Human Mobility Analytics and Services Initiative (HUMANS)

Ian Lane, Ole Mengshoel, John Paul Shen, Pei Zhang
ECE Department, Silicon Valley Campus

Hae Young Noh, Zhen (Sean) Qian
CEE Department, Pittsburgh Campus

Carnegie Mellon University

A New Research Initiative at CMU-SV

- Research Vision:
  "Distilling Human Mobility Data into Valued Services for Societal Good"
- Research Horizon:

1–2 years	2–3 years	3–5 years	5–10 years
Industry product roadmap	Industry R&D	Nearer-term academia research	Longer-term academia research

- Research Approach:
  - In-situ deployment of end-to-end experimental systems
- Research Theme:
  - Human mobility data from industry partners
  - Behavior models and predictive intelligence
  - Services for mobile users, communities, and enterprises

Human Mobility Analytics and Services Systems

SENSING → **ANALYTICS** → **SERVICES**

Data from diverse sensing platforms:

- Human sensing
- Vehicle sensing
- Environment sensing
- Infrastructure sensing

Behavior models and predictive intelligence:

- Mobile population
- Connected vehicles
- Transportation systems
- Wireless networks

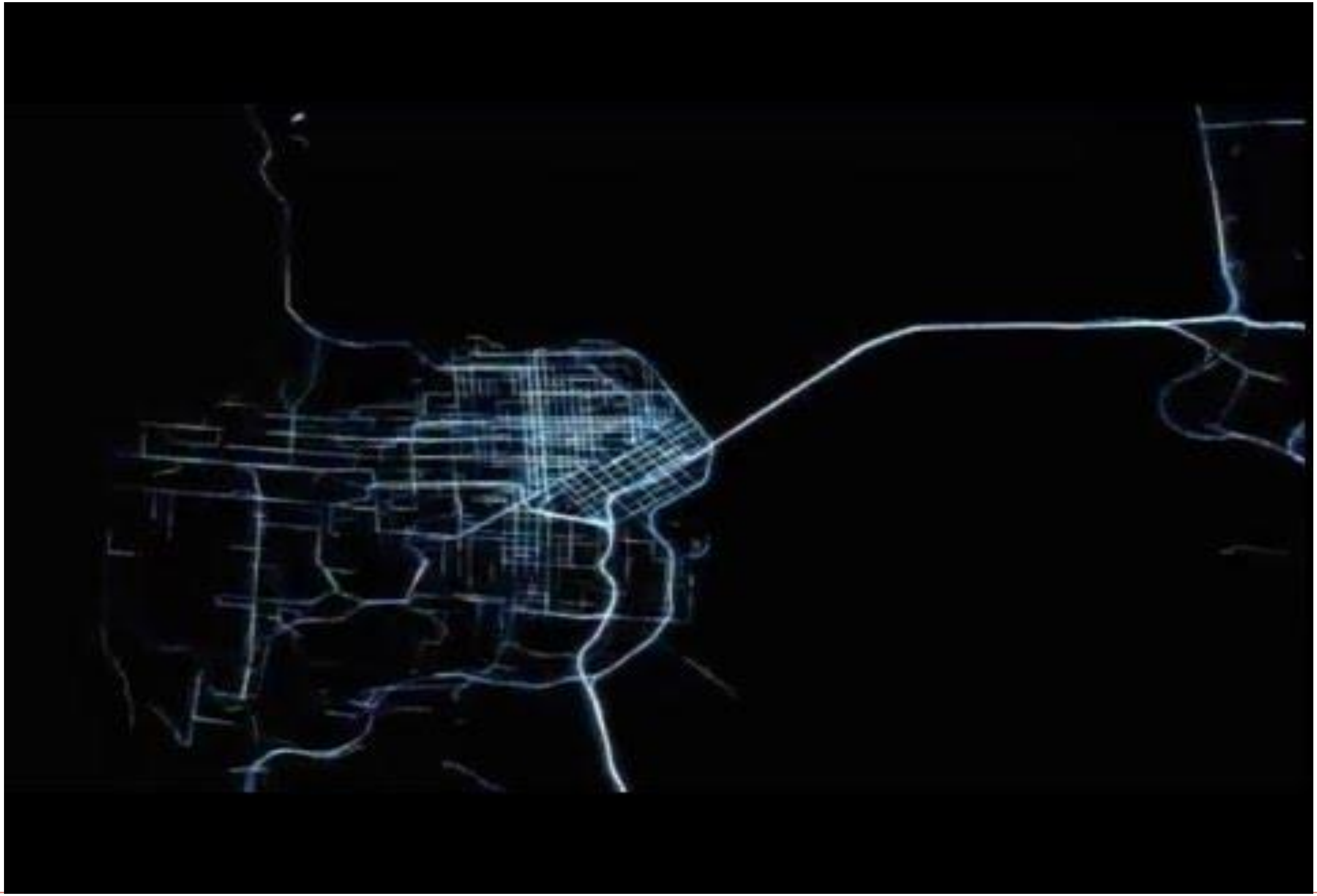
Diverse forms of useful & specialized services:

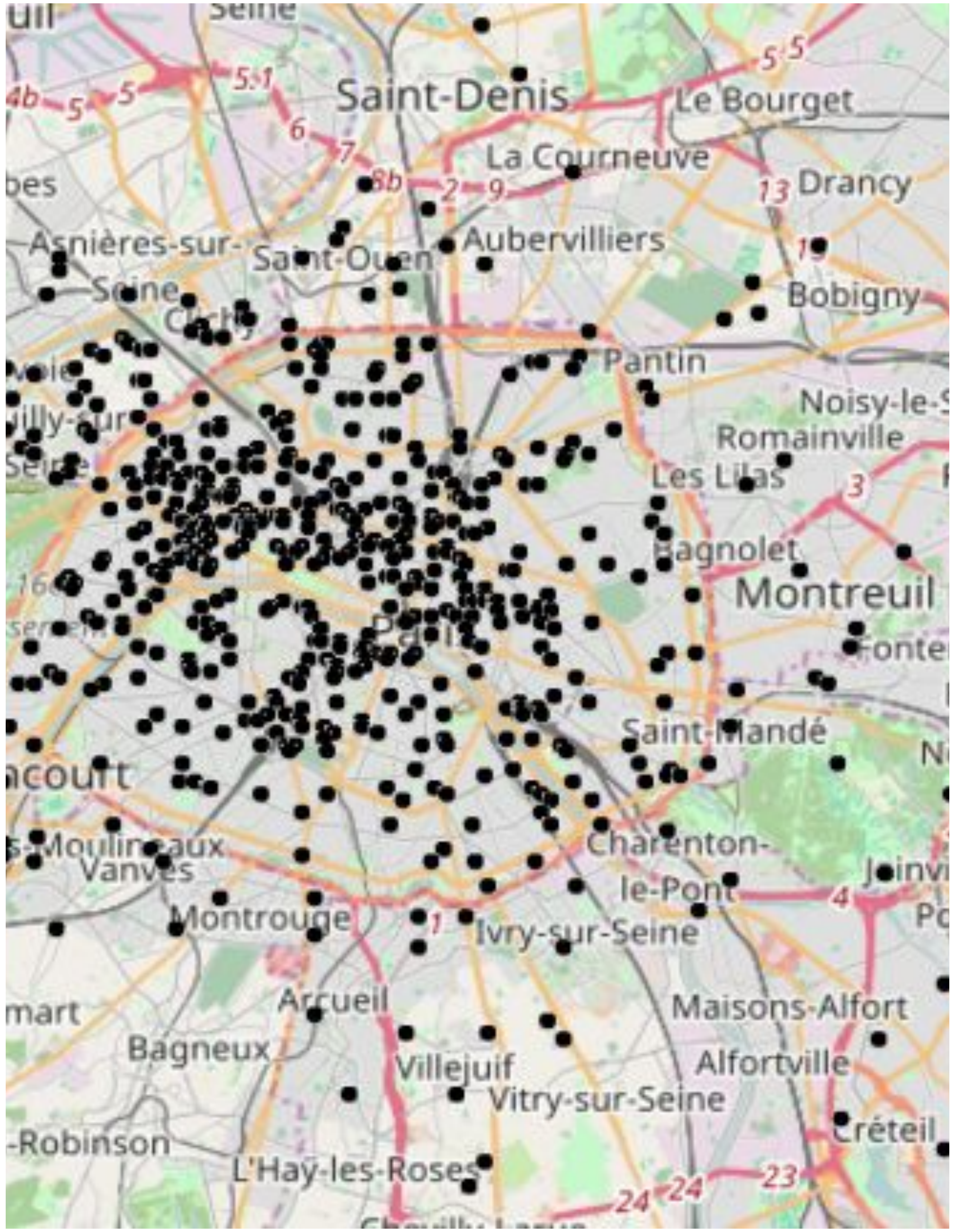
- Mobile users services
- Public services and policy management
- Enterprise business intelligence

“Human Mobility Analytics and Services: Insights from a Ride-Sharing Service”

John Paul Shen & Abhinav Jauhri
ECE Department, CMU (Silicon Valley Campus)
November 17, 2016 (Traffic21 Seminar)

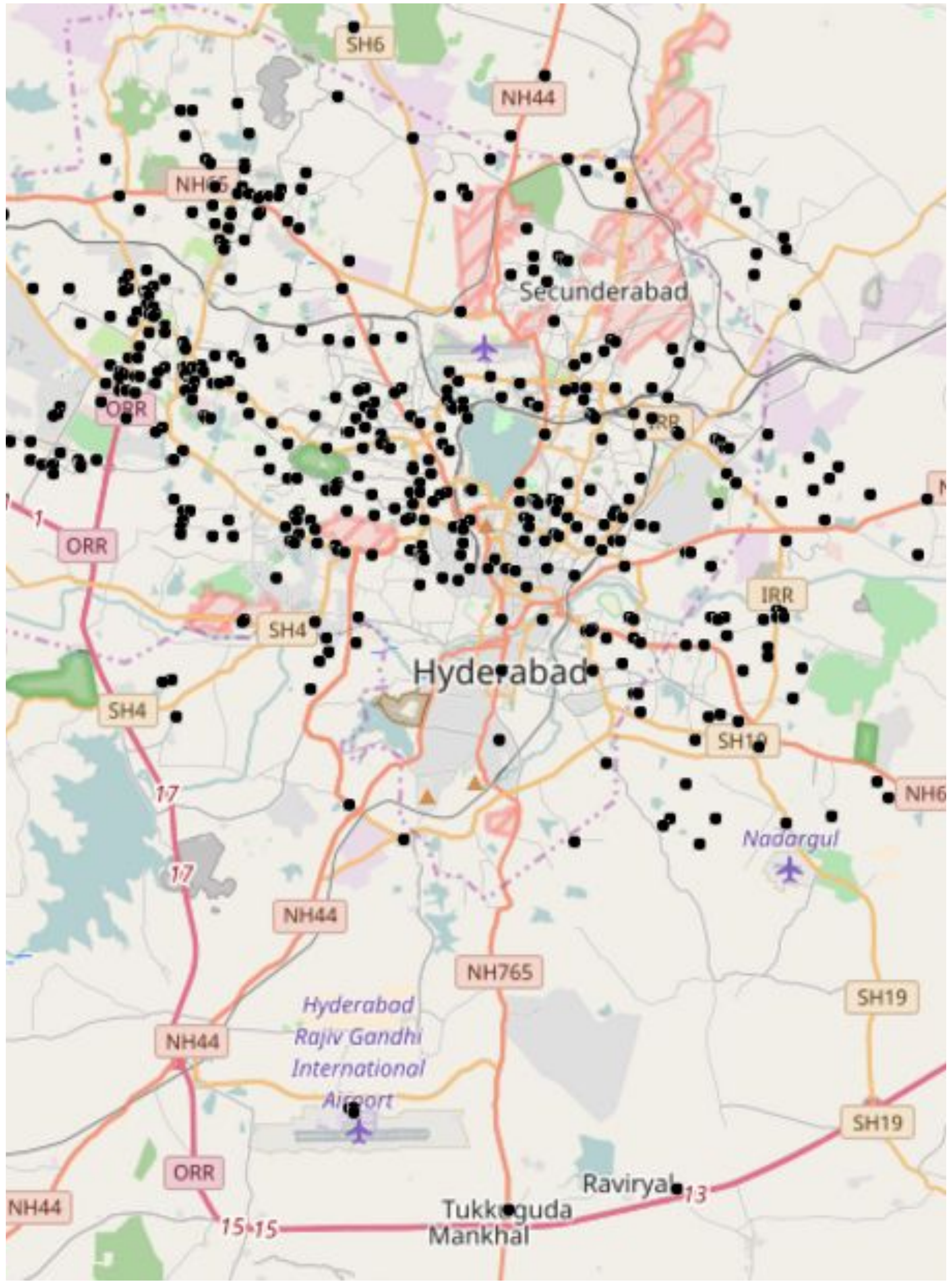
ABSTRACT:
This talk presents some initial results and insights from our collaboration with a global ride sharing service company. Through this collaboration we have access to ride request data from 400+ cities in the world. We focused on ride requests from the top 40 cities and developed a space-time graph model that captures the spatial and temporal variations of ride requests in a city. Based on this graph model, we can characterize the “*poolability*” of a city (i.e. the % of ride requests that can be pooled). Many cities exhibit potential poolability in the 30% range. We are developing a ride pooling algorithm that can exploit this potential, with the goal of reducing the number of vehicles on the roads and potentially alleviating city traffic congestions.



