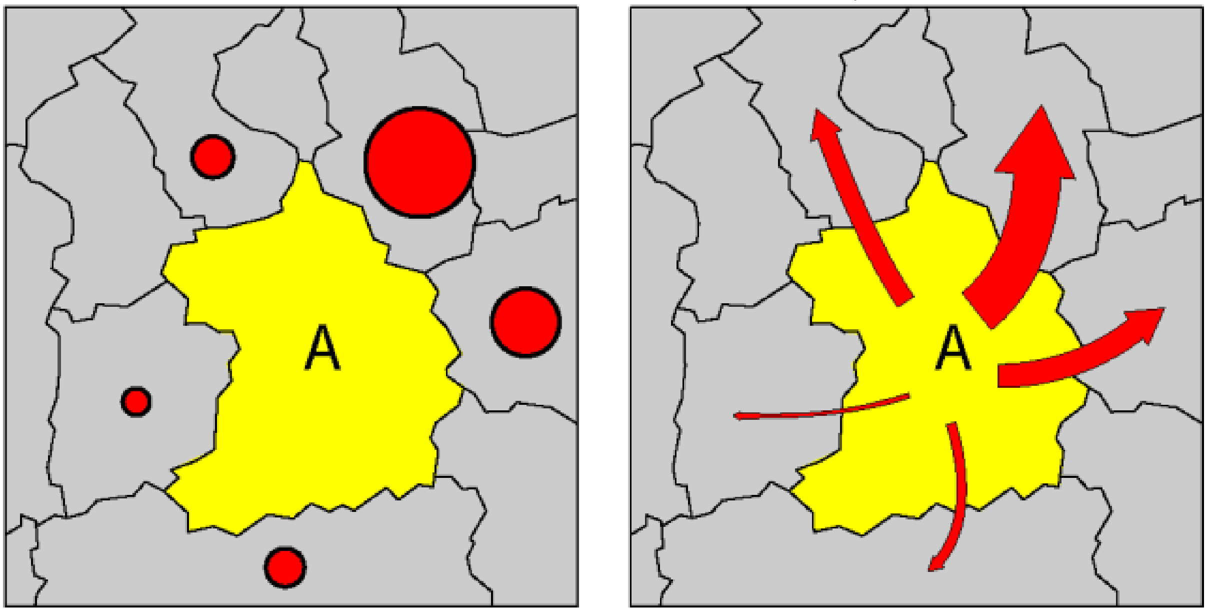

Figure 1: These two maps convey fundamentally different messages, even though they are based upon identical statistical data.

The pre-attentive perception of the visual system processes the weight/strength and length of linear symbols in the very early stages of perception (Ware 2004). This makes both of these properties strong cartographic variables for lines. The map in Figure 2 demonstrates the problem that arises, when we encode the statistical data, our primary focus, to the width of linear symbols. The wanted information is encoded, but the length of the symbols is dominating the map. The "short vs. long" message of the map is interfering with the primary "many vs. few" message. If we follow Tuftes (1983) idea of a "data-ink ratio", we are using a lot of "ink" for the redundant visualization of distance too (by position and line length).

