Pacific Graphics (2018)
H. Fu, A. Ghosh, and J. Kopf (Editors)

Short Paper

A Visual Analytics Approach for Traffic Flow Prediction Ensembles

Kezhi Kong[1], Yuxin Ma[1], Chentao Ye[1], Junhua Lu[1], Xiqun Chen[2], Wei Zhang[1] and Wei Chen[1]†

[1]State Key Lab of CAD&CG, Zhejiang University, China
{durantkong,mayuxin,yct,akiori,zhangwei}@zju.edu.cn, chenwei@cad.zju.edu.cn
[2]College of Civil Engineering and Architecture, Zhejiang University, China
chenxiqun@zju.edu.cn

Abstract

Traffic flow prediction plays a significant role in Intelligent Transportation Systems (ITS). Due to the variety of prediction models, the prediction results form an intricate structure of ensembles and hence leave a challenge of understanding and evaluating the ensembles from different perspectives. In this paper, we propose a novel visual analytics approach for analyzing the predicted ensembles. Our approach models the uncertainty of different traffic flow prediction results. The variations of space, time, and network structures of those results are presented with the visualization designs. The visual interface provides a suite of interactions to enhance exploration of the ensembles. With the system, analysts can discover some intrinsic patterns in the ensemble. We use real-world urban traffic data to demonstrate the effectiveness of our system.

CCS Concepts

•***Human-centered computing*** → ***Visual analytic;***

1. Introduction

Traffic data analytics and visualization have been extensively studied for decades [AAD*08, CGW15] and applied in many areas such as urban planning and city security. In this work, we focus on traffic flow prediction, a major functional component in Intelligent Transportation Systems (ITS) during the past decades. Various types of traffic prediction models were developed, including time-series models [LCZ14, CZZ17, JZC14] and deep learning models [TP15].

Understanding predictive outputs from these models is a challenging task. With massive computation resources, analysts can often run multiple models with different parameter settings simultaneously, thereby harvesting an ensemble of prediction results for the same road network. However, this brings another cumbersome task, that is, evaluating, comparing, and analyzing the resultant ensemble data. Visualization and visual analytics methods are widely used to depict "commonalities, differences and trends" in ensemble data [OJ14]. Despite a large amount of work dedicated to visual analysis of ensemble data, especially graph ensemble data [YC17, GHL15, SNG*17], it remains challenging to visually analyze ensembles of spatio-temporal graph structures, including 1) non-trivial visual design of spatio-temporal graphs and 2) comprehensive and contemporary analysis of the outcome ensembles from multiple prediction models.

In this paper, we present an iterative visual analytics approach for traffic flow network ensembles generated by multiple prediction models. By cooperating with domain experts, we derive a list of tasks to guide our design. To support visual exploration of flow patterns, we employ advanced pattern mining methods based on k-nearest neighbor (k-NN) graphs. Our approach leverages a visual analysis pipeline to support presenting and exploring well-established traffic flow prediction models. The key components of the visual interface are a set of visual designs, i.e., a k-NN graph view of clustered ensembles, a multi-level map view of ensembles on the basis of road networks, and a flow ensemble view of a set of similar ensembles. In summary, the main contributions of our research are in the following three folds: 1) a novel scheme that uses spatio-temporal graph ensembles to analyze the uncertainty induced by traffic flow prediction models, 2) a suite of visual designs to depict spatio-temporal and structural characteristics of multiple prediction models, and 3) an effective and scalable visual analytics system that allows users to summarize, investigate, and compare predicted traffic flows.

2. Analytical Pipeline

Ensemble Modeling and Analysis Predicted flows from multiple models are imported. We first employ general uncertainty measurements of flow network ensembles on all roads. Then, we use distance-based methods to extract k-NN graphs of ensemble flows.

Visual Comparative Exploration In this visual analysis stage,

† Wei Chen is the corresponding author.

