

# A Visual Analytics Approach to Exploration of Large Amounts of Movement Data

Gennady Andrienko & Natalia Andrienko

Fraunhofer Institute Intelligent Analysis and Information Systems, Germany

<http://geoanalytics.net/>

Thanks to the recent progress in positioning and tracking technologies, data about various mobile objects or agents are currently collected in growing amounts. Analysis of such data can yield valuable knowledge about the movement behaviors of the objects and about the environment in which they move. Since movement data by themselves lack semantics (essentially, the records consist of coordinates and time stamps), their analysis requires the involvement of a human expert, who can interpret the data on the basis of his/her knowledge. This includes the general knowledge of the (geographic) space, time, and movement as well as territory- and domain-specific knowledge.

In order to enable human interpretation of movement data, it is necessary to represent the data in a proper, i.e. visual, way. However, traditional approaches to visualization and interactive exploration of movement data cannot cope with the large amounts of movement data. There is a pressing need in adequate analysis methods. A general approach to handling large datasets includes aggregation and summarization. Aggregation means putting together data items that are close and/or similar. Summarization means deriving characteristics of so formed aggregates (i.e. groups of data items) from the characteristics of their members.

To understand how movement data can be appropriately aggregated and summarized, we have introduced a formal model of collective movement of multiple entities as a function  $\mu: E \times T \rightarrow S$  where  $E$  is the set of moving entities,  $T$  (time) is the continuous set of time moments and  $S$  (space) is the set of all possible positions. As a function of two independent variables,  $\mu$  can be viewed in two complementary ways, which are called *traffic-oriented view* and *trajectory-oriented view*. In the traffic-oriented view (the term “traffic” is used in an abstract sense), the movement is seen as a time-ordered sequence of *traffic situations*, where a traffic situation consists of the positions and movement characteristics of all entities at some time moment. In the trajectory-oriented view, the movement is seen as a set of *trajectories* of all entities, where a trajectory describes the movement of a single entity during the time period under analysis.

For each of the two views, different methods of aggregation and summarization are required. In the traffic-oriented view, it is necessary to aggregate and summarize traffic situations. These basically consist of points in space and point-related characteristics. Therefore, the aggregation and summarization methods suitable for point data can be applied here. In particular, the points can be aggregated by spatial compartments (e.g. cells of a regular grid), by time intervals, which may be defined according to the linear or cyclical model of time, and by values of movement attributes such as direction and speed. The resulting aggregated data can be visualized by means of animated or static maps with the use of coloring or shading, graduated symbols, or diagrams, and non-cartographic displays such as temporal histograms. We particularly suggest two cartographic

visualization techniques: mosaic diagrams for the exploration of cyclical patterns in traffic variation (Fig.1) and directional bar diagrams for the exploration of movements in different directions.