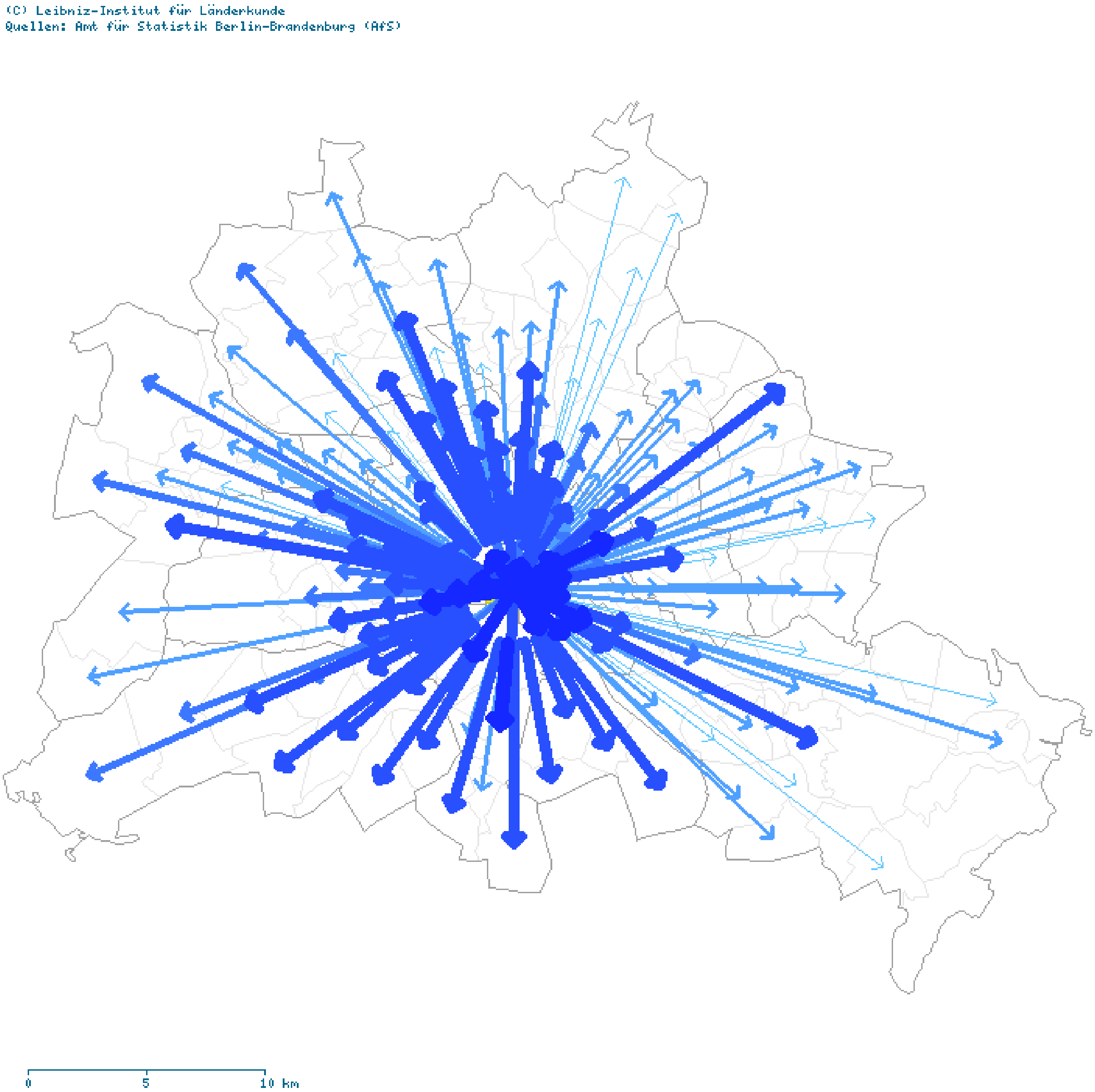

Figure 2: Visualization of out flows from an urban district using linear symbols. Screenshot from an unpublished visualization system.

Point symbols (area diagrams) are free of those side effects. They are effectively visualizing the statistical value of the relationship through symbol size at the position of its origin or destination. Spatial distribution patterns are much clearer, and the distance of the relationship is now irrelevant for the visualization efficiency of the symbol: The short and strong movements are now actually dominating the map. But something essential is lost: The notion of movement. Figure 3 demonstrates a way to overcome this and therefore offers mnemonic movement cues by color and rotation of the symbols.