DOI: 10.2312/pg.20181281

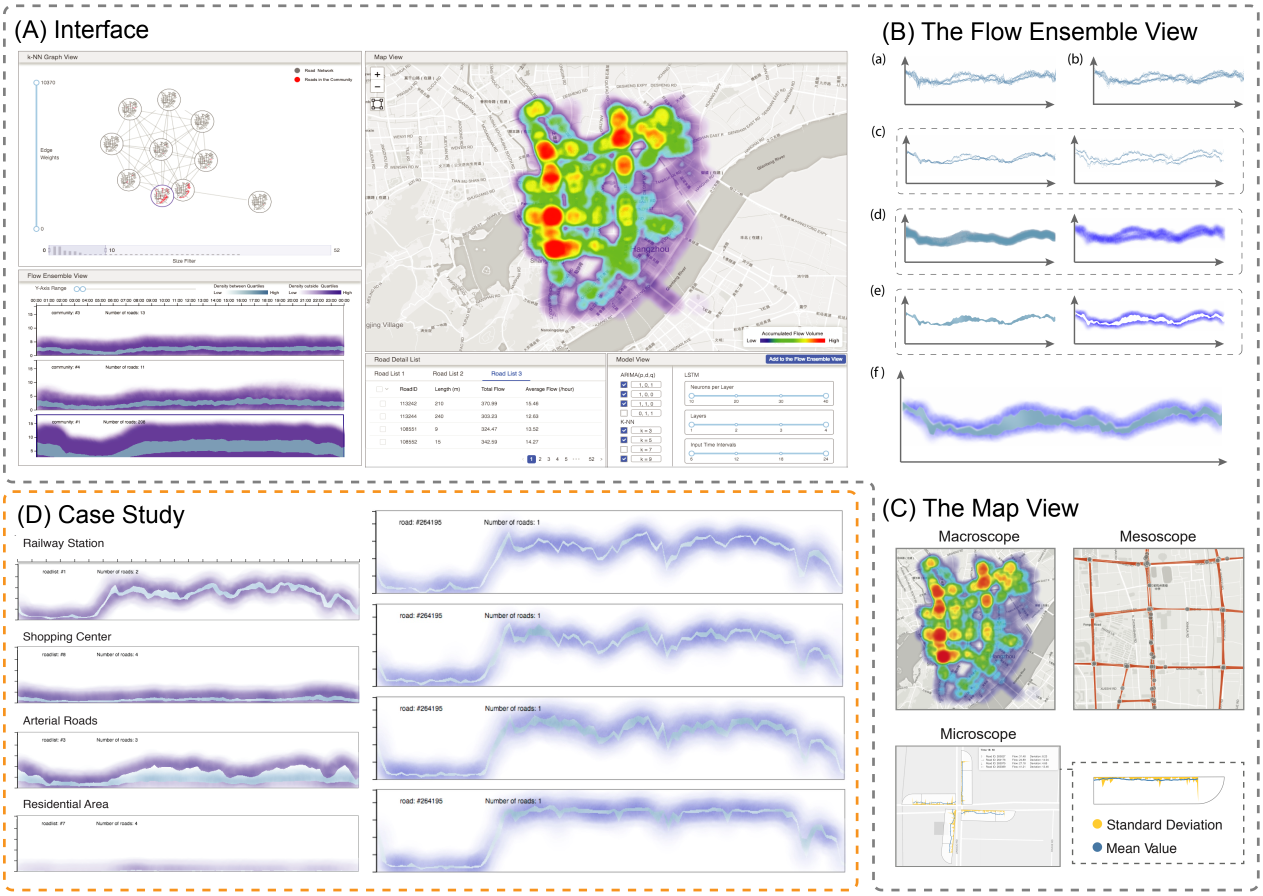

Figure 1: *(A) The visual interface of our system. (B) An illustration of the flow ensemble view. (C) Visual design of the multi-level temporal map view. (D) The flow ensemble views (Left) of four functional regions selected in the case study, and (Right) four different parameter configurations.*

we provide two analytical schemes to explore ensemble flows from two different aspects. We use a map view with heatmaps to present a global picture of road attributes, such as flow volumes and uncertainty. To examine and explore k-NN relations among all roads, a k-NN graph view is provided to show the top-k correlations and summarize patterns in the k-NN graph. When specific road groups are selected, the flow ensemble view can be adopted to visualize and reveal detailed differences in temporal, spatial, topological, or uncertainty-related features between the two groups. All views are tightly coupled through the strategy of multiple coordinated views in order to enable a comprehensive analysis of ensemble flows and fulfill various analytical tasks.