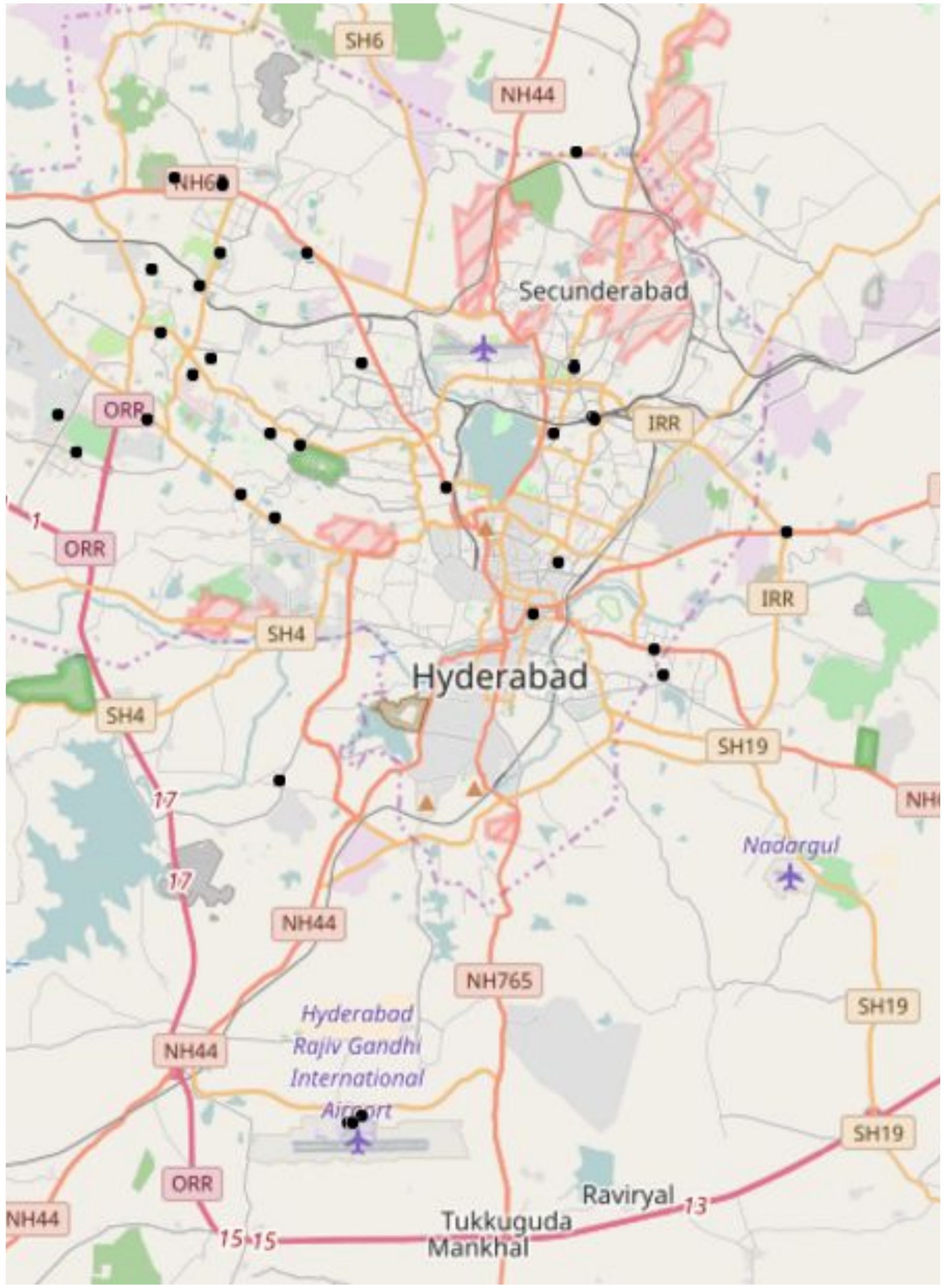


Distribution of ride requests at 2pm in Paris

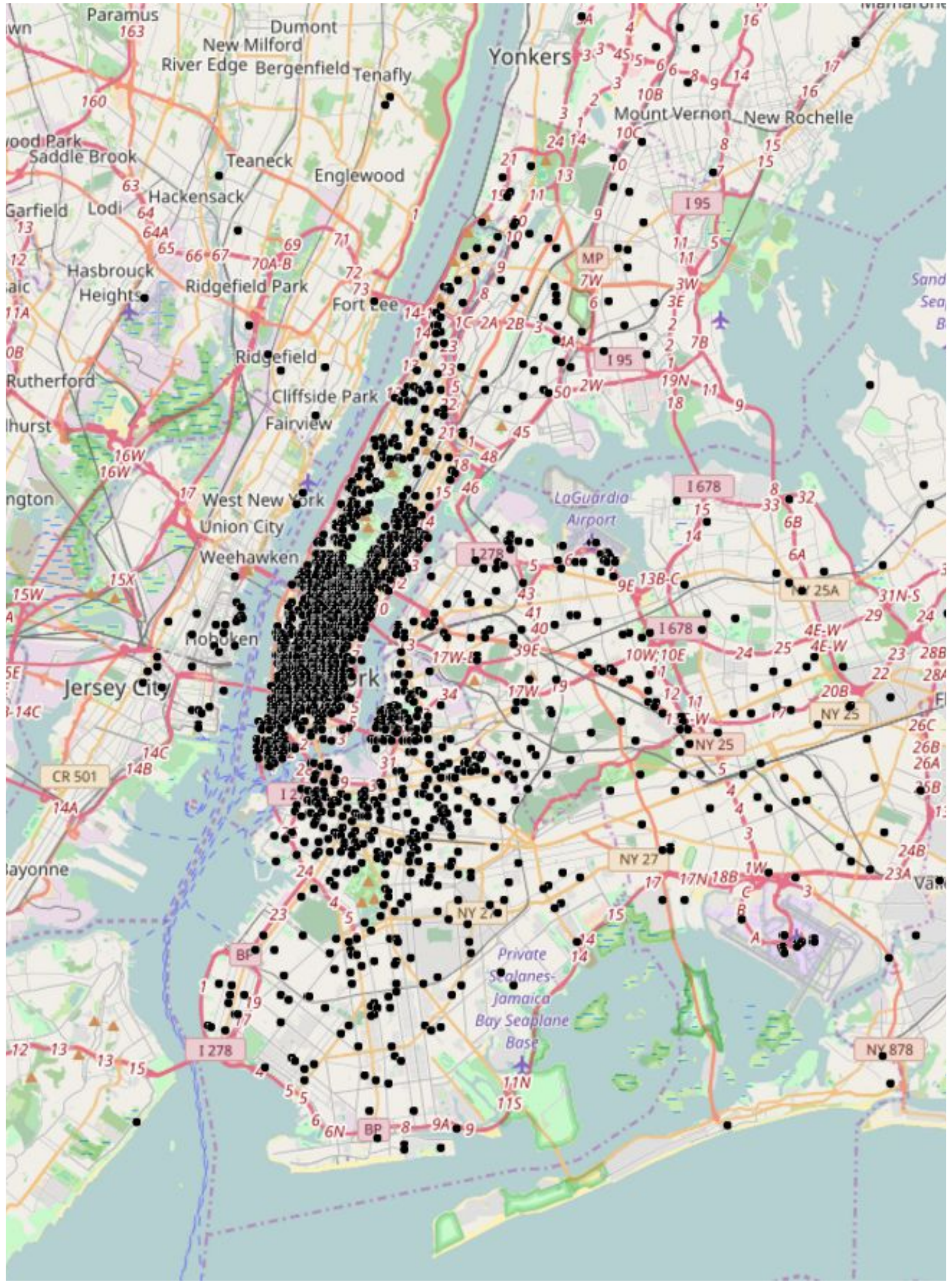
Distribution of ride requests at 4am in Paris



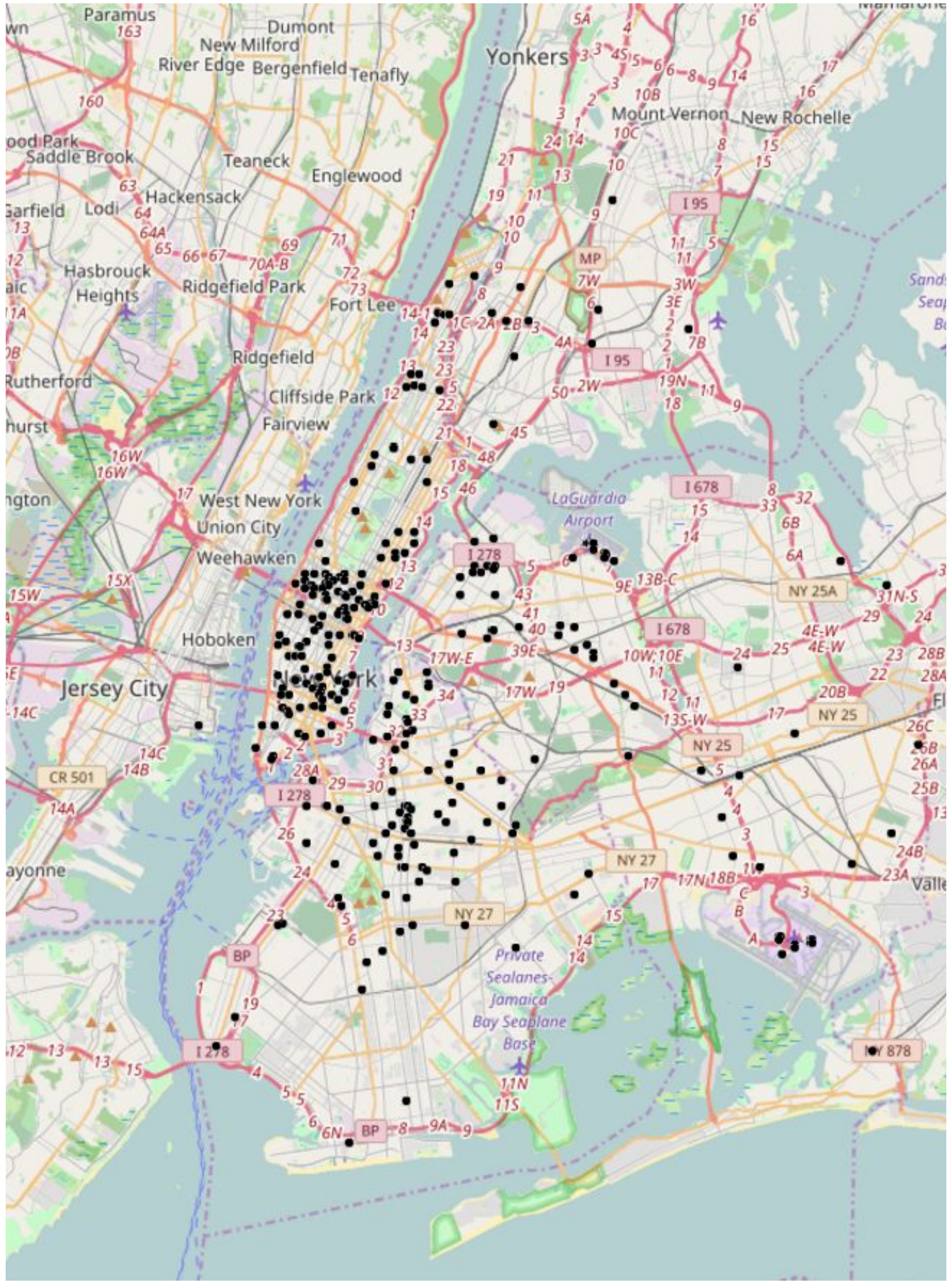
Distribution of ride requests at 11am in Hyderabad



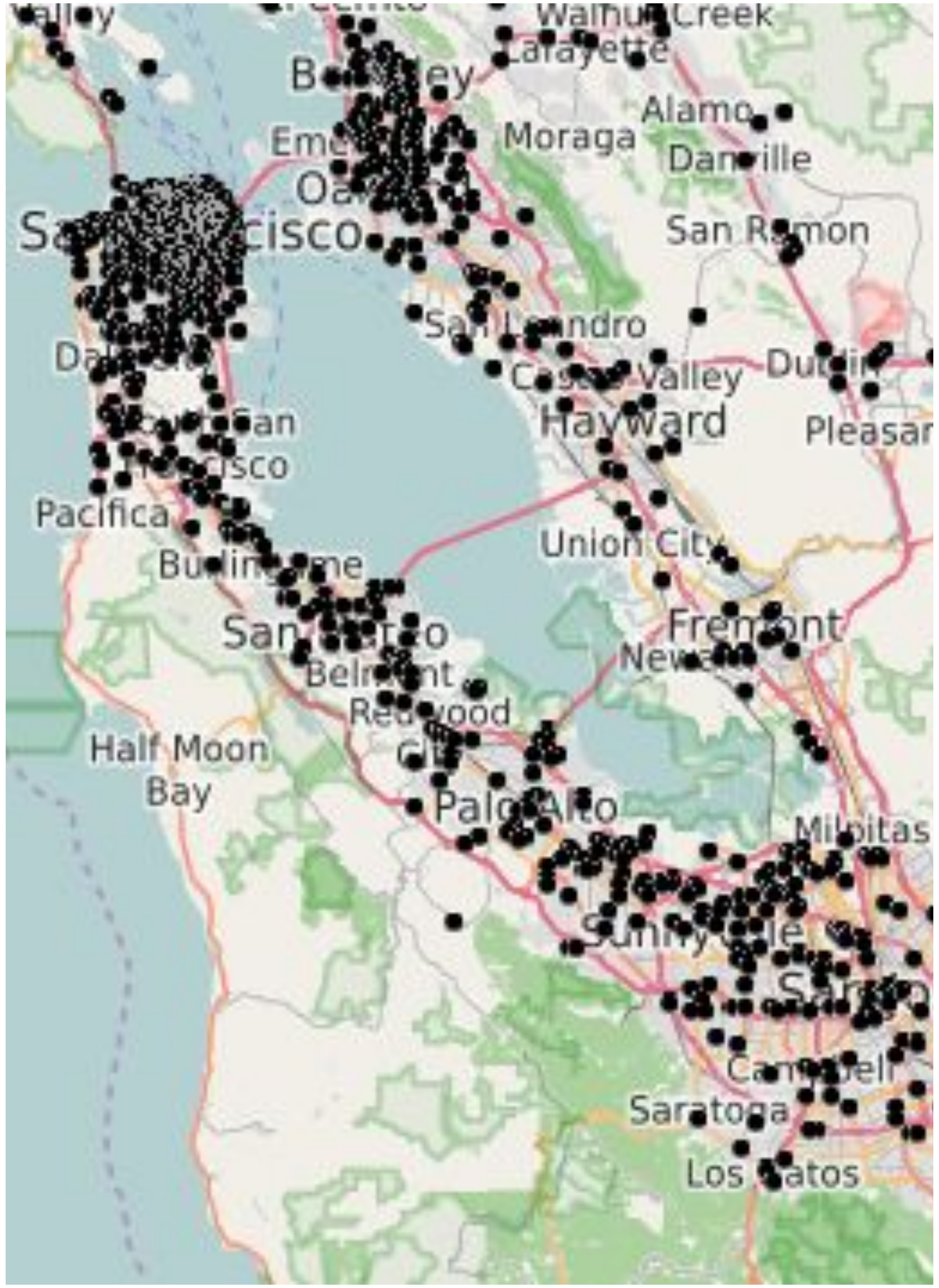
Distribution of ride requests at 5am in Hyderabad



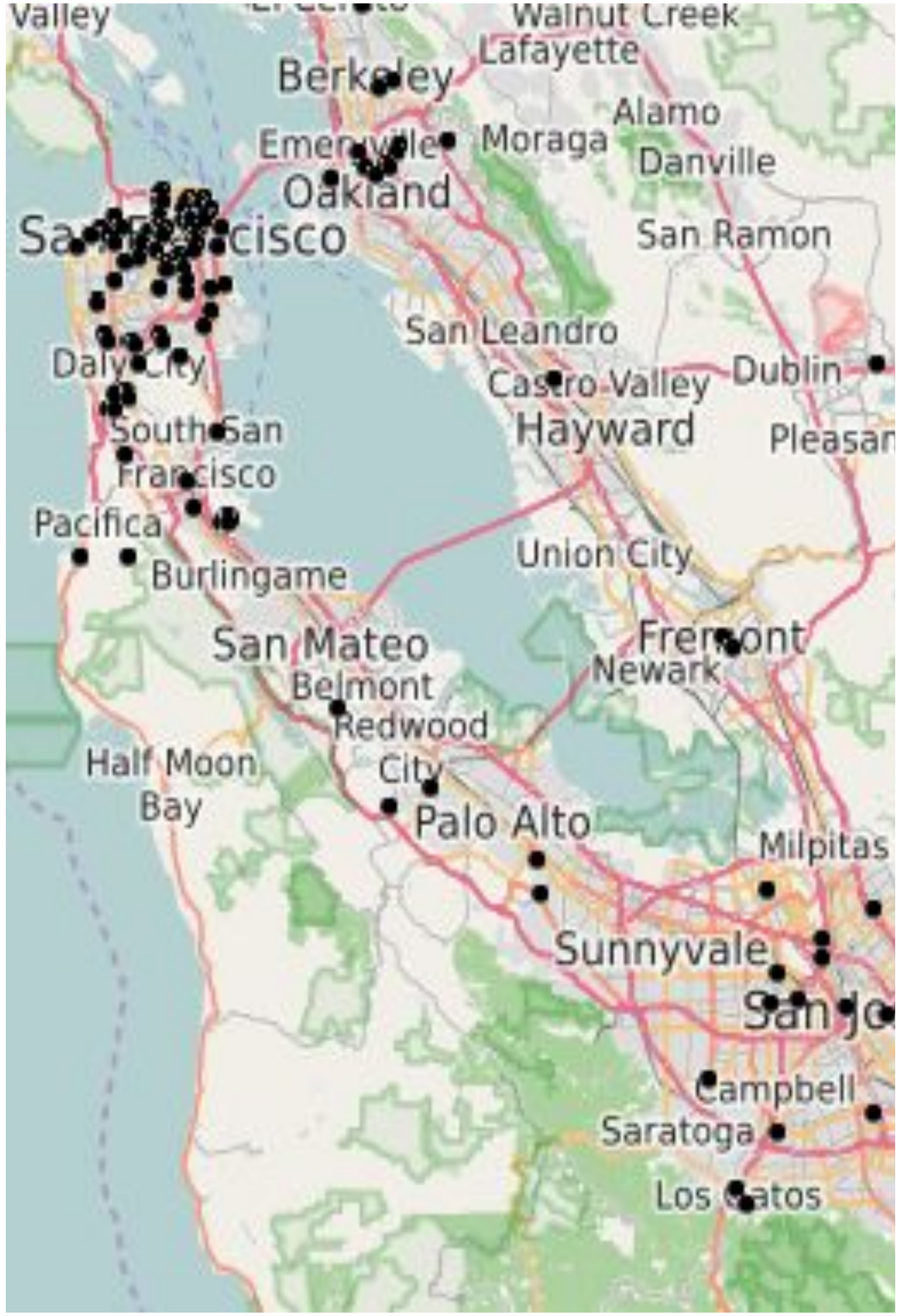
Distribution of ride requests at 8pm in NY