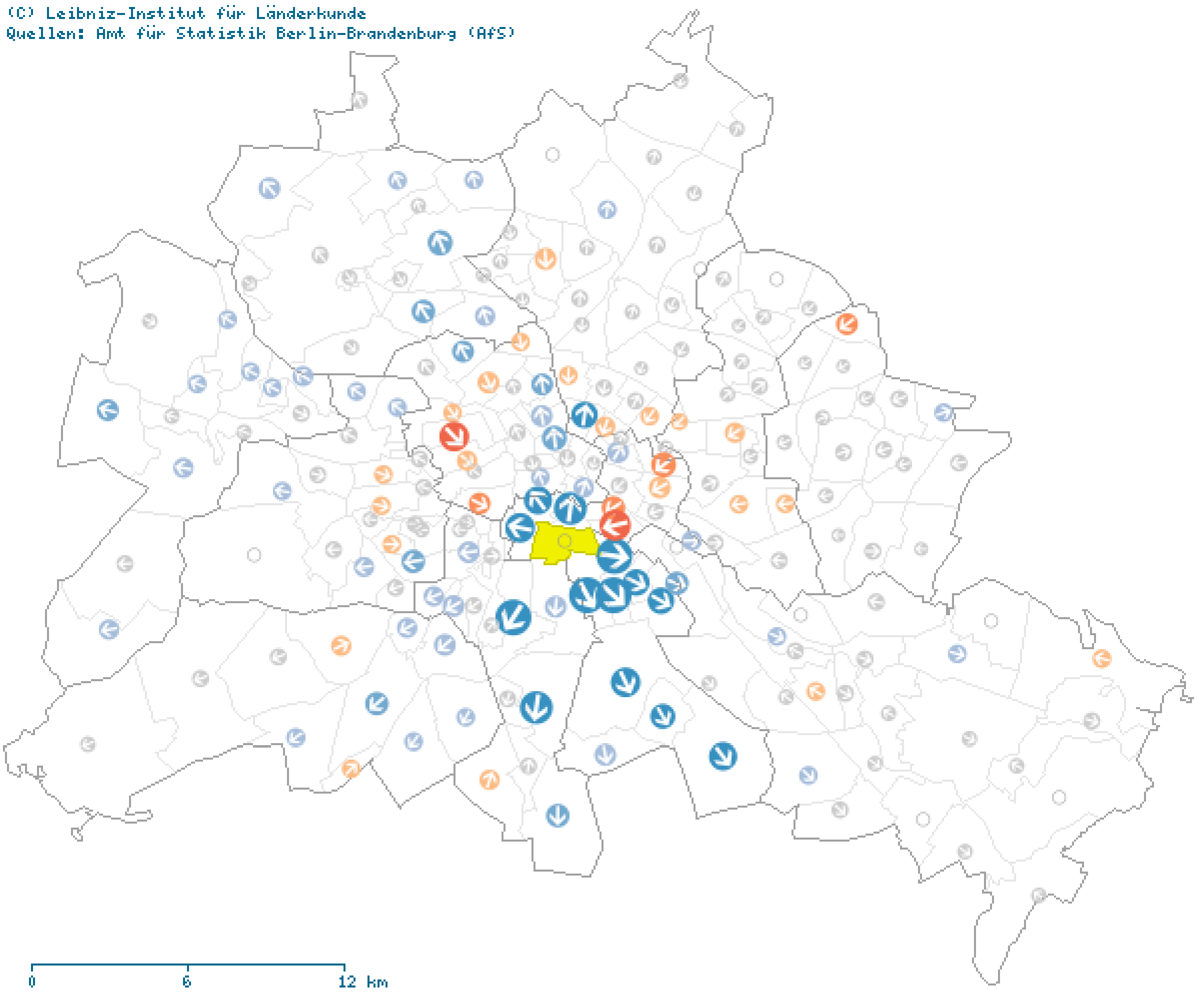

Figure 3: A method offering mnemonic cues for point symbolization. The map shows urban migration flows (migration balance) for a city district of Berlin. The blue and red color corresponds to negative and positive balance values. The direction of the arrows encodes the same information, pointing towards the selected entity for a surplus of in-migration on that relation.

Our experience shows that the loss of a notion of movement is acceptable under certain conditions that depend on the target audience of the map. The domain experts, being familiar with the statistical data, prefer the point symbolization for a visual, heuristic data analysis. After short training they take advantage of its robustness to reveal spatial distribution patterns. They would, however, prefer maps with linear symbolization for communication purposes with colleagues, politicians or the public. As experts of a specific geographical space they are in a position to simplify the complexity of those maps drastically by the aggregation of a number of spatial entities into regions, effectively bundling many movements into one. The feedback we get from end user experts is similar. The creators of exhibitions or atlases for a broader audience of sporadic users prefer maps with linear symbols too. The simple and clear message "something is moving from here to there" is unmatched and often preferred over a higher level of statistical and spatial detail.

PROTOTYPING A COMMUTER FLOW VISUALIZATION

In order to make the visualization of commuter flows to be of value to planners, it must be done on the highest available level of spatial detail. On a regional level the indicators in-commuters, out-commuters and balance need to be depicted for municipalities. For the visualization of long-distance relationships the administrative level of counties (Kreise) and states (Länder) as source or target of a relation is appropriate. Taking the number of 12,500 German municipalities into account, we have a number 150,000 potential maps to create. This number of maps can only be handled by an interactive cartographic system (Hanewinkel 2009).

As we created such a cartographic system, our existing visualization application "hin&weg" served as a basis for our prototype. This existing web application processes and visualizes urban migration data as service for some partnering municipalities (Specht 2009). A guest access to this system together with a version of our commuter flow visualization prototype is available at [http://hinundweg.ifl-leipzig.de].

The prototype offers visualizations of the two fundamental categories of indicators mentioned earlier. The first class of maps provides summarizing information on the spatial entities of all of the administrative levels, using the first group of indicators. This set of simple maps builds the exploratory and interpretative framework for the whole application. These maps provide the necessary entry points for the visualization of commuter flow data, because they allow for a simple heuristic classification of regions. At start-up the prototype presents an overview map containing in-commuter indicators, clearly highlighting the regions with a surplus of in-commuters (Figure 4). Additional maps of the first class (out-commuter ratio, work place density etc.) highlight other aspects of the problem domain, but serve the same purpose: They provide an overview and invite the user to dive deeper into the commuter flow data.