3. Ensemble Analysis

Traffic flows comprise a road network and vehicles' passing records in the network. In a road network presented, the nodes are a set of fixed points which typically consist of road intersections, the starting or ending points of a road, entrances/exits of some places, etc. Formally, we define the road network as a graph $G = \{N, E\}$ where $N = \{n_1, n_2, \cdots, n_p\}$ and $E = \{e_1, e_2, \cdots, e_q\}$ represent the set of p nodes and q connected roads, respectively. Traffic flow $f_{e,t}$ is defined as the number of vehicles (denoted by $\#veh$) that pass over a road e during a time interval t: $f_{e,t} = \#veh_{e,t}$. Specifically, we use $F_{G,t} = \{f_{e,t}, e \in E\}$ to represent the set of flows in a road network G, and F_{e,t_1,t_k} as the flow series from time t_1 to t_k on road e, namely $\{f_{e,t_i}, i = 1, \cdots, k\}$.

With historical traffic flow records F_{e,t_1,t_k} for a specific road or road network, the purpose of a flow prediction model is to forecast the future flow volumes $\{f_{e,t_j}, j = k+1, \cdots, k+h\}$. Here, h denotes the prediction horizon of the model, which indicates the length of the predicted time series. Two domain experts suggested a list of model types in research and applications, including 1) Autoregressive Integrated Moving Average model (ARIMA), 2) k-nearest neighbors prediction (k-NN), and 3) Long short-term memory networks (LSTM) [TP15]. Specifically, we define $Pre_{m,e,h} = \{F_{m,e,t}, t \in [k+1, \cdots, k+h]\}$ as the flow predicted by a model m on road e for the next h time intervals. We consider $Pre_{m,e,h}$ as an ensemble member, and the flow ensemble of all predictions on a specific road e is defined as $En_M(e,h) = \{Pre_{m,e,h} | m \in M\}$ for all trained models M.

To represent distributions and variations of all the model outcomes, we use the standard statistical measures to summarize the flow prediction ensembles. Within each time interval, the mean flow volume of all M models on road e is represented as $\overline{Pre}_{M,e,t}$, and the standard deviation as $\hat{Pre}_{M,e,t}$. To compute the k-NN graphs of different flow ensembles, a proper algorithm should be able to quantify the similarities between ensembles and perform advanced pattern mining. In our work, the MUNICH algorithm [AKKR09] is selected. Based on the similarities derived from the algorithm, we employ k-NN graphs to model top-k relations among the flow ensembles in the following two steps:

Construction of the k-NN Graph In the context of our work, the roads are considered as nodes in the target k-NN graph. For each node that represents a road, a directed edge is connected from the road to its k most similar roads based on the similarity of flow ensembles. Thus, a graph with $|E|$ nodes and $k \cdot |E|$ directed edges can be constructed by iterating this procedure on every road.

Simplification and Clustering It should be noted that the k-NN graph is a directed graph. To simplify the graph structure and enhance correlations among vertices (roads), we transform the k-NN graph into an undirected one by keeping edges between all mutually-connected node pairs, that is to say, retaining the relations where two roads appear in each other's top-k list. The simplified undirected graph reflects strong connections of similar flow ensembles. We further perform clustering with the Girvan-Newman community detection algorithm [GN02] to retrieve strongly-connected partitions.

4. Visualization Design

The Map View The map plays a crucial role in presenting geospatial information and tieing up other views which utilize spatial relations among roads or road networks. Our core idea is to embed temporal glyphs into the road network and support smooth transitions and rich interactions on multiple levels. As illustrated in Figure 1 (C), our visual design is separated into three different levels: 1) On the *microscopic (road-wise)* level, the viewport of map view only shows several connected roads that come from one or two blocks. The sparsity of roads enables visual encoding of additional information in the free space among different roads. Thus, a glyph is embedded next to the starting point of a road, which contains a line chart with the time axis alongside the road. In the line chart, mean values of the flow ensemble for the road are marked as a blue line, while the filled area indicates the standard deviation. A white bounding box is drawn under the glyph as the background. 2) On the *mesoscopic (community-wise)* level, the road network becomes denser than those in microscopic viewports, which consequently reduces the space to place the glyphs. Hence, on this level only the node-link diagram of the road network is presented. 3) The *macroscopic (city-wise)* level is intended to provide an overview of the flow volume distribution in the entire urban area. We use the 2-D kernel density estimation (KDE) to generate a density map of the entire road network by uniformly sampling points along each road. The number of sample points is proportional to the accumulated flow volume of the road. Thus, a 2-D KDE is generated for each sample point. The analysts can zoom in and out among the three levels. Additional details are displayed by hovering on roads on mesoscopic and microscopic levels. Moreover, highlighting a coordinate on glyphs in the same viewport is linked in order to facilitate identification and comparison of mean and deviation values along time axes.