Distribution of ride requests at 3am in NY



Distribution of ride requests at 7pm in SF



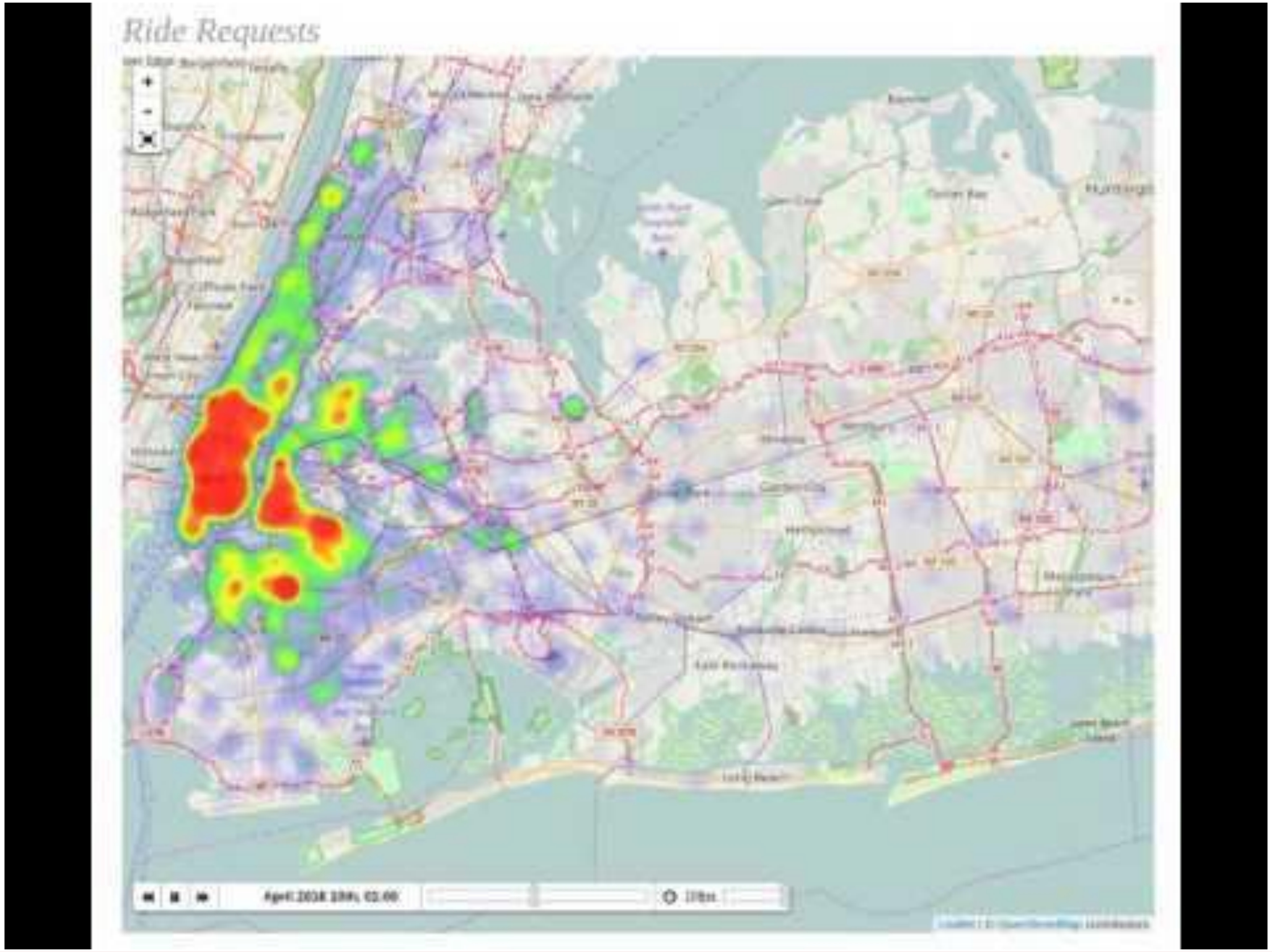
Distribution of ride requests at 5am in SF

Average Ride Requests per Hour

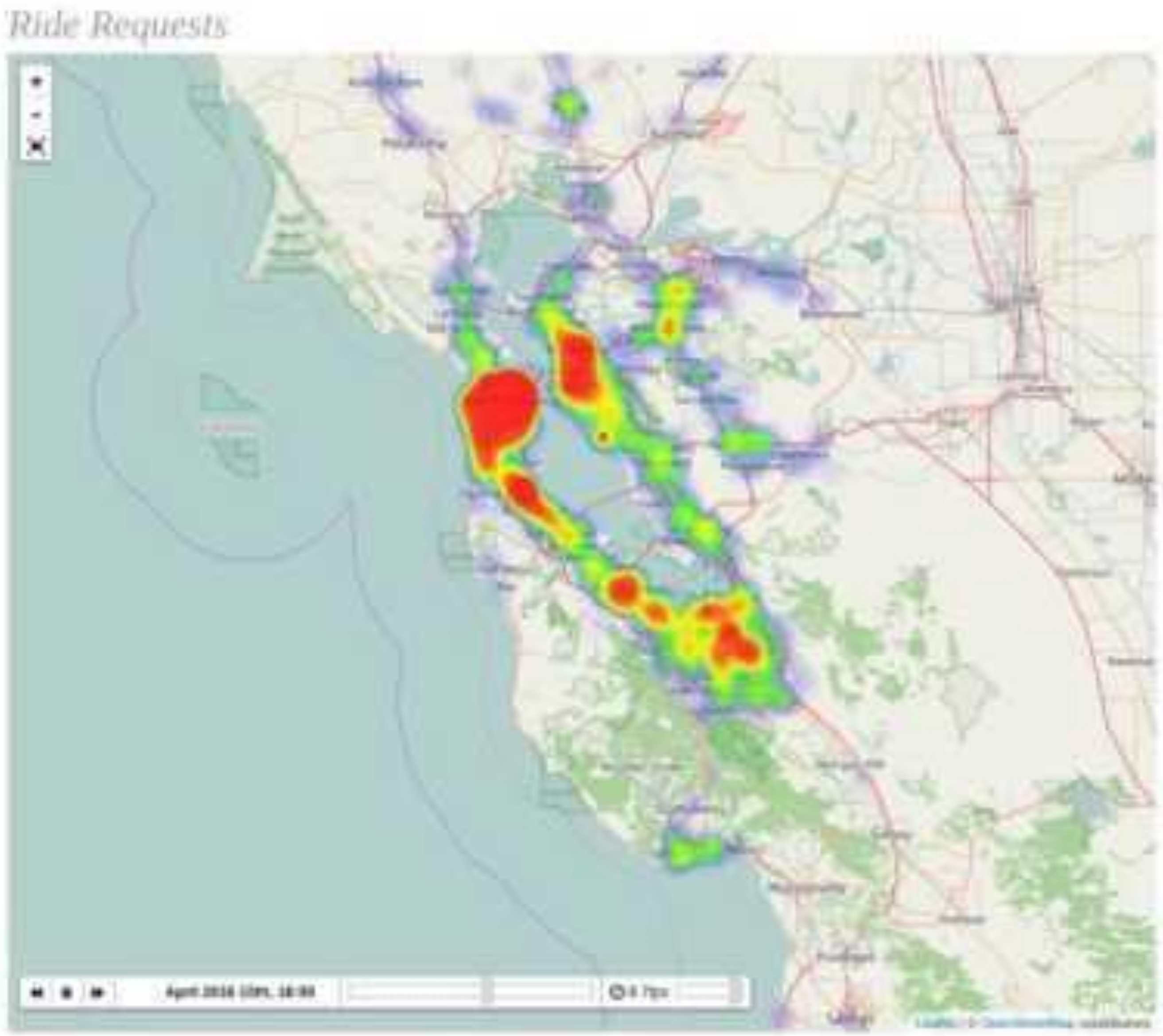
City	Hourly Average
Paris	3.5K
Hyderabad	1.5K
New York	7.7K
San Francisco	6.4K

Recent emergence of ride-sharing services is transforming human mobility and transportation in major cities of the world (Buzzfeed 2016). In December 2015, Uber Technologies, Inc. reported completion of a billion rides (Fortune 2015) within five years since it started operations. Didi alone in China reported 1.4 billion ride requests in 2015 (Wired 2016). They are expecting to reach 6 billion ride requests in 2016. There is huge potential for such services to transform urban transportation and automotive industries.