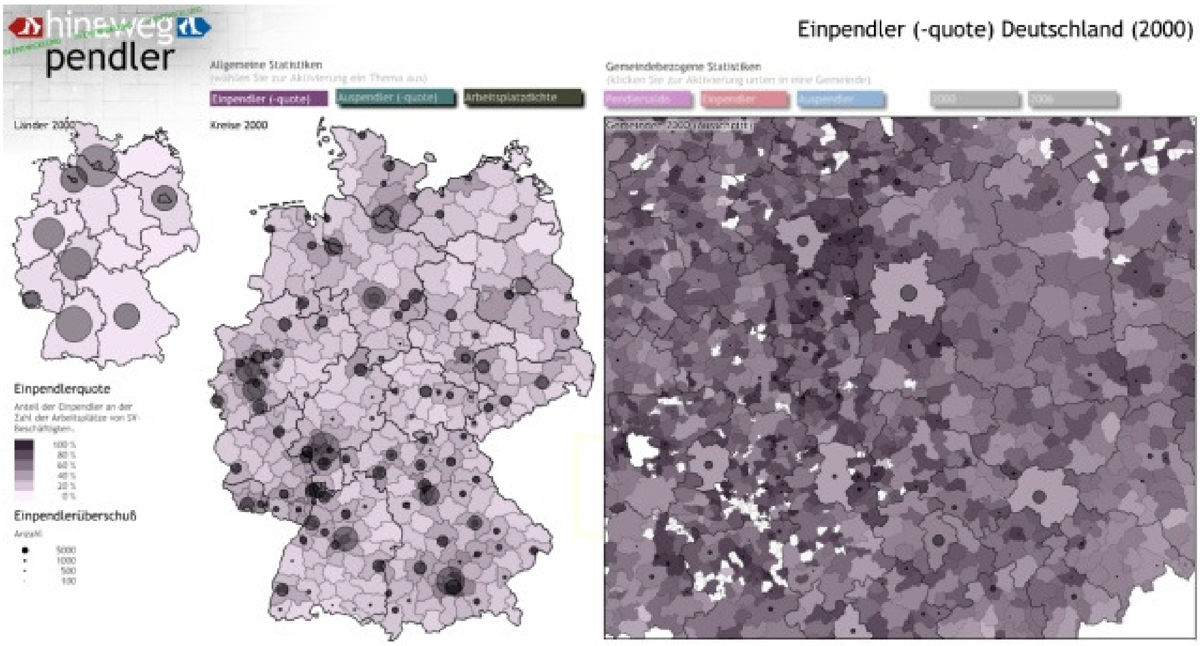

Figure 4: Screenshot of the prototype at start-up. The choropleth map shows the in-commuter ratio, the point symbols visualize the surplus of in-commuters.

The maps on the commuter flow data, representing the second class of maps and our main interest, depict the relationship between spatial entities (in-commuter, out-commuter, commuter balance). These maps are enabled at any time by selecting a municipality on the detailed map at the right of the screen. The commuter count between two spatial entities can be regarded as the atomic data level for the problem domain. The visualization of the balance between in- and out-commuters is especially helpful and reveals interesting spatial patterns (Figure 5).

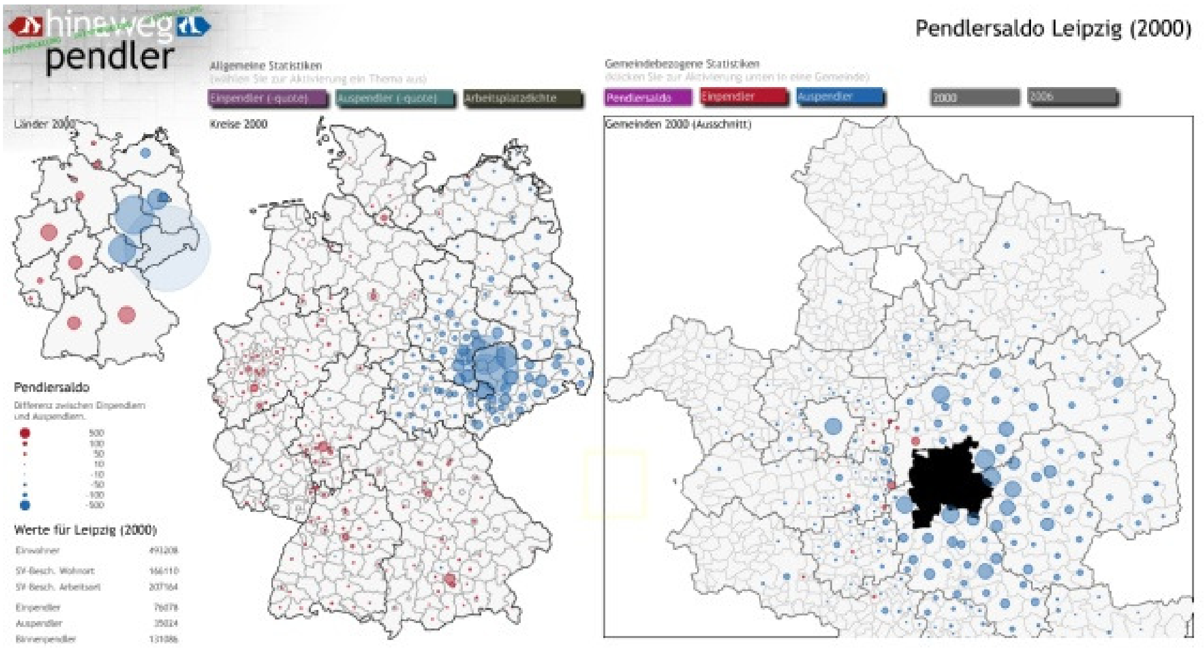

Figure 5: Another screen shot of the prototype, showing a display of the commuter balance of the City of Leipzig (black).

The prototype shows the selected indicator using one map for each of the administrative levels. The map scale has been adapted to fit the level of detail. With the simultaneous visualization of overview and detail we hope to ease the use of the application and allow top-down as well as bottom-up reading of the maps. The multiple-maps-strategy evens out another problem for the visualization of the commuter streams: Privacy matters lead to a loss of data on the level of commuter streams between municipalities. However, these “lost” commuters are registered at the next higher administrative level and can be made visual on those maps.

The maps depicting the administrative level of states and counties constantly display the whole of Germany. Since the majority of commuters travel relatively short distances, the administrative level of municipalities is additionally displayed on a framed map, showing a small, enlarged extract of Germany only. The center and extent of the municipality map is selected simply by clicking into the state map (left) or the county map (in the middle). This design entails a positive user experience, as it avoids heavy loading of the full municipality geometry data at application start-up and, more important, frees the user from the cognitively expensive processes of map navigation by pan and zoom (Ware 2004).