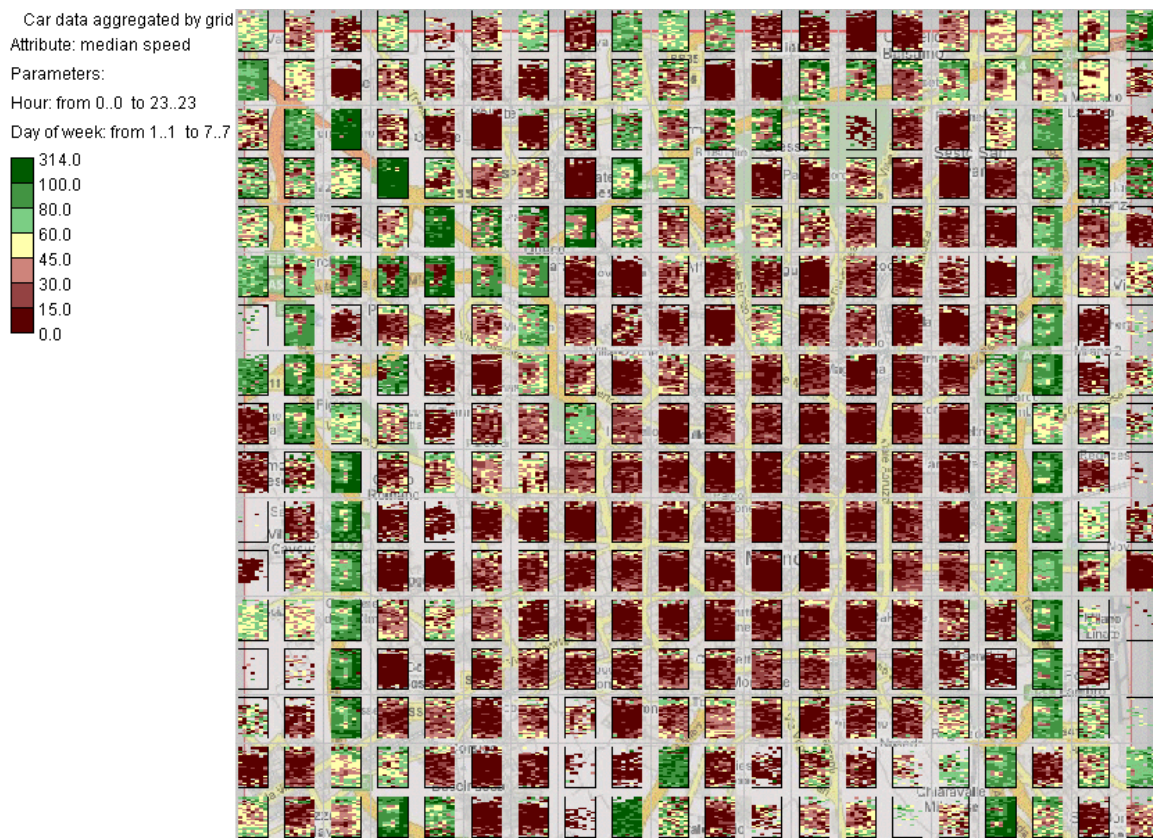


Fig.1: A map with mosaic diagrams.

In the trajectory-oriented view, it is necessary to aggregate and summarize trajectories, which are much more complex objects than points. We have considered two methods of spatial aggregation of trajectories: (1) by the spatial closeness of their starts and ends; (2) by the spatial closeness and geometrical similarity of the routes. The first aggregation uses a previously defined partitioning of the space into areas. The aggregation is done by putting together the trajectories with the starts and the ends fitting in the same areas. The aggregates can be visualized by means of an origin-destination matrix and by a map with vectors (directed lines) varying in their widths and/or colors or shades according to the characteristics of the aggregates.

The second aggregation is based on clustering of the trajectories according to the similarity of the routes. For the clustering, a special similarity (distance) function is used. For the visualization of the results, we suggest a method based on treating trajectories as sequences of moves between small areas, which are defined automatically using characteristic points of the trajectories, i.e. starts, ends, turns, and stops. The areas are built as circles around clusters of characteristic points from multiple trajectories and around isolated points. The aggregation is done by putting together moves between the same areas. To visualize a cluster of trajectories, only the moves from the trajectories of this cluster are aggregated. The aggregated moves are shown on a map by vectors, as in

the case of aggregation by starts and ends. The visualization can be interactively manipulated. Thus, the user may choose to see only the moves occurring in at least  $k$  trajectories, where the parameter  $k$  can be dynamically changed (Fig.2).

