Leveraging Spatial Abstraction in Traffic Analysis and Forecasting with Visual Analytics

Natalia Andrienko[1,2], Gennady Andrienko[1,2], and Salvatore Rinzivillo[3]

[1] Fraunhofer Institute IAIS, Sankt Augustin, Germany
1[st]name.2[nd]name@iais.fraunhofer.de
[2] City University London, UK
[3] Istituto di Scienza e Tecnologie dell'Informazione, Area della Ricerca CNR, Pisa, Italy
rinzivillo@isti.cnr.it

Abstract. By applying spatio-temporal aggregation to traffic data consisting of vehicle trajectories, we generate a spatially abstracted transportation network, which is a directed graph where nodes stand for territory compartments (areas in geographic space) and links (edges) are abstractions of the possible paths between neighboring areas. From time series of traffic characteristics obtained for the links, we reconstruct mathematical models of the interdependencies between the traffic intensity (a.k.a. traffic flow or flux) and mean velocity. Graphical representations of these interdependencies have the same shape as the fundamental diagram of traffic flow through a physical street segment, which is known in transportation science. This key finding substantiates our approach to traffic analysis, forecasting, and simulation leveraging spatial abstraction. We present the process of data-driven generation of traffic forecasting and simulation models, in which each step is supported by visual analytics techniques.

1 Introduction

The topic of this presentation, based on [4], is derivation of traffic forecasting and simulation models from traffic data. Traffic data in the form of trajectories of vehicles are currently collected in great amounts, but their potential remains largely underexploited. By means of visual analytics methods, we discovered fundamental patterns of traffic flow dynamics that are common for different areas and spatial scales. On this basis, we created interactive visual interfaces for representing these patterns by mathematical models and devised a lightweight traffic forecasting and simulation algorithm that exploits these models. We developed interactive visual embedding for defining initial conditions, running simulations, and analyzing the outcomes. Since simulations could be prepared and performed very fast, thus allowing interactive operation, our tools allow the users to imitate various interventions altering network properties and/or traffic routes and investigate their impacts on the traffic situation development, including comparative analysis of various “what if” scenarios.

adfa, p. 1, 2011.

2 Approach summary

Given a set of trajectories, we apply a method [2] that derives an abstracted network consisting of territory compartments (further called cells) and links between them. In brief, the method organizes points sampled from the trajectories into groups fitting in circles of a user-specified maximal radius. The medoids of the groups are taken as generating seeds for Voronoi tessellation. Smaller or larger cells (Voronoi polygons) can be generated by varying the maximal circle radius, thus allowing traffic analysis at a chosen spatial scale. Moreover, it is possible to vary the spatial scale across the territory depending on the data density [1]. Next, the trajectories are transformed into *flows* (aggregate movements) between the cells by time intervals. For each pair of neighboring cells (C_i, C_j) and each time interval T_k, the flow is an aggregate of all moves from C_i to C_j that ended within the interval T_k and started within either T_k or T_{k-1}. The flow is characterized by the number of moves and the mean speed (velocity) of the movement. The number of moves (traffic volume) per time interval is called *traffic intensity* (a.k.a. *traffic flow* or *flux*). Since available trajectories typically cover only a sample of vehicles that move within a network and not the entire population, the computed traffic intensities need to be appropriately scaled, to approximate real intensities. Appropriate scaling parameters or functions can be derived by comparing the computed vehicle counts with measured counts obtained from traffic sensors [7].

To study and quantify the relationships between the traffic intensities and mean speeds, the data are further transformed in the following way. Let A and B be two time-dependent attributes associated with the same link and defined for the same time steps. The value range of attribute A, which is taken as an independent variable, is divided into intervals. For each value interval, all time steps in which values from this interval occur are found, and all values of attribute B occurring in the same time steps are collected. From these values of B, summary statistics are computed: quartiles, 9th decile, and maximum. For each statistical measure, a sequence of values of B corresponding to the value intervals of A is constructed. These sequences are called *dependency series*. We first take the traffic intensity as the independent variable and derive dependency series of the mean speed. Then we take the mean speed as the independent variable and derive dependency series of the traffic intensity. Dependency series may be derived using either the absolute or relative traffic intensities, the latter being the ratios or percentages of the absolute intensities to the maximal intensities attained on the same links.