From a planners perspective the map navigation could be extended using circular extracts of the municipality map. The radius of the extract and the selection of the municipalities displayed could be defined by the distance thresholds, usually at 50 km and 150 km (Einig and Pütz 2007). This allows for a differentiation of short distance, long distance and weekly commuters.

THE QUEST FOR AN OVERVIEW

The creation of the prototype is best described as a circular, explorative design and research process. The reflection about it left us with a number of ideas and research questions.

The prototype exhibits excessive detail information through commuter flow maps for all of the municipalities of Germany. There are summary indicators to frame this detail, yet it lacks a country-wide overview presentation of the actual raw flow data. Is it actually possible to visualize all of the commuter relationships between all entities in one map on screen?

The most simple solution is to visualize all relationships through lines connecting the entities, or through line segments pointing towards the target. Rae (2009) shows a selection of solutions and examples for the visualization of the commuter volume. This serves well as an indicator, since it carries no information about direction (opposing to commuter balance, in-commuter or out-commuter values). But then again, this indicator falls short for our purposes, as it provides no insight into the quality of the commuter streams. The commuter balance is an indicator suited far better for such a visualization. But it creates a case even more complex, as it requires the encoding of direction (positive vs. negative balance values).

Rae (2009) makes clear that a map with thousands of linear thematic map symbols is incomprehensible and imperceptible, even with the most careful symbolization and on the largest computer screens. If there is no satisfying way of presenting the complexity, how can the complexity be reduced than? We regard selection and aggregation as two fundamental strategies serving this purpose. However, by reducing the complexity, both trade a loss of information for an optimized perception.

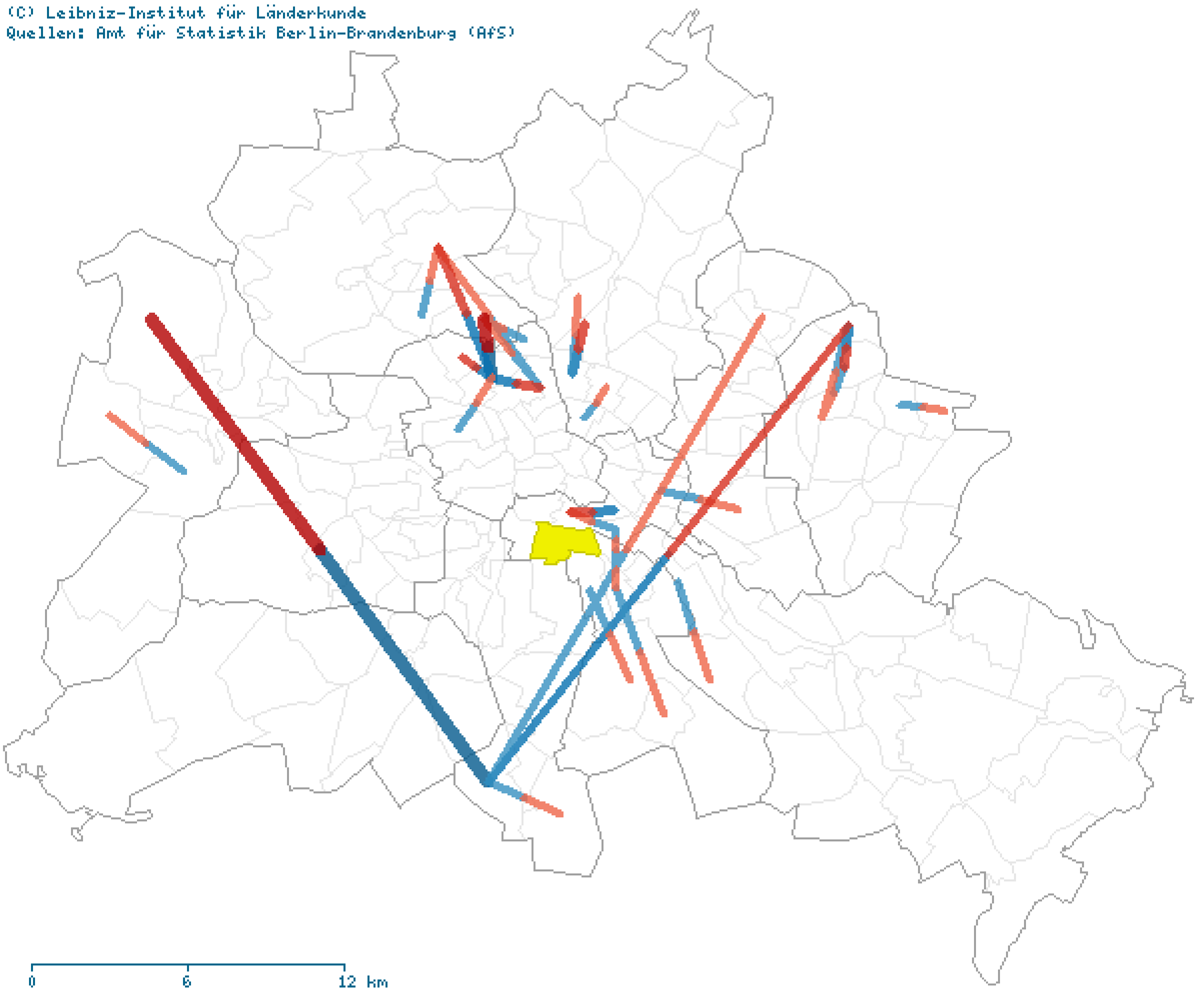

Figure 6: A visualization of the largest migration balance values within the City of Berlin (selection mechanisms applied). The lines start in blue color at the city district with the negative balance and end in red color at the positive side of the relation. Screen shot taken from "hin&weg".