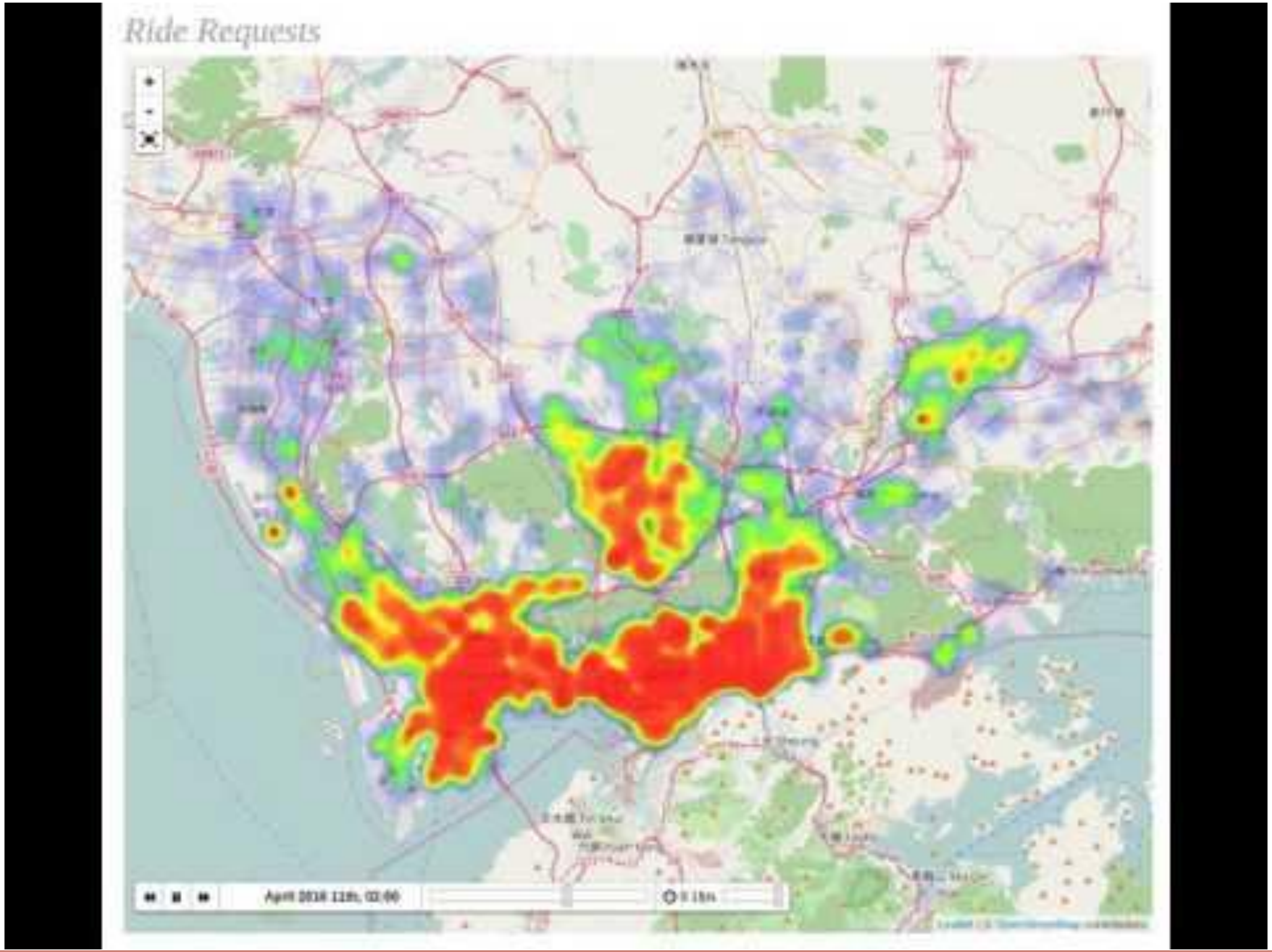
Ride Requests in New York



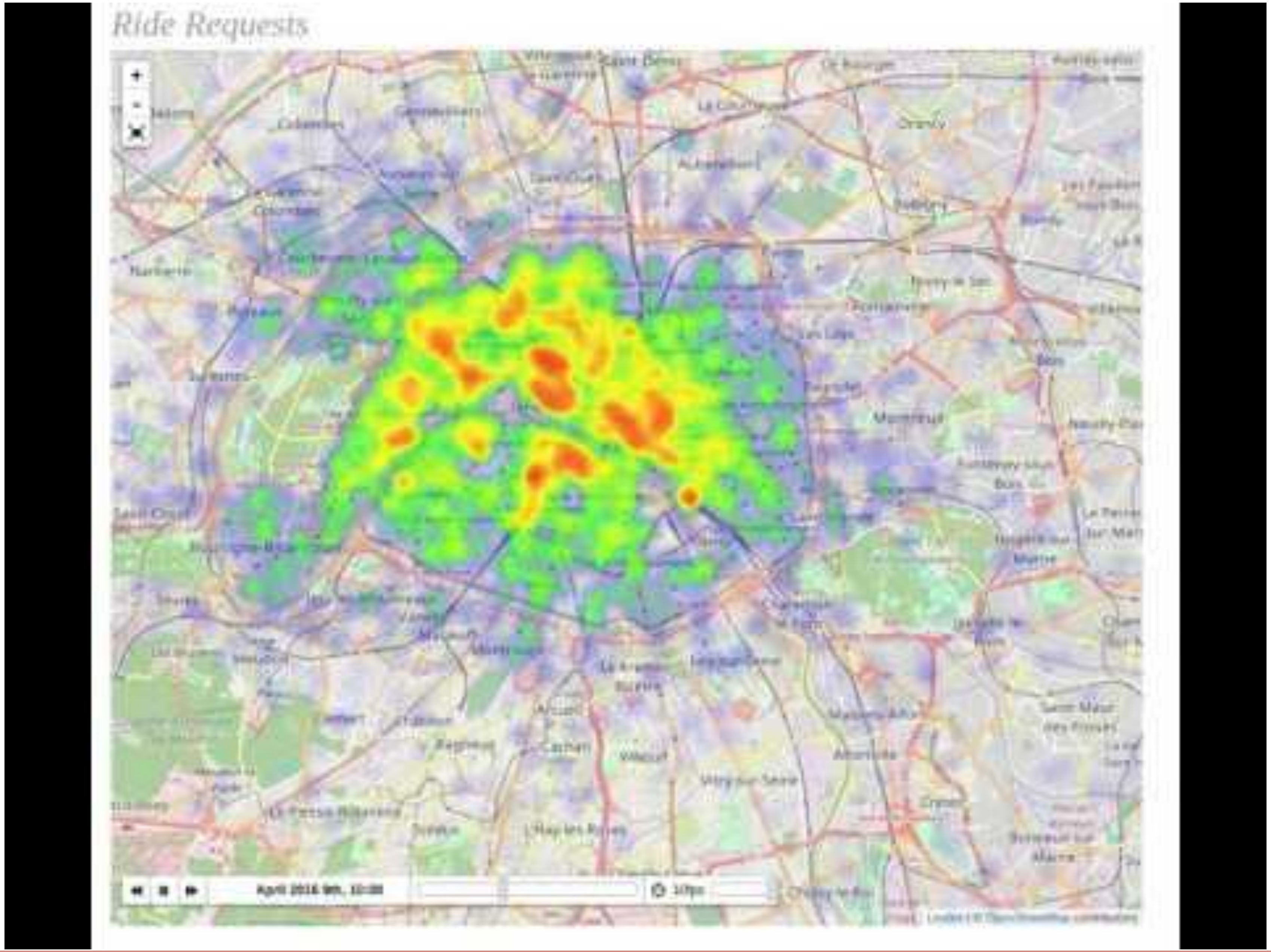
Ride Requests in San Francisco



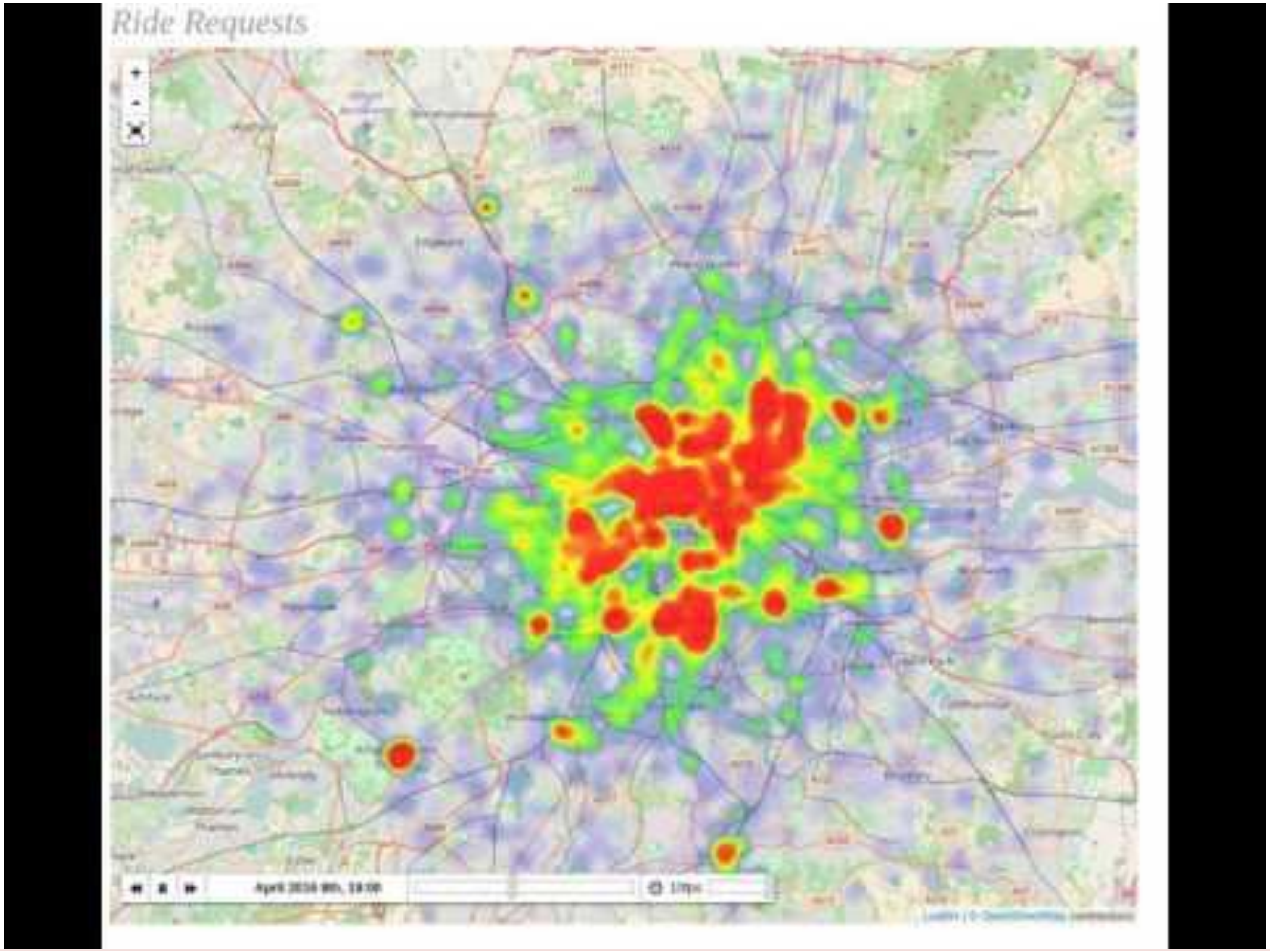
Ride Requests in Shenzhen



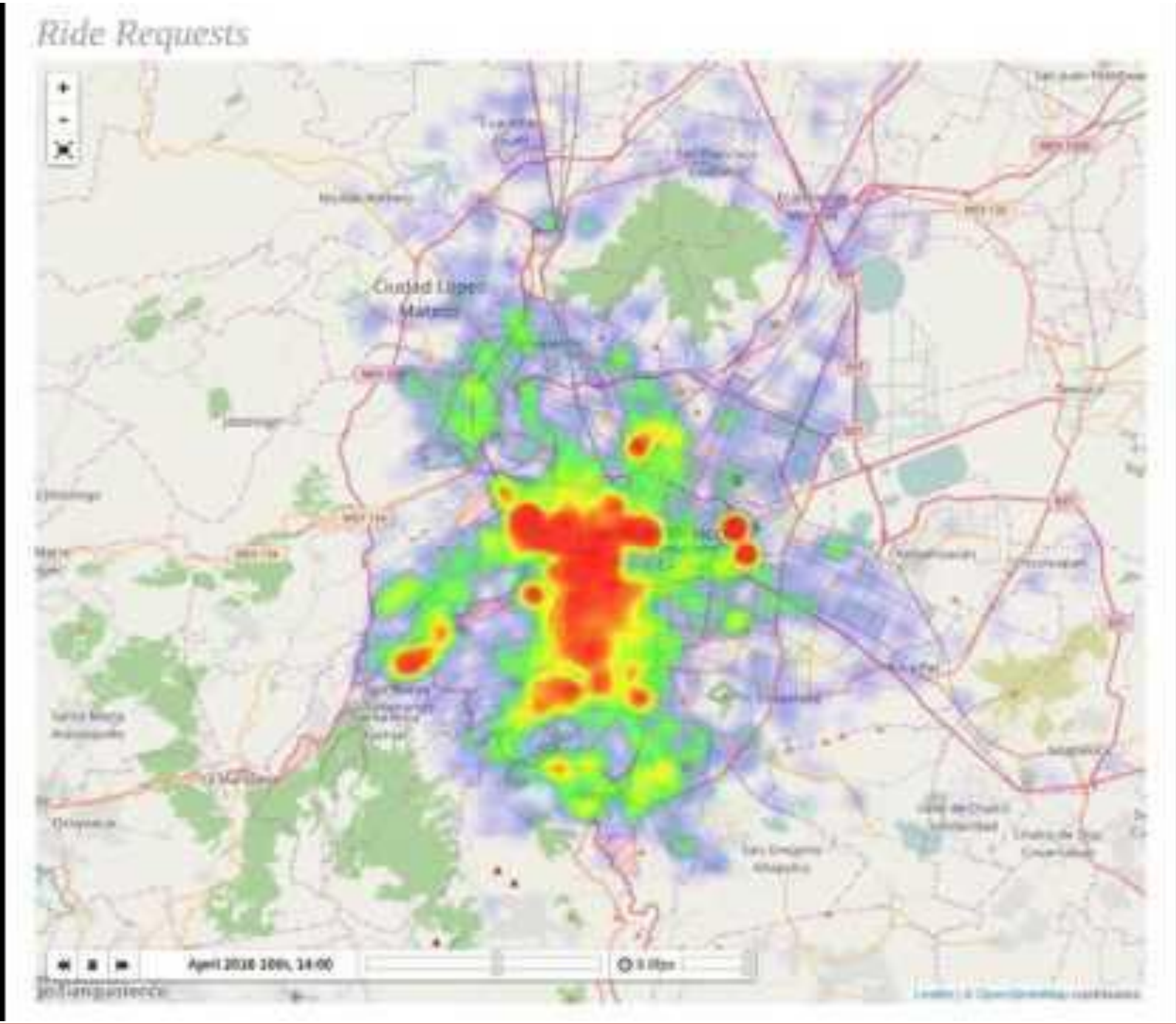
Ride Requests in Paris



Ride Requests in London

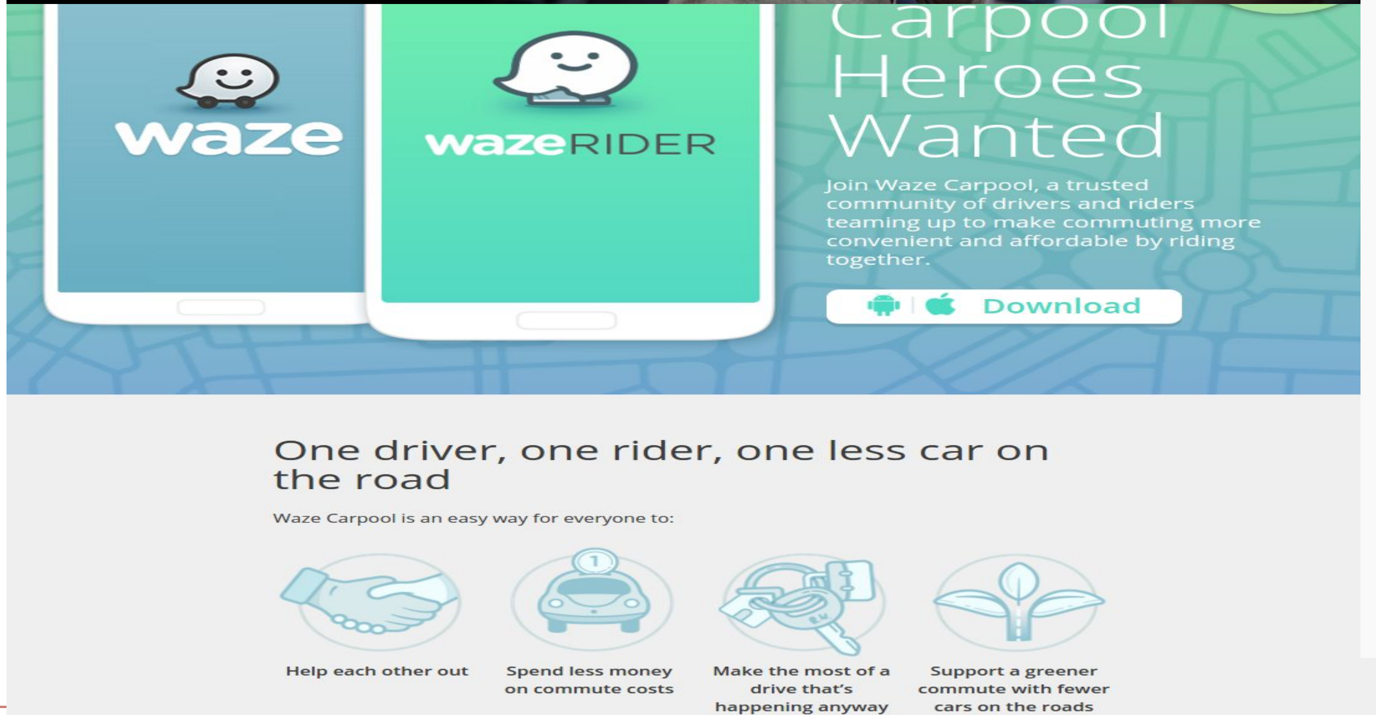


Ride Requests in Mexico City



Observation 1: There is significant variability in the **ride request patterns** from city to city, and across space and time within each city.

All major ride-sharing services are introducing **“ride pooling”** to reduce number of vehicles needed on the roads and to increase the overall efficiency of ride-sharing services.



Didi Kuaidi, China’s Dominant Taxi App Firm, Launches Carpooling Service

Posted Jun 1, 2015 by Jon Russell (@jonrussell)

CrunchBase

Didi Chuxing

FOUNDED
2012

OVERVIEW
Founded in 2012, DiDi is the world's largest mobile transportation service platform, offering a broad range of mobile technology-based transportation options across over 400 major Chinese cities, including taxi hailing, private car hailing, Hitch (social ride-sharing), Chauffeur, DiDi Bus, DiDi Test Drive, and DiDi Enterprise Solutions. As the leader in China's sharing economy initiative, DiDi ...