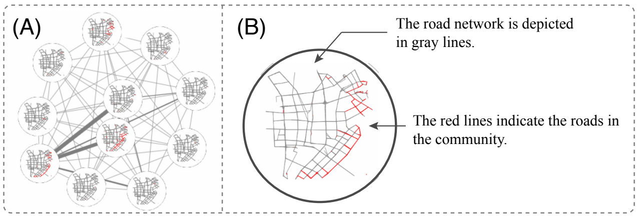
(A)
(B)
The road network is depicted in gray lines.
The red lines indicate the roads in the community.

Figure 2: *(A) An illustration of the k-NN Graph View. (B) The design of the glyph.*

The k-NN Graph View The k-NN graph view is designed to present top-k correlations and neighboring patterns in a specific time period. We employ snapshots of maps to visualize the corresponding flow ensembles of roads in communities, therefore the task to summarize flow ensemble patterns is directly supported. Figure 2 (B) illustrates the design of the glyph that encodes a community. Inside a circle, a snapshot of the entire road network is depicted with gray lines. The roads that are contained in the community is highlighted in red in the snapshot network. Thus, the glyph provides an overview of spatial distributions for the roads. We use the layout of glyphs to encode relations of corresponding communities in the k-NN graph. The k-NN graph is firstly summarized into a graph of communities. The edge weights between communities are equal to the number of edges that connect roads from different communities in the original k-NN graph. The summarized graph, as shown in Figure 2 (A), is visualized with the force-directed layout algorithm with edges connected between community glyph. The line widths are adopted to indicate the edge weights. Basic panning and zooming operations are supported in the canvas. The detailed information of a community can be expanded by clicking on the corresponding glyph. To filter out small communities, the analysts can use a range selector on a bar chart where the communities are ordered by the number of contained roads.

The Flow Ensemble View To visualize temporal trends and distributions of the roads, we design a density-based flow ensemble view. Our design is a hybrid combination of line charts and box plots. As shown in Figure 1 (B), each time series is presented by using one line segment per time series. The visual clutter caused by a vast number of line segments can be eliminated by using the curve density estimation [LH11], yielding a density map of lines. To further depict the statistical values, two layers of curve density maps are employed. First, line segments in the original line charts are uniformly sampled into a set of sample points. Then, all the first and third quartiles at each time step are composed to form a stripe-like contour. The sample points are naturally split into two groups: the ones that lie inside the contour between two quartiles, and their counterparts. Finally, two 2-D KDE-based density maps are generated with respect to the two sample point groups and then composed into a single density map. For each density map, the configurations of grouped roads like the number of roads and the average flow volume of all the roads are shown. In terms of interactions, the analysts are allowed to hover the mouse pointer on a density map. All the time series that pass through the hovering position will be highlighted as lines.

Other Widgets The model view shows all the model types and parameters used for flow prediction. To summarize the flow volumes in the entire urban area, an area chart for describing the accumulated mean volume of all flow ensembles is rendered as the background. When a road or a road group is highlighted in other views, details of the roads are listed in the road detail view.

5. Evaluation

We conducted a case study on a real city-wise dataset. The traffic flow dataset for our evaluation contains taxi trajectory logs from a city in China for 30 days. It contains 1,253 roads and about 8,000 taxis. We set 15 minutes as the time step, yielding 2,880 time intervals. We adopt 88 different models from various types (including the ARIMA, *k*-NN, and LSTM) with different parameter settings to generate prediction results for the 31st day (96 time intervals).