The strategy of selection basically involves the application of a filter to the flow data. This process involves no transformation of the raw data. The visualization in Figure 6 is an example for a radical filtering for the largest balance streams, showing only about 0,05% of the streams of that specific location. Both Pütz (2007) and Rae (2009) use selection and decide for relatively high filter threshold values as well. Can this grave loss of information be minimized? An interactive cartographic system might manipulate the filter threshold in accordance with map scale and symbol density, striving for the best balance between perception and information loss. It might as well animate the threshold, successively showing more and more streams. The indicator value does not need to be the only criteria for selection. For a variety of research questions the distance between the source and the target of a stream is of great help.

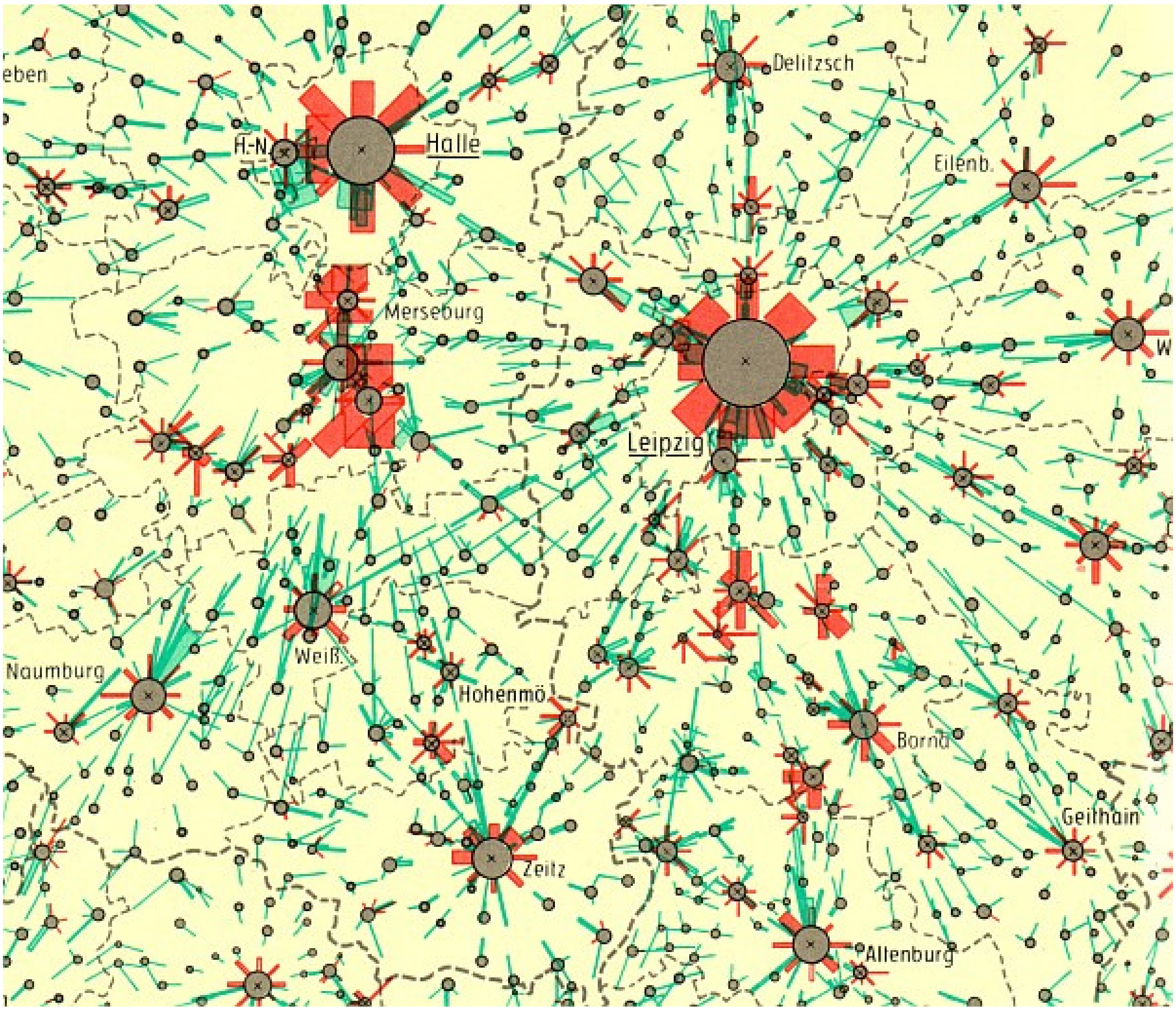

Figure 7: This extract of a map displays the commuter relationships of the East German municipalities (see text, Bölter 1983).