LOCATION
Beijing, 22

CATEGORIES
Public Transportation, Mobile Apps, Transportation

FOUNDERS
Cheng Wei

WEBSITE
http://www.xiaojukeji.com

Full profile for Didi Chuxing

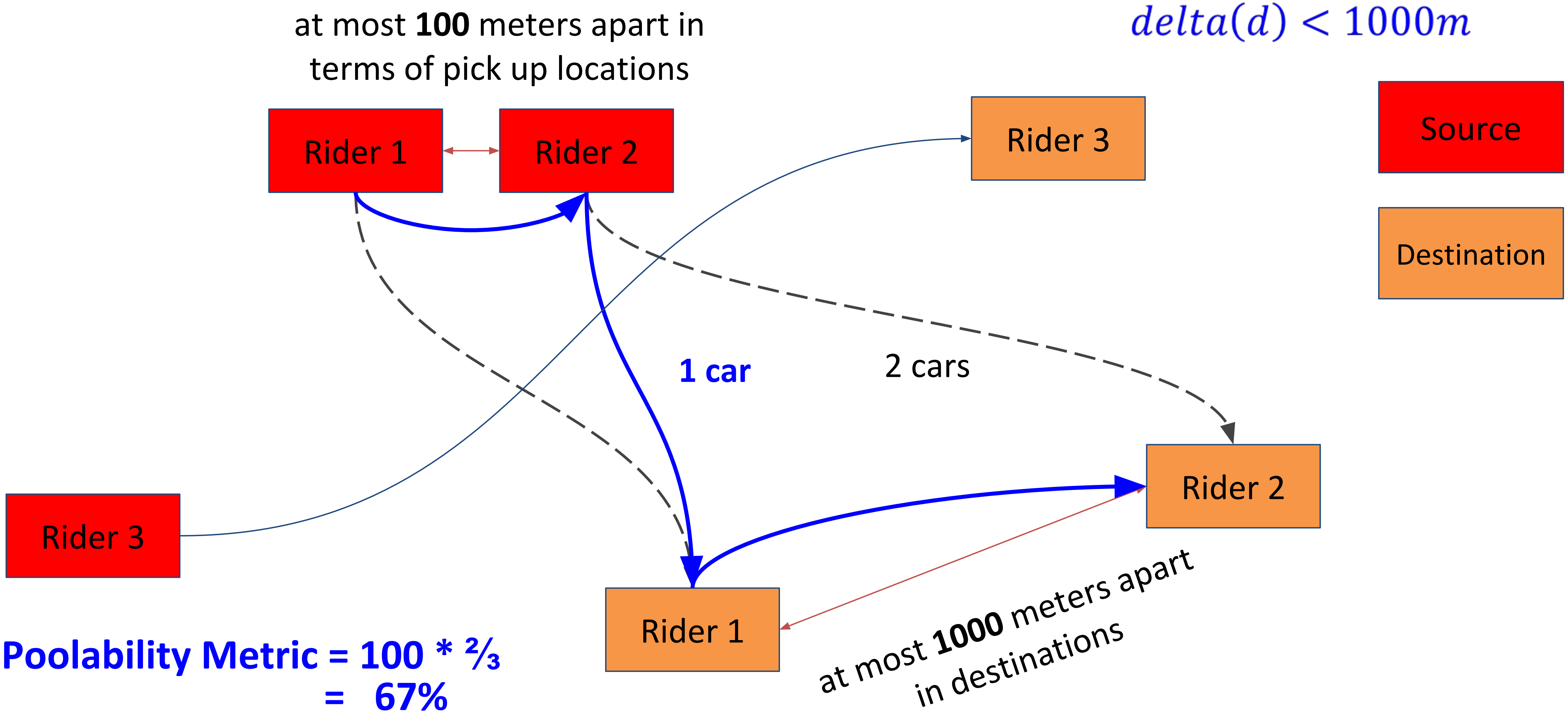
Uber launched a nonprofit car-pooling service in China called People’s Uber last year, and now its biggest rival — and China’s largest ride-sharing service — has followed suit with a

What is “Ride Poolability”?

Each Ride Request: <*t,s,d*>

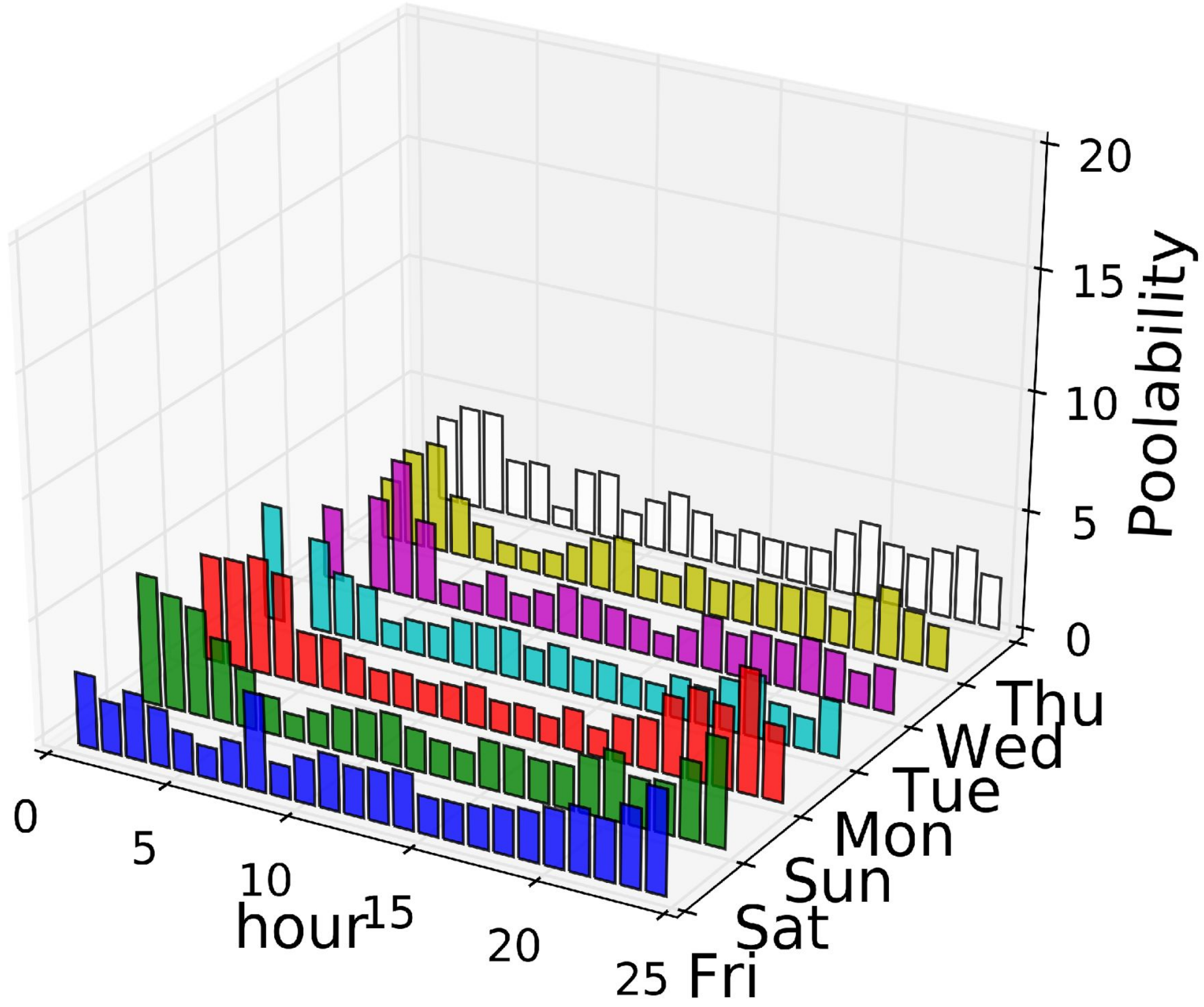
Proximity Constraints for Pooling:

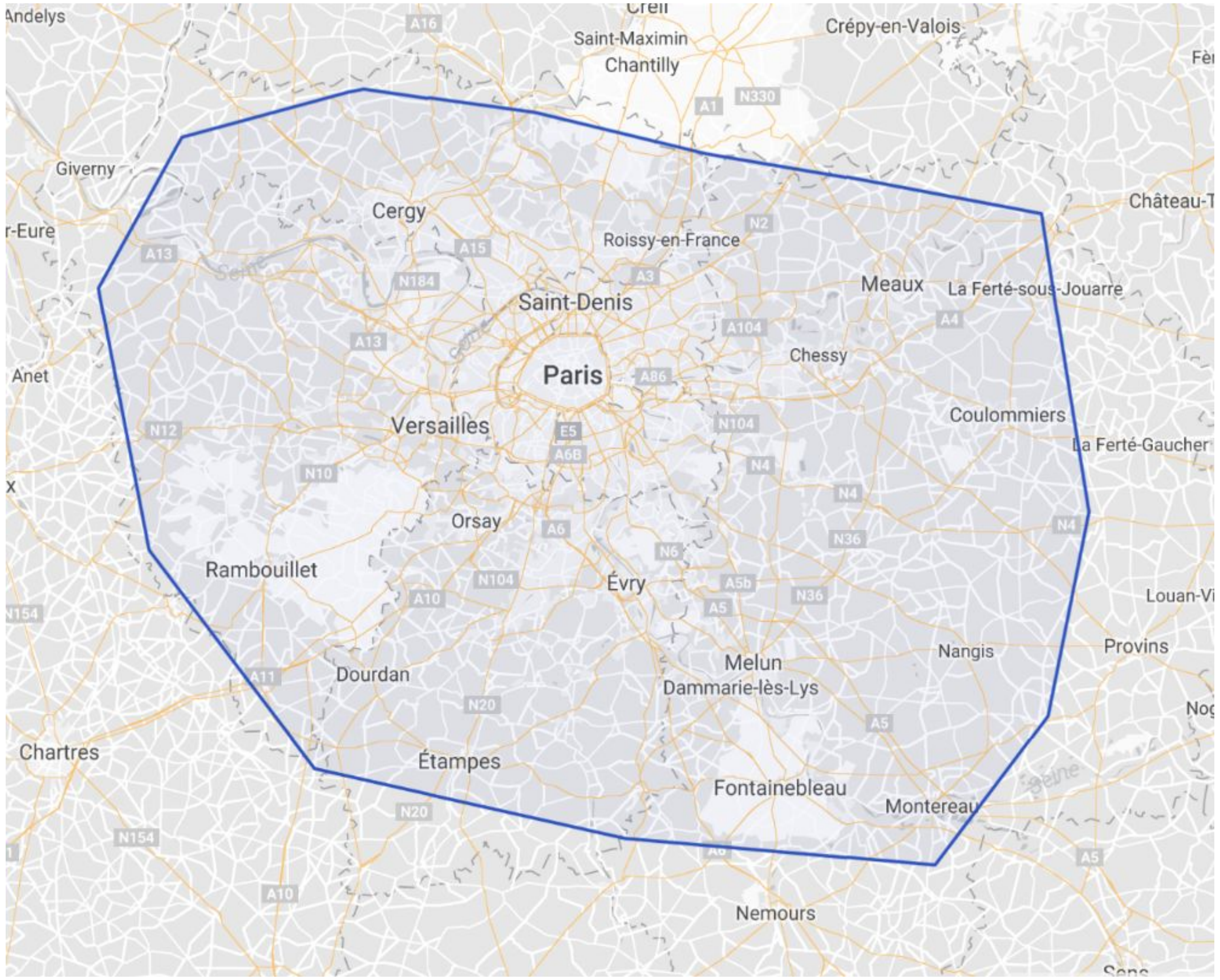
$$delta(t) < 5min$$
$$delta(s) < 100m$$
$$delta(d) < 1000m$$