In Fig. 1, two maps on the left represent abstracted transportation networks of Milan with different levels of spatial abstraction. Curved lines in the upper map and half-arrow symbols in the lower map represent the network links. On the right of each map, the upper graph shows the dependencies of the mean speed on the relative traffic intensity. The horizontal axis corresponds to the traffic intensity and the vertical axis to the 9th decile of the mean speed (this statistical measure is less sensitive to outliers as the maximum). The lower graph shows the dependencies of the relative traffic intensity on the mean speed. The horizontal axis corresponds to the mean speed and the vertical axis to the maximal traffic intensity. The network links have been clustered by similarity of the speed-intensity dependencies. The coloring of the link sym-

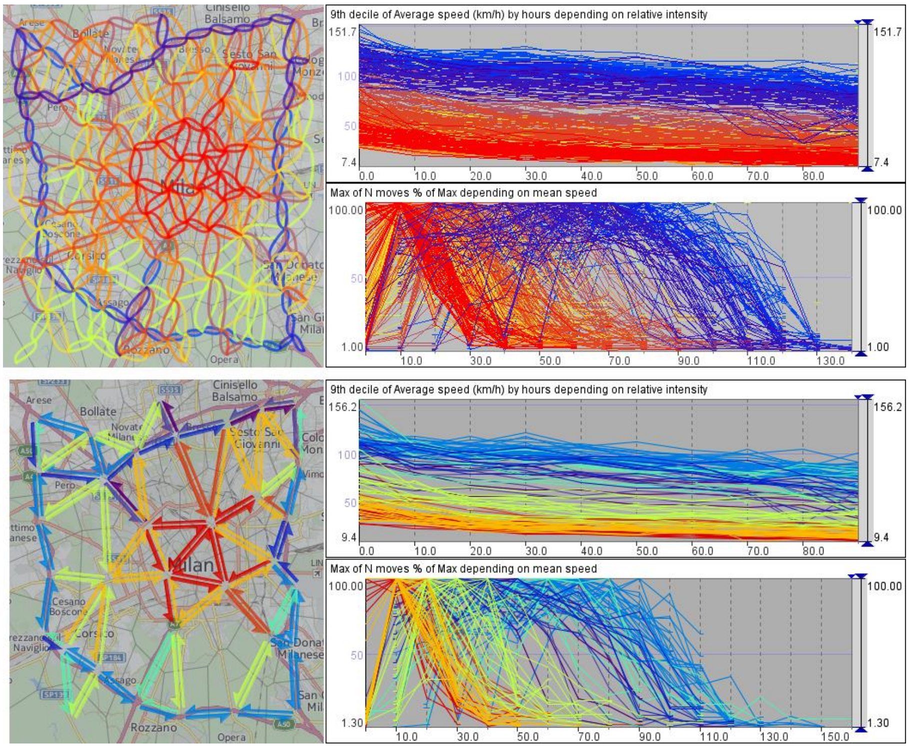

Fig. 1. The maps show spatially abstracted transportation networks of Milan with cell radii about 2km (top) and 4 km (bottom). The graphs to the right of each map represent the dependencies between the relative traffic intensities and the mean speeds on the network links.

bols on the map and lines in the graphs represents the cluster membership. The shapes of the dependency lines are very similar to the curves in the fundamental diagram of traffic flow describing the relationship between the flow velocity and traffic flux [5, 6] in a physical transportation network consisting of street segments. We see that the same relationships exist also in a spatially abstracted network. Moreover, we have found that the relationships conforming to the fundamental traffic diagram exist on different levels of spatial abstraction, as illustrated in Fig.1.