From our point of view aggregation of data and entities is another technique towards a reduction of complexity. In opposition to selection it reduces details of the flow data without biasing the display of the total flow value. The process of aggregation involves the creation of new, non-atomic statistical indicators and requires a common feature for group building. Such a feature can be found in the hierarchy of administrative structure or, as Figure 7 demonstrates, simply in the position in geometric space. This commuter map, from the archive of our institute, dates back to the 1980s and shows in- and out-commuter streams of East German municipalities as red and green line segments, pointing in the direction of source and target respectively. For municipalities with a relatively high surplus of in-commuters, the in-commuter relations are aggregated into sectors of 45° width, based upon the direction of their origin. The commuter relationships are then visualized together as one line per sector. An adoption of this idea for the computer screen can be found in Figure 8, showing an overview of aggregated in- and out-commuter relationships.

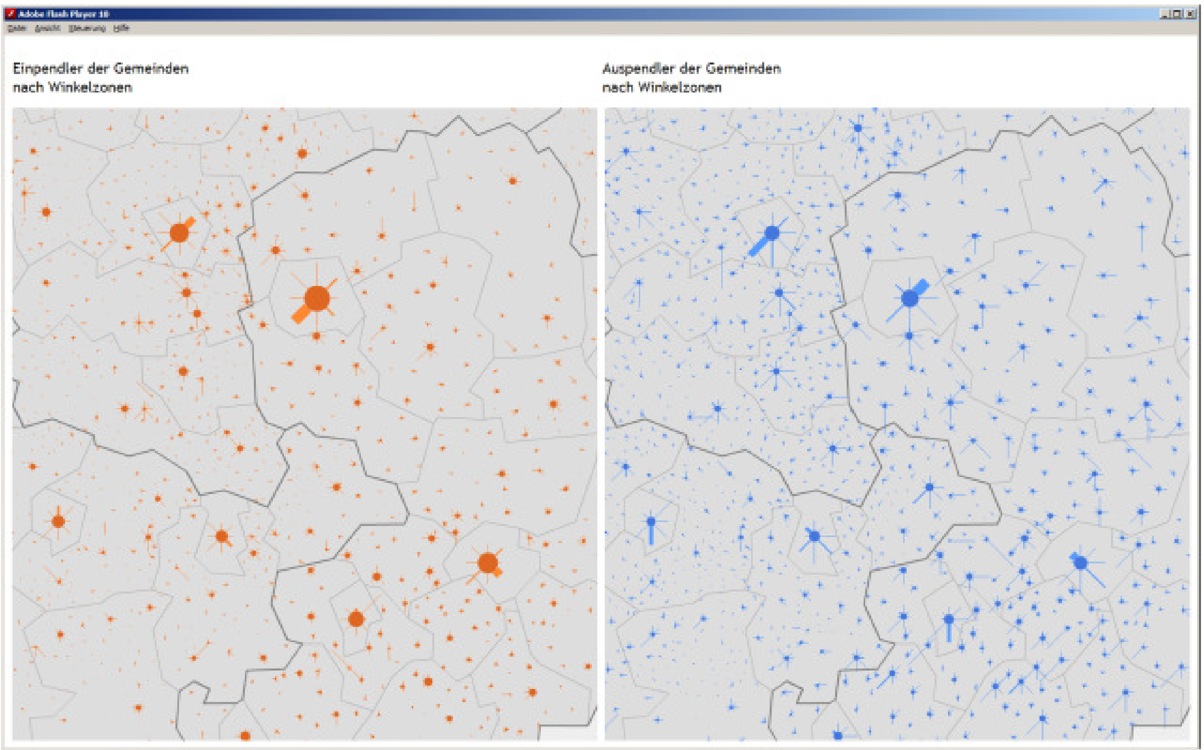

Figure 8: In- and out-commuter streams aggregated by sectors of 45°. Screenshot of a prototype.

OUTLOOK

For us, the purpose of any overview visualization remains in the goal to provide access to detailed information, even to the atomic data set that constitutes the visualization. One way to integrate this idea is a kind of a map scale based aggregation with a high level of aggregation for maps of the entire nation or region and no aggregation when zoomed in to the individual municipalities. Another approach is sketched in Figure 9. It is based upon the idea to trace commuter streams down to their origin. Utilizing interactive map symbols, the user drills down multiple levels of aggregation that involve the geometric position as well as the administrative structure.