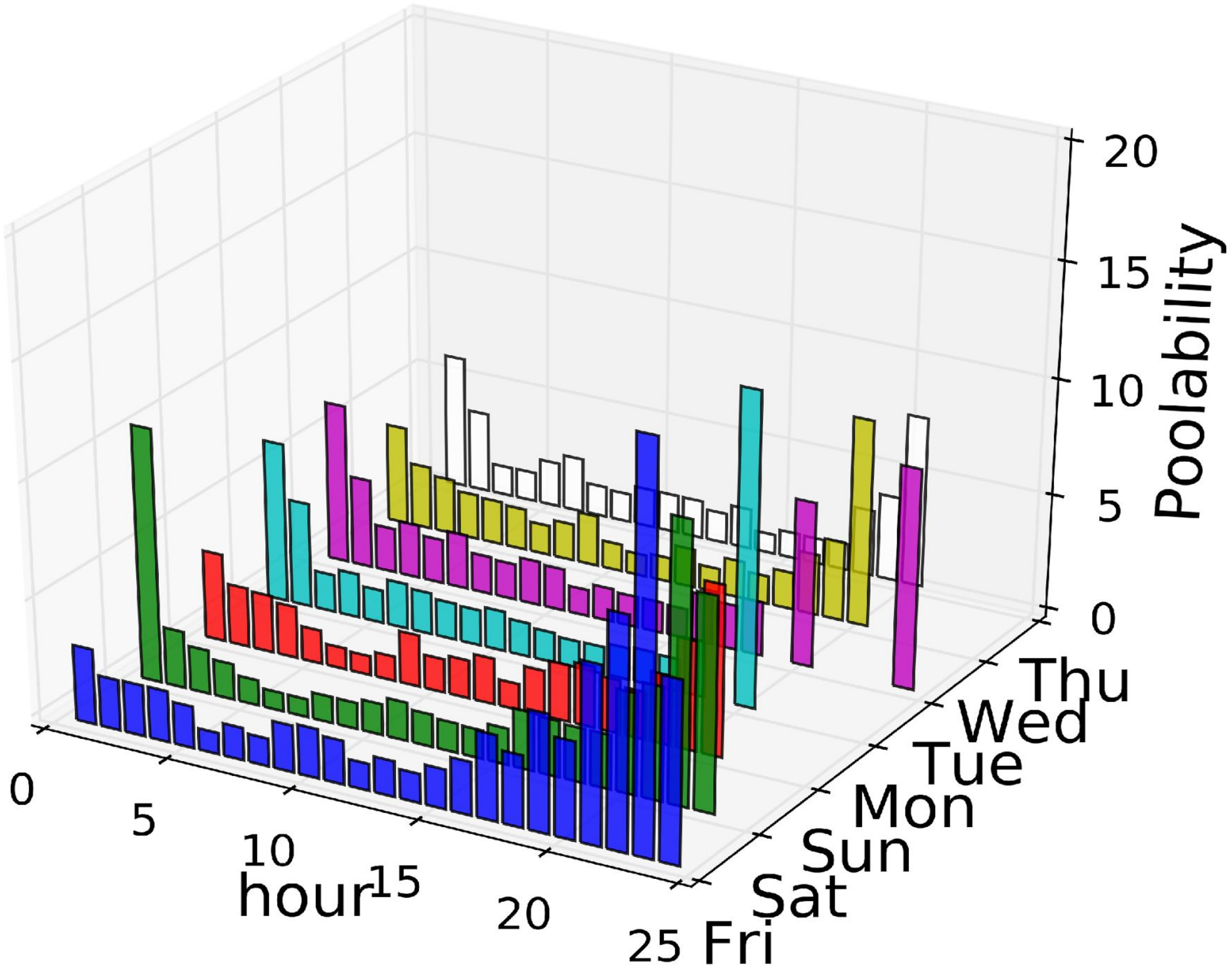
Poolability Metric = 100 * ⅔
= 67%

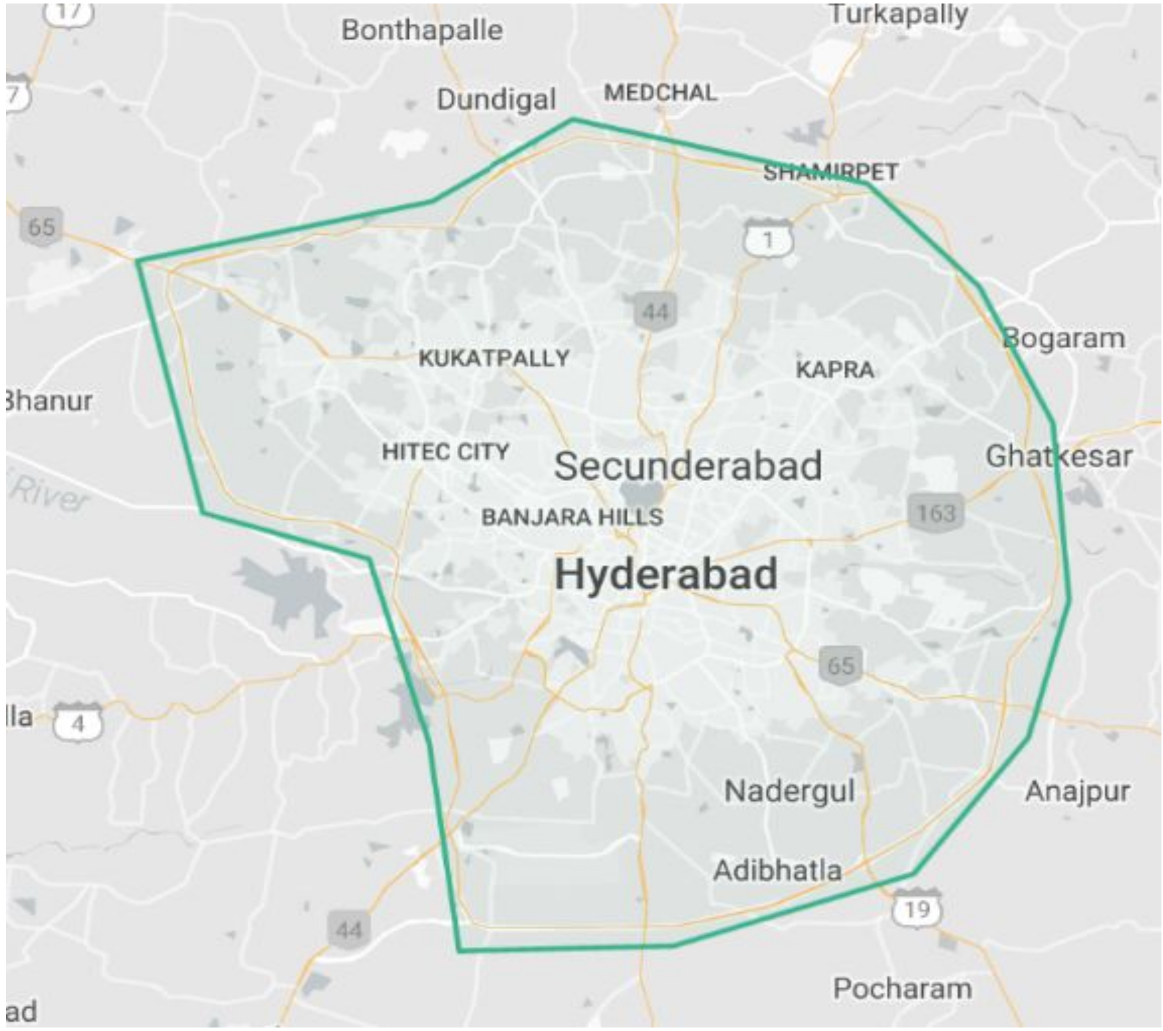
Ride Poolability Profile for Paris



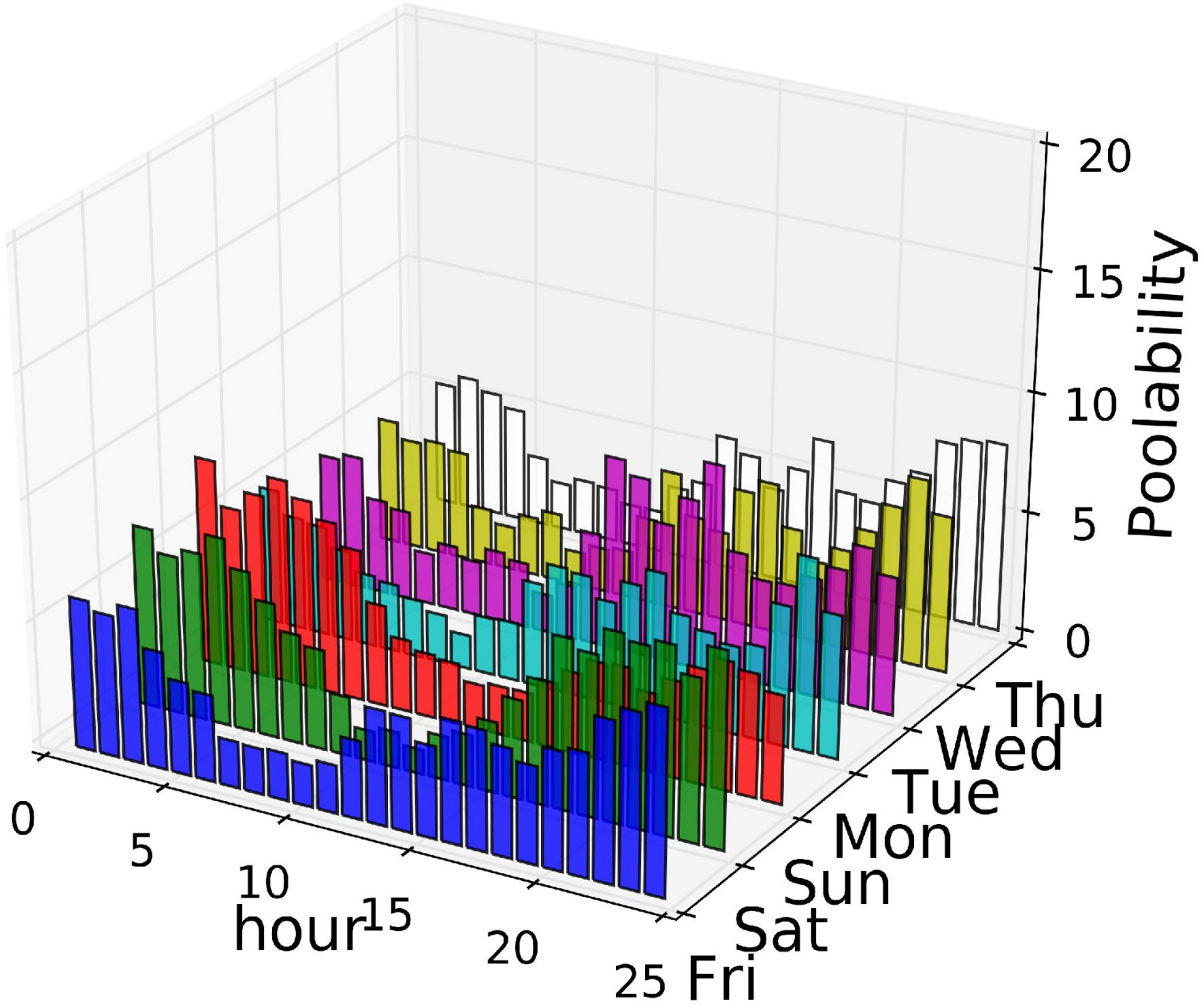


Ride Poolability Profile for Hyderabad

