Understanding Traffic using Network Science
2016 IGERT Bootcamp Project
Mentor: Arlei Silva (arlei@cs.ucsb.edu)

1 Overview

Understanding human mobility is one of the greatest challenges for improving quality of life in big cities [1, 2]. Recently, several efforts have tried to address mobility related problems using a data-driven approach. For instance, in traffic planning, many cities (e.g. San Francisco[1], Beijing[2]) have applied traffic data in order to design more effective transportation infrastructure. Another related problem is monitoring, which consists of identifying not only the main traffic patterns but also major disruptions (e.g. accidents, a music festival), as means to support better traffic management in the so called Smart Cities [3, 4].

In this project, we will investigate how network science can help us to understand human mobility. We provide you a road network and speed data collected by Caltrans' sensors (cameras) spread over CA roads[3] (see Figure 1). The data is collected along one month and the time granularity is 5 minutes. The idea is to use the dataset and your data/network science skills to answer some initial questions and then give you freedom to explore further questions that might be of your interest. The sky is the limit!

2 Questions

1. Can you distinguish the main speed patterns in the data (e.g. weekday/weekend, day/night)? How to visualize/characterize these patterns?

2. How easy is it to predict the speeds? Pose one or more prediction tasks and try some simple approaches (e.g. linear regression, moving average). Now assume only some pairs location-time are missing, how can you fill the missing data? How the quality is affected by the amount of information available?

[1] http://www.sfcta.org/modeling-and-travel-forecasting

[2] https://www.technologyreview.com/s/425553/gps-data-on-beijing-cabs-reveals-the-cause-of-traffic-jams/

[3] http://www.dot.ca.gov/dist07/travel/traffic/

Figure 1: Caltrans camera.

3. How to cluster locations in terms of speeds? Compare a completely structural approach (e.g. graph partitioning) against a feature-based one. Try to interpret the clusters.

4. How would you find anomalous patterns in the data? Try a simple approach and show the results.

3 Data and Code

Please check the project's github repository: `https://github.com/arleilps/IGERT-Traffic`. You can download the dataset at: `https://drive.google.com/open?id=0B69hCaWL5B4ARkl3YjEyY19QT1E`.

3.1 Data

There are two versions of the dataset: a small one with 100 sensors and a larger one with 2K sensors. There are two different types of file:

- traffic.graph: Edge list of the road network (assume undirected)
- traffic_X.data: Snapshots with speed measurements (vertex,speed), where X is a timestamp

3.2 Code

We provide some python code to read the data, visualize the network and to extract date information from timestamps. See `https://nbviewer.jupyter.org/github/arleilps/IGERT-Traffic/blob/master/igert-traffic.ipynb` for usage.

4 What is Expected

Answers to some of the questions proposed and maybe other related questions in the form of a presentation. Please include plots, visualizations, and main lessons learned during the project. You can also add questions you would like to explore in the future.

5 Skills Required

- Python basics (syntax, loops, functions etc.).
- Basic Statistics (average, distribution etc.)
- Data science basics (clustering, anomaly detection etc.)

6 What will you Learn?

- Hands-on network science
- Python (scipy/numpy, notebooks etc.)
- Basic data visualization (graphviz)

7 Mentoring

The mentor will meet with the students daily for one hour. Extra meetings can be scheduled if needed. Mentor can also be contacted via email.

References

[1] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[2] C. Guo, C. S. Jensen, and B. Yang. Towards total traffic awareness. *ACM SIGMOD Record*, 43(3):18–23, 2014.

[3] A. Ihler, J. Hutchins, and P. Smyth. Adaptive event detection with time-varying poisson processes. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 207–216. ACM, 2006.

[4] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi. Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1):22–32, 2014.