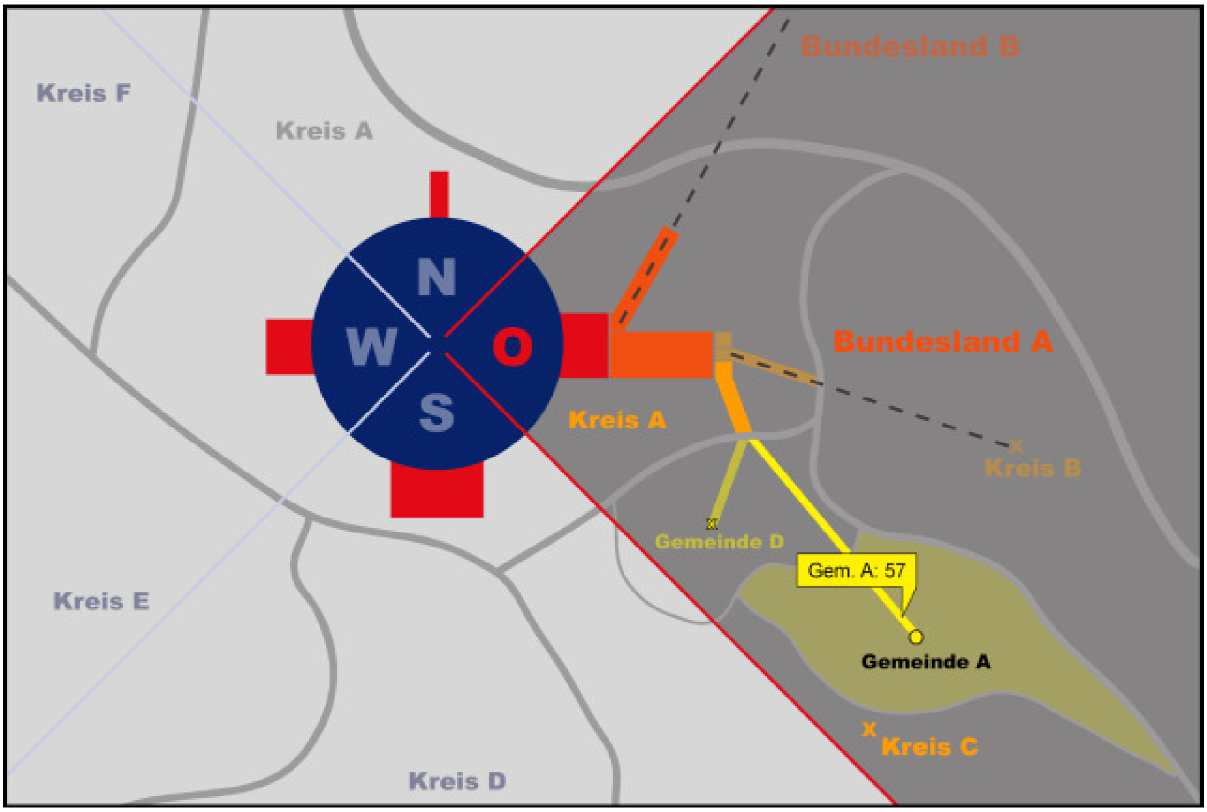

Figure 9: Sketch of an interactive map symbol that allows to trace hierarchical aggregation (red and orange) down to the atomic value (yellow).

REFERENCES

Bölter, P.; Breitfeld, K.; Grundmann, L.; Jeschke, I. (1983) Arbeitspendelwanderung in der DDR. Karte 1:500000. [= Generalschema für die Standortverteilung der Produktivkräfte der DDR, Nur für den Dienstgebrauch], Institut für Geographie und Geoökologie der Akademie der Wissenschaften der DDR, Leipzig; map archive of the IfL.

Einig, K.; Pütz, T. (2007) Regionale Dynamik der Pendlergesellschaft. In: Bundesamt für Bauwesen und Raumordnung (ed.): Siedlungsstruktur und Berufsverkehr, Informationen zur Raumentwicklung, Heft 2/3.2007, p. 73-91.

Hanewinkel, C. (2009) Interaktive Karten. In: Bundesinstitut für Bau-, Stadt- und Raumforschung (ed.): Kartographische Visualisierung in der Raumplanung, Informationen zur Raumentwicklung, Heft 10/11.2009, p. 741-750.

Pütz, T. (2007) Siedlungsverflechtungen im Großraum Leipzig. Materialien zu Veröffentlichungen des BBSR, http://www.bbsr.bund.de/cln_016/nn_495386/BBSR/DE/Fachpolitiken/VerkehrMobilitaet/Verflechtungen/Fachbeitraege/Pendler/PendlerLeipzig,templateId=raw,property=publicationFile.pdf/PendlerLeipzig.pdf, accessed 25.01.2010.

Rae, A. (2009) From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. In: Computers, Environment and Urban Systems, 33, p. 161–178

Siedentop, S. (2007) Auswirkungen der Beschäftigungssuburbanisierung auf den Berufsverkehr. In: Bundesamt für Bauwesen und Raumordnung (ed.): Siedlungsstruktur und Berufsverkehr, Informationen zur Raumentwicklung, Heft 2/3.2007, p. 105-124.

Specht, S. (2009) Hin und weg – Herausforderungen bei der Visualisierung komplexer Wanderungsprozesse. In: Kriz, K.; Kainz, W.; Riedl, A. (eds.): Geokommunikation im Umfeld der Geographie; Wiener Schriften zur Geographie und Kartographie, Band 19, Wien, p.41-46.

Tufte, E. (1983) The Visual Display of Quantitative Information. Graphics Press, Cheshire.

Ware, C. (2004) Information Visualization. Perception for Design. Morgan Kaufmann Publishers, San Francisco.