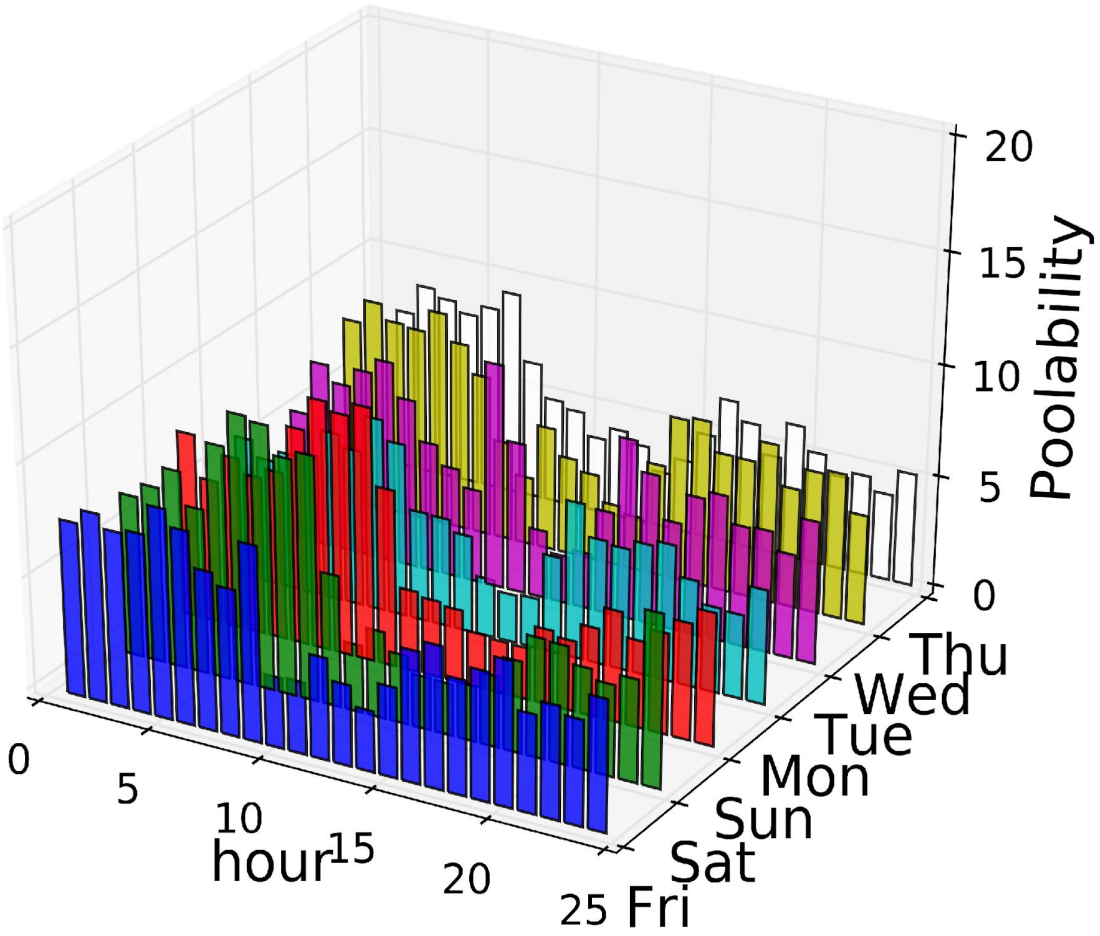


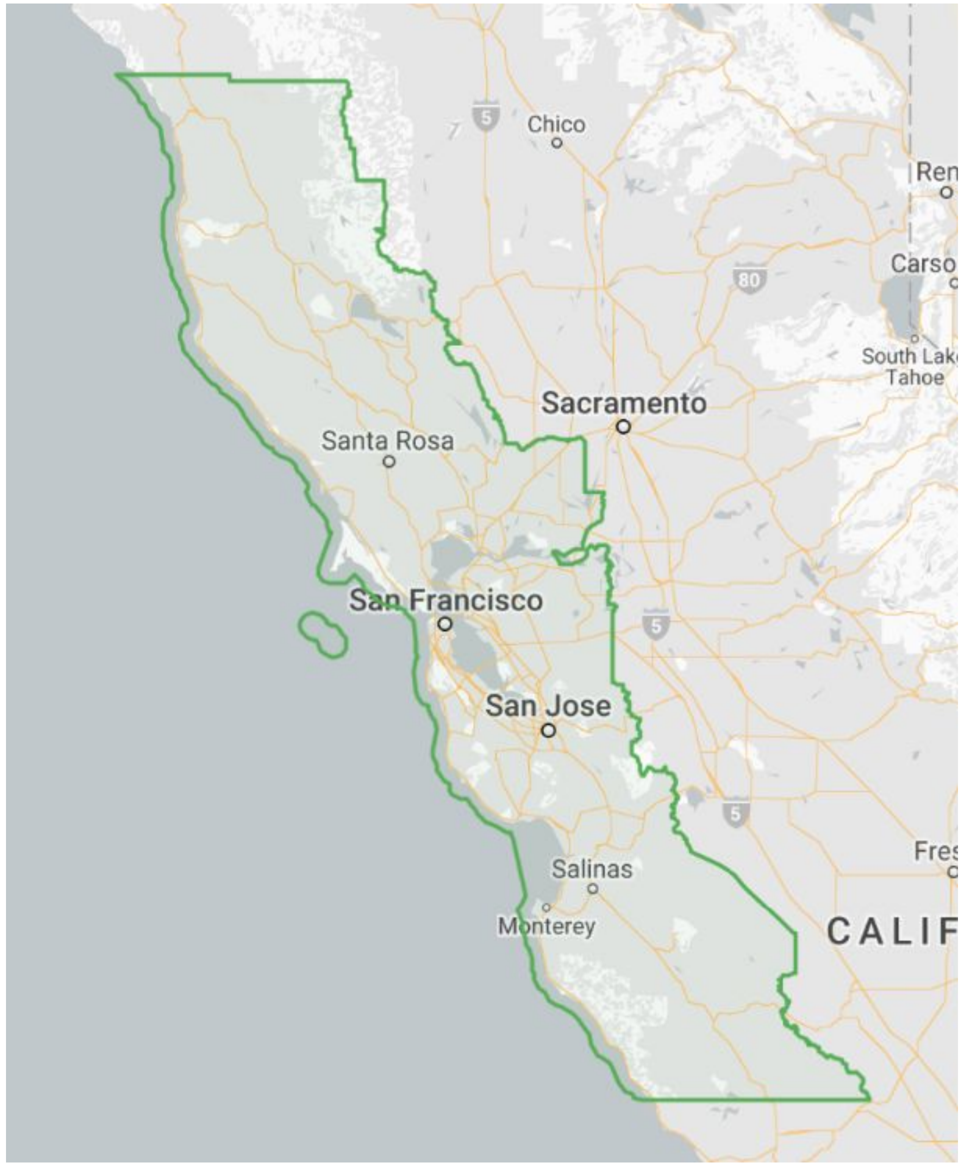


Ride Poolability Profile for New York

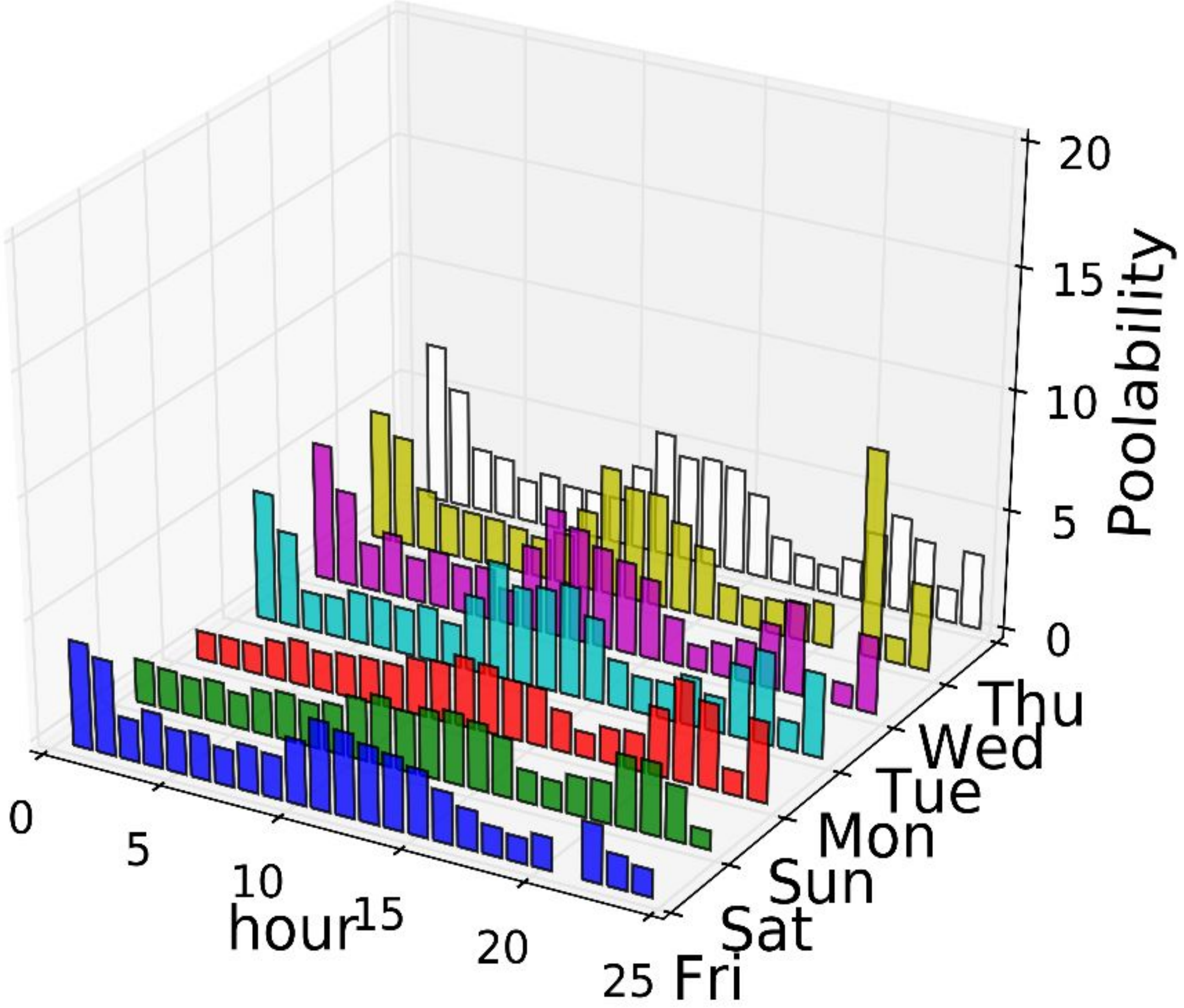


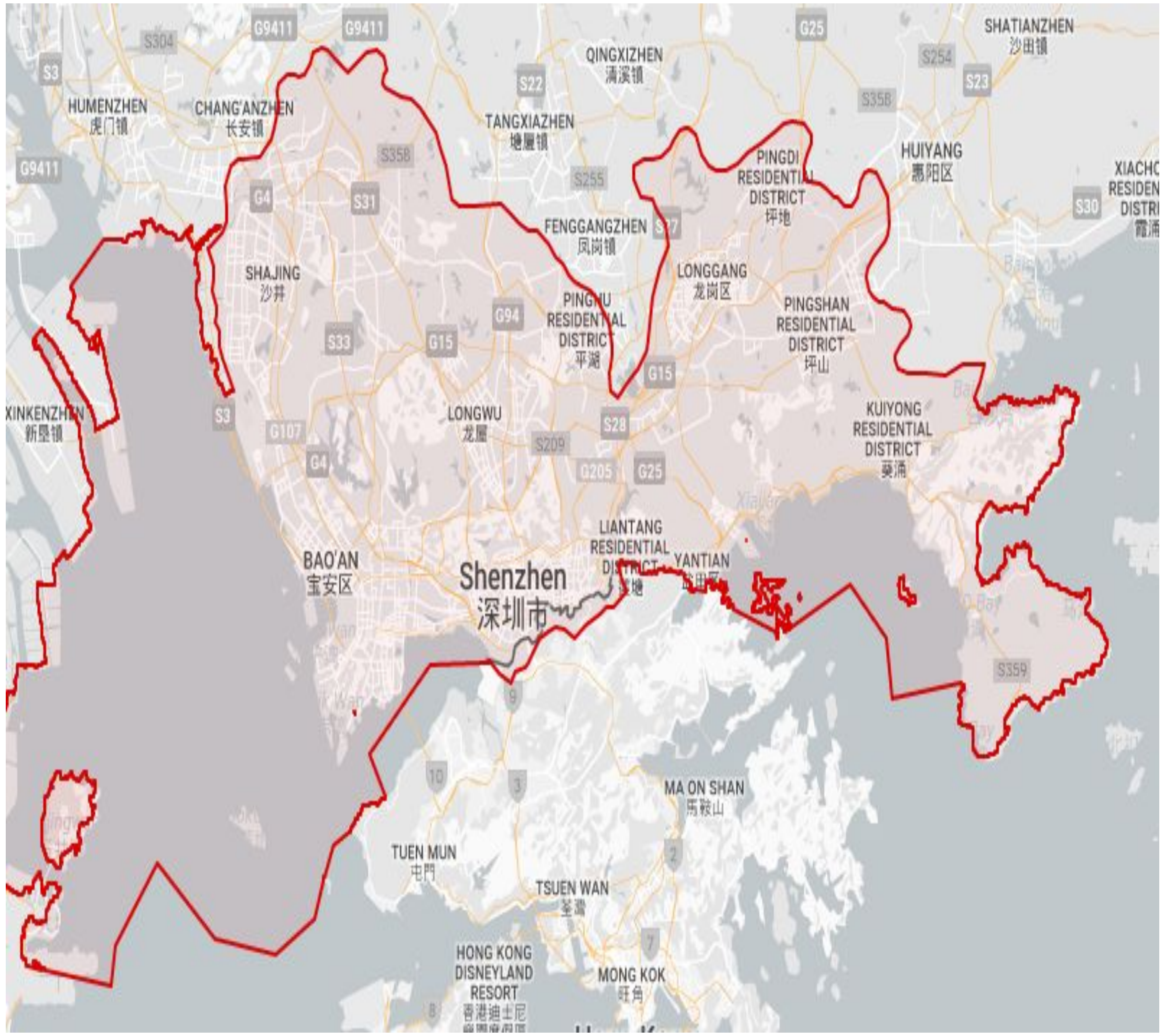
Ride Poolability Profile for San Francisco