We have developed interactive visual tools supporting derivation of formal models from the time series of flow characteristics and from the dependency series [3]. Models are built for clusters of links rather than individual links, to avoid over-fitting and reduce the impacts of noise and local outliers. Predictions made for link clusters are individually adjusted for each link based on the statistics of its original values [3]. We have also developed a novel traffic simulation algorithm that can directly work with the derived models. The main idea is following: for each link, the algorithm finds how many vehicles need to move through it in the current minute, determines the mean speed that is possible for this link load (using the dependency model from the traffic intensity to the mean speed), then determines how many vehicles will actually be able to move through the link in this minute (using the dependency model from the mean

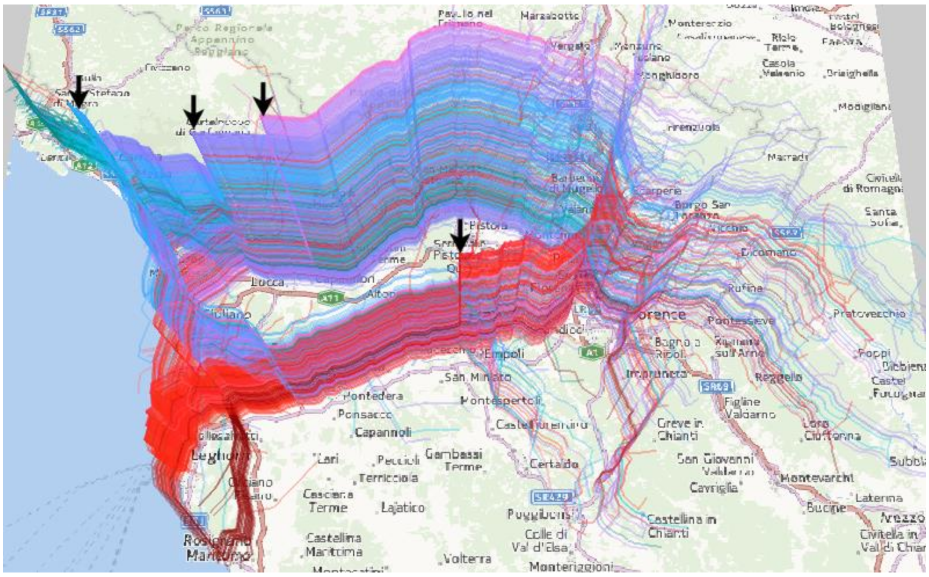

Fig. 2. For a scenario of mass evacuation from the coastal areas in Tuscany (Italy), simulated car trajectories are represented in a space-time cube, where two dimensions represent geographical time and one dimension time. The arrows point at the places of major traffic suspensions.

speed to the traffic intensity), and then promotes this number of vehicles to the end place of the link and suspends the remaining vehicles in the start place of the link.

To perform a simulation, the analyst defines a scenario. A wizard guides the analyst through the required steps and providing visual feedback at each step. We describe the simulation of a scenario of mass evacuation from the coastal area in Tuscany (Italy). The appendix to the paper (http://geoanalytics.net/and/is2015/) includes a video demonstration of the process of model building, scenario definition, simulation, and exploration of results supported by interactive visual interfaces.

References

1. Andrienko, G., Andrienko, N., Bak, P., Keim, D., and Wrobel, S. *Visual Analytics of Movement. Springer*, 2013.
2. Andrienko, N., Andrienko, G. Spatial Generalization and Aggregation of Massive Movement Data, *IEEE Trans. Visualization and Computer Graphics*, 17(2): 205-219, 2011.
3. Andrienko, N., Andrienko, G. A Visual Analytics Framework for Spatio-temporal Analysis and Modeling, *Data Mining and Knowledge Discovery*, 27(1): 55-83, 2013.
4. Andrienko, N., Andrienko, G., and Rinzivillo, S. Leveraging Spatial Abstraction in Traffic Analysis and Forecasting with Visual Analytics, *Information Systems*, 57(1): 172-194, 2016; Appendix: http://geoanalytics.net/and/is2015/
5. Gazis, D.C. *Traffic Theory*, Kliwer Academic, Boston, USA, 2002.
6. http://en.wikipedia.org/wiki/Fundamental_diagram_of_traffic_flow
7. Pappalardo, L., Rinzivillo, S., Qu, Z., Pedreschi, D., and Giannotti, F. Understanding the Patterns of Car Travel. *European Physics Journal Special Topics*, 215, 61–73, 2013.