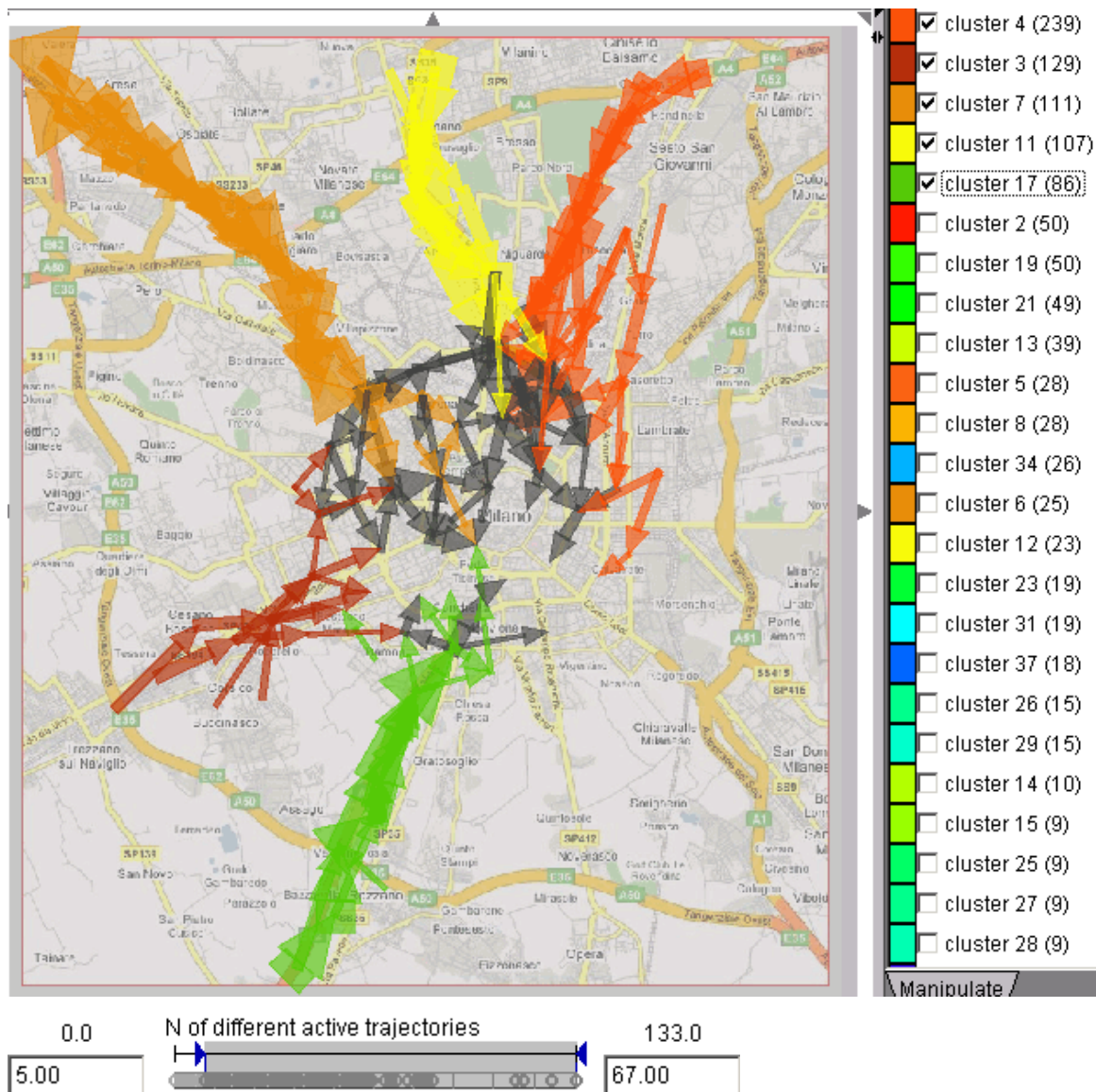


Fig.2: Visualization of trajectories aggregated according to route similarity by means of clustering. Only the moves occurring in at least 5 trajectories are visible. Five biggest clusters, shown in different colors, are selected for viewing. Dark grey is used for the moves occurring in two or more of the selected clusters.

When a dataset is too big for processing in the main computer memory, all the operations on data aggregation and summarization need to be performed in the database, and only aggregated data loaded in the main memory for visualization and interactive exploration. All the suggested aggregation methods except for the route-based aggregation of trajectories can be implemented with the use of standard database operations. For the route-based aggregation, the clustering procedure needs to be implemented so as to work in the database.

However, when the data are not yet too large for in-memory processing but just too large for non-aggregated visualization (due to display clutter), it is possible to do the aggregation in the main memory. This may be done by means of *dynamic aggregators*, which create very interesting possibilities for interactive analysis. A dynamic aggregator keeps references to its members (i.e. the objects it aggregates) and reacts to various interactive operations on the set of the objects such as filtering and classification. In response, it adjusts the values of aggregate attributes and, as a consequence, alters its appearance in visualizations. In particular, the aggregate moves and the generalized places (areas) can be implemented as dynamic aggregators keeping references to trajectories or fragments of trajectories. They react to the temporal filter (selection of a time interval), attribute filter (selection of trajectories by attributes such as duration and length), cluster filter (selection of clusters), and assignment of colors to groups of trajectories resulting from clustering or classification. Being represented on a map as vector symbols, aggregate moves can change their thickness or color and hide from the view when all their members are filtered out. Aggregate moves and places also control the appearance of symbols or diagrams representing values of aggregate attributes on a map or in a matrix display.

### **Related literature**

Andrienko, G., N.Andrienko and S.Wrobel, (2007). “Visual Analytics Tools for Analysis of Movement Data”. ACM SIGKDD Explorations, 9(2): 38-46.

Giannotti, F., and D. Pedreschi, eds. (2007). “Mobility, Data Mining and Privacy - Geographic Knowledge discovery”. Springer, Berlin