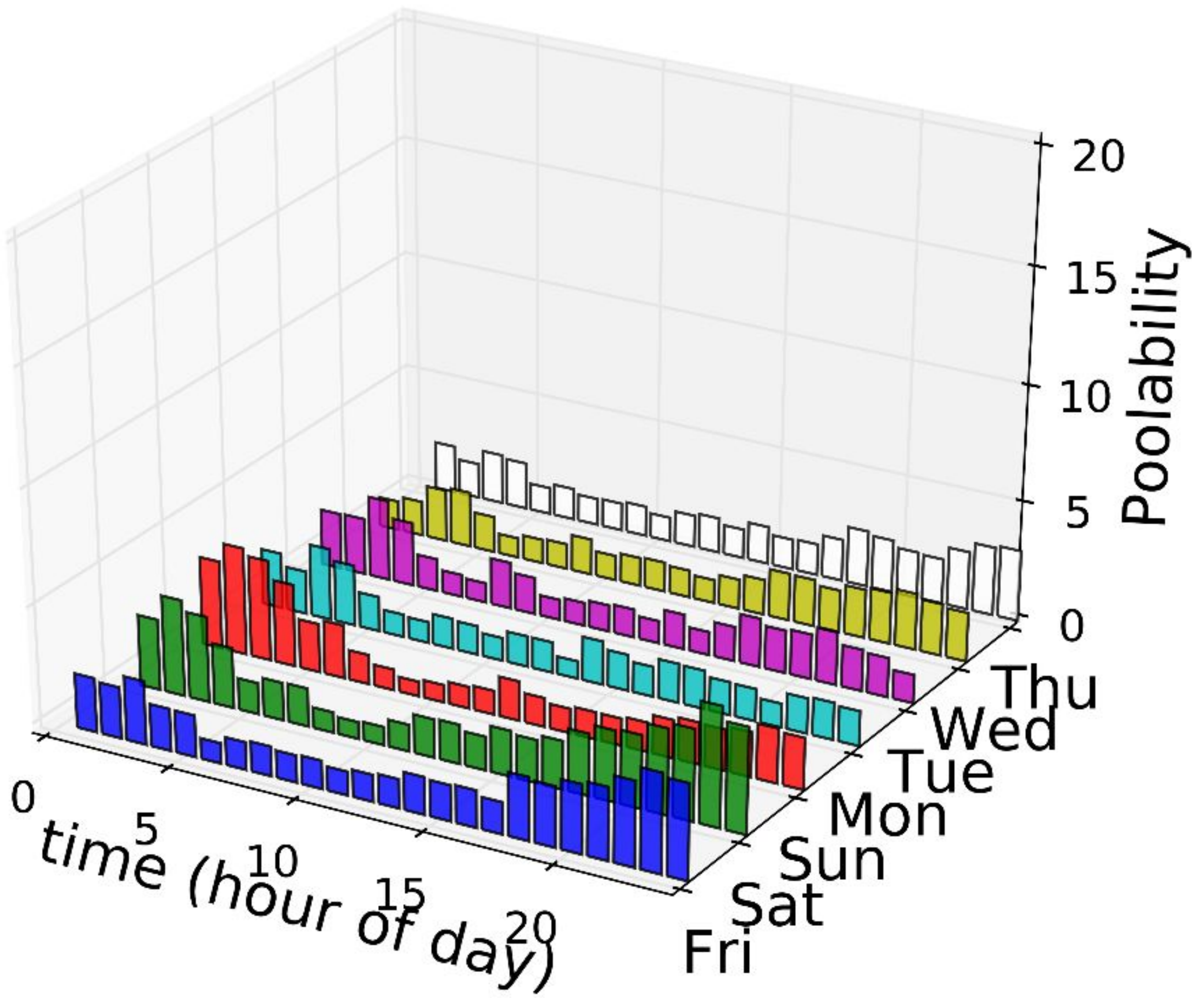


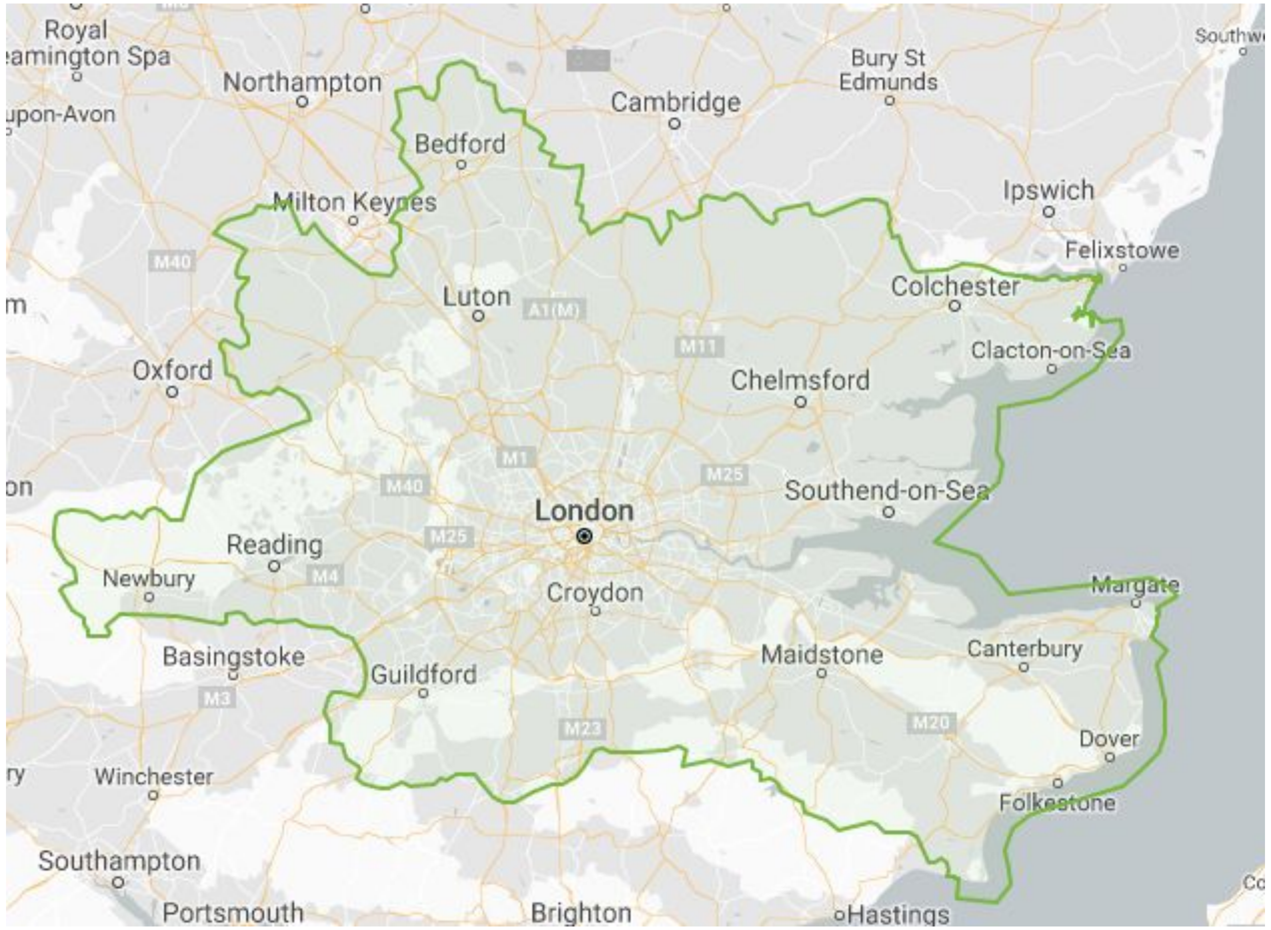
Ride Poolability Profile for Shenzhen



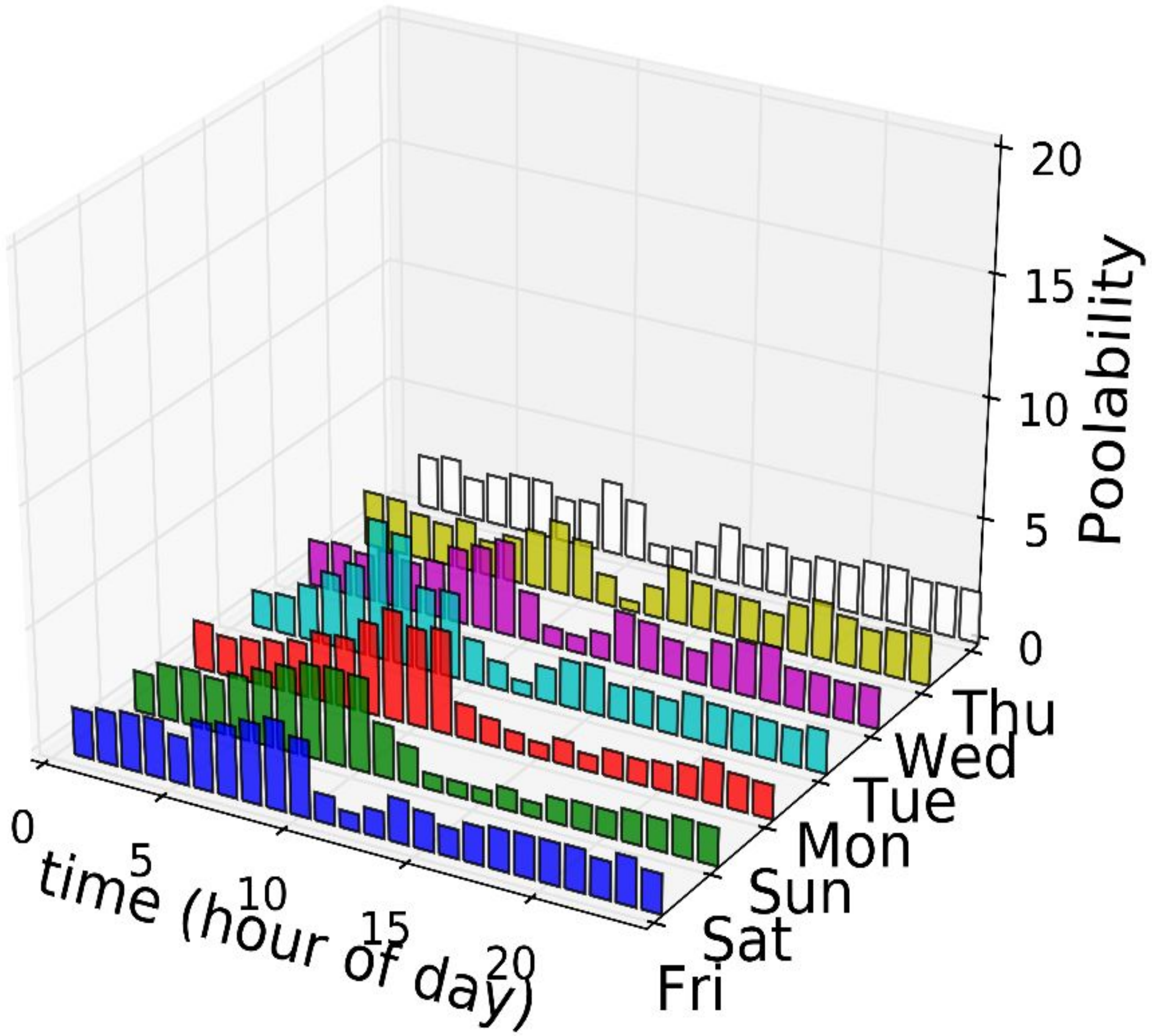


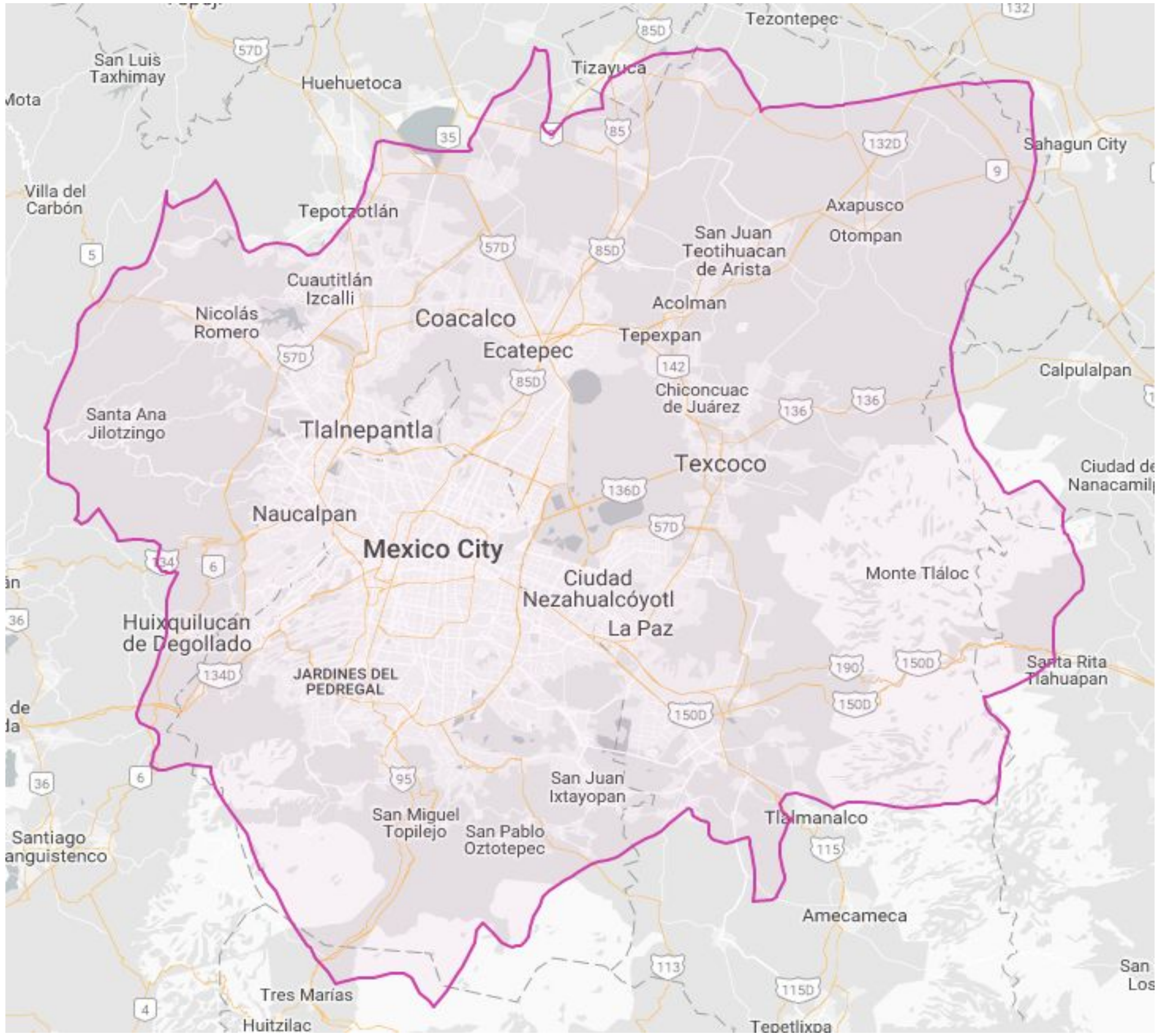
Ride Poolability Profile for London



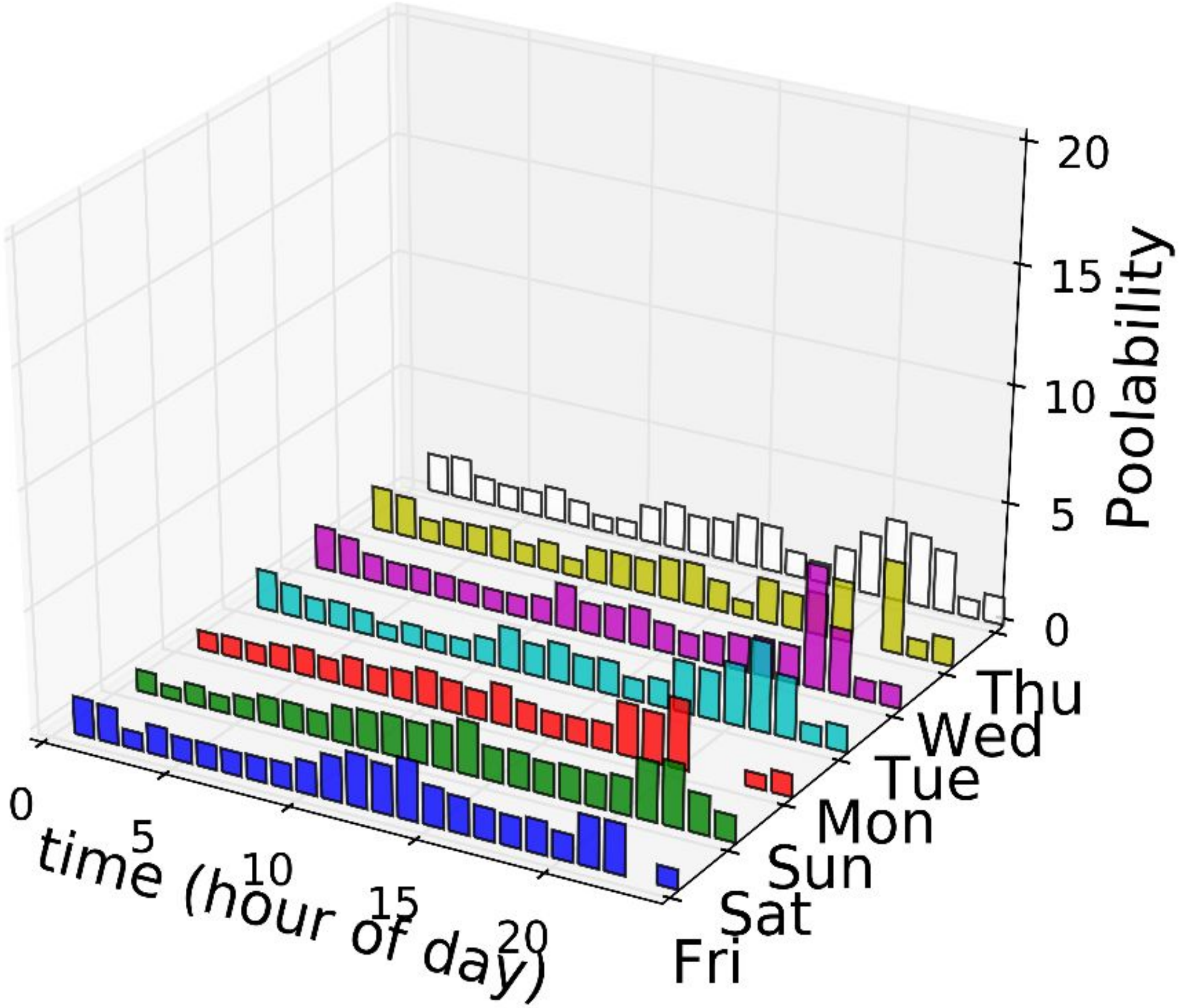


Ride Poolability Profile for Mexico City





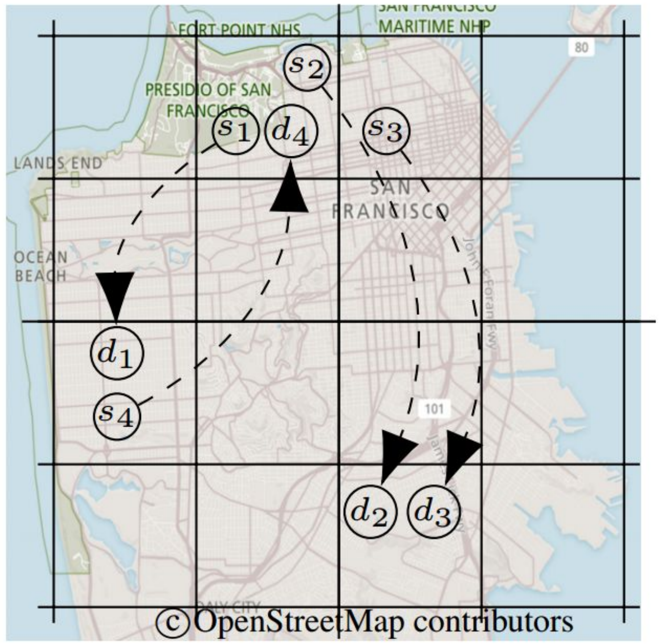
Ride Poolability Profile for Shanghai



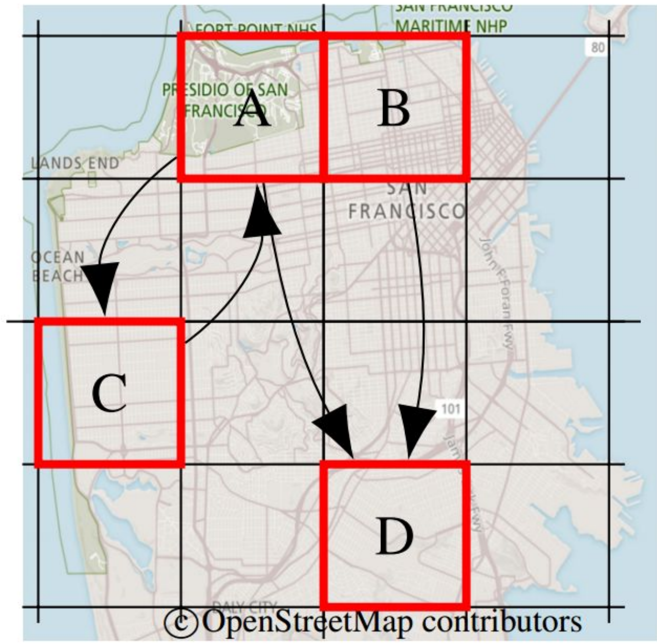
Observation 2: There is similar variability in the **ride poolability profiles** from city to city, and across space and time within each city.

Question 1: Is there a rigorous model that can capture both the spatial and temporal variations of **ride request patterns** in a city?

Ride Request Graph (RRG)



(a) Four ride requests distributed spatially over a map

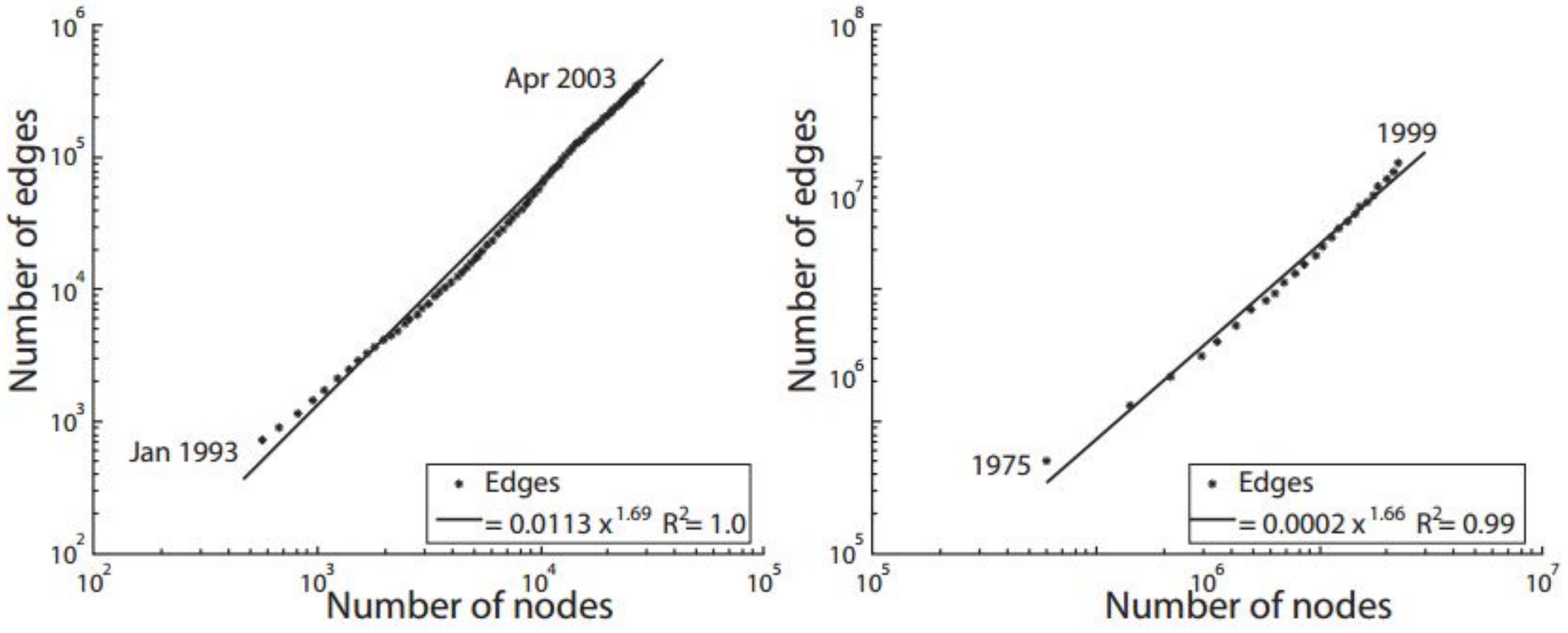


(b) Corresponding Ride Request Graph with four nodes (marked by red boxes) and directed edges.

Multiple instances of RRG can be created at every time instance, let's say every 5 minutes.

Densification Power Law (DPL)

Time-evolving graph like Arxiv citation graph, the Patent citation graph, and many others share a common property i.e. DPL.