The mobility patterns of taxis, e.g., the spatial distributions and temporal patterns, are relevant to the land use pattern in an urban region. These relations are implicitly presented in the data, and further influence the training of the prediction models and the parameter sensitivity. In this case, we seek to study the variances of the model stability in regions with distinctive land use patterns. Based on the daily life experiences, we choose several representative regions with different developments like shopping centers, residential area, railway stations, and arterial roads shown in the left side of Figure 1 (D). Generally, there are regional variances in terms of the total flow volume. Compared to the residential area, the other three types of regions have much larger flow volumes, larger variances of flow volumes, and more branches in the flow ensembles.

We further study the parameter sensitivity of prediction models on a representative road of a region. The flow ensembles of all prediction models for a main corridor near a railway station are displayed in the right side of Figure 1 (D). It illustrates the results generated by modulating different parameter configurations of LSTM. It is shown that the number of neurons and the number of layers have few influences on the results, while the window size has a significant influence on the stability of the prediction results.

6. Conclusion and Future Work

In this paper, we propose a visual analytics approach for analyzing spatio-temporal graph ensembles generated by traffic flow prediction results. By tackling the features of space, time, graph structures, and a variety of models, we conduct a comprehensive analytical pipeline for presenting, exploring, and comparing different patterns. A suite of visual design from multiple perspectives is adopted.

One promising extension is to incorporate other data sources such as traffic events, and other types of sensors or GPS logs into the analysis process. Currently, our system can be used as a validation tool for traffic flow prediction models. In the future, model design components can be embedded into our system to support online examination and design of models in an integrated environment.

Acknowledgements

This work is supported by the National Key Research and Development Program (2018YFB0904503) and National Natural Science Foundation of China (61772456 and U1609217).

References

[AAD*08] Andrienko G., Andrienko N., Dykes J., Fabrikant S. I., Wachowicz M.: Geovisualization of Dynamics, Movement and Change: Key Issues and Developing Approaches in Visualization Research. *Information Visualization 7*, 3 (2008), 173–180. 1

[AKKR09] Assfalg J., Kriegel H.-P., Kröger P., Renz M.: Probabilistic similarity search for uncertain time series. In *Proceedings of International Conference on Scientific and Statistical Database Management* (2009), pp. 435–443. 3

[CGW15] Chen W., Guo F., Wang F.-Y.: A survey of traffic data visualization. *IEEE Transactions on Intelligent Transportation Systems 16*, 6 (2015), 2970–2984. 1

[CZZ17] Chen X. M., Zahiri M., Zhang S.: Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. *Transportation Research Part C: Emerging Technologies 76* (2017), 51 – 70. 1

[GHL15] Guo H., Huang J., Laidlaw D. H.: Representing Uncertainty in Graph Edges: An Evaluation of Paired Visual Variables. *IEEE Transactions on Visualization and Computer Graphics 21*, 10 (2015), 1173–1186. 1

[GN02] Girvan M., Newman M. E.: Community structure in social and biological networks. *Proceedings of the National Academy of Sciences 99*, 12 (2002), 7821–7826. 3

[JZC14] Jiang X., Zhang L., Chen X. M.: Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. *Transportation Research Part C: Emerging Technologies 44* (2014), 110 – 127. 1

[LCZ14] Li L., Chen X., Zhang L.: Multimodel Ensemble for Freeway Traffic State Estimations. *IEEE Transactions on Intelligent Transportation Systems 15*, 3 (2014), 1323 – 1336. 1

[LH11] Lampe O. D., Hauser H.: Curve density estimates. *Computer Graphics Forum 30*, 3 (2011), 633–642. 3

[OJ14] Obermaier H., Joy K. I.: Future Challenges for Ensemble Visualization. *IEEE Computer Graphics and Applications 34*, 3 (2014), 8–11. 1

[SNG*17] Schulz C., Nocaj A., Goertler J., Deussen O., Brandes U., Weiskopf D.: Probabilistic Graph Layout for Uncertain Network Visualization. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 531–540. 1

[TP15] Tian Y., Pan L.: Predicting short-term traffic flow by long short-term memory recurrent neural network. In *Proceedings of IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)* (2015), pp. 153–158. 1, 2

[YC17] Yan K., Cui W.: Visualizing the uncertainty induced by graph layout algorithms. In *Proceedings of IEEE Pacific Visualization Symposium* (2017), pp. 200–209. 1