New York Traffic Patterns

Vaibhav Karve, Derrek Yager ¹

University of Illinois at Urbana-Champaign

vkarve2@illinois.edu, yager2@illinois.edu

October 5, 2016

Supervisors: Richard Sowers and Daniel Work

¹Support from National Science Foundation grant DMS 1345032 MCTP: PI4: Program for Interdisciplinary and Industrial Internships at Illinois is gratefully acknowledged.
Motivation

- Better tracking algorithms
- Takes into account traffic dynamics
- Taxi data for NYC
The Raw Data

- Nodes = Intersections (95,581 total)
- Links = Roads, one for each direction (260,855 total)
- For each link, we have the number of taxis/hr and the average speed
- Amount of data
  \[
  24\text{hr} \times 365\text{ days} \times 4\text{ years} \times 260,855\text{ links} \\
  \approx 9 \times 10^9\text{ entries}
  \]
Goals

Compression:
- Compress all this data to some traffic signatures
- Use NMF and get “reasonable” error
- Represent links as linear combination of signatures
- Insist on certain sparsity conditions

Estimation:
- Fill in missing data using matrix factorization
Daily Traffic Patterns

Figure: Comparison on link #169017 vs. top 3 signatures
Future Work

- Continuing as IGL project
- Visualization of results: 
  *what does all this numerical information look like spatially?*
- Characterize links having less data: 
  *recognize holes so data collectors know where to go.*
- Extend to other cities/data: 
  *does Chicago, LA, London, SF traffic behave the same as NYC?*
- Our results could be useful to city planners for traffic management and disaster mitigation.
- Could also be useful to NYC Taxi Association, Google and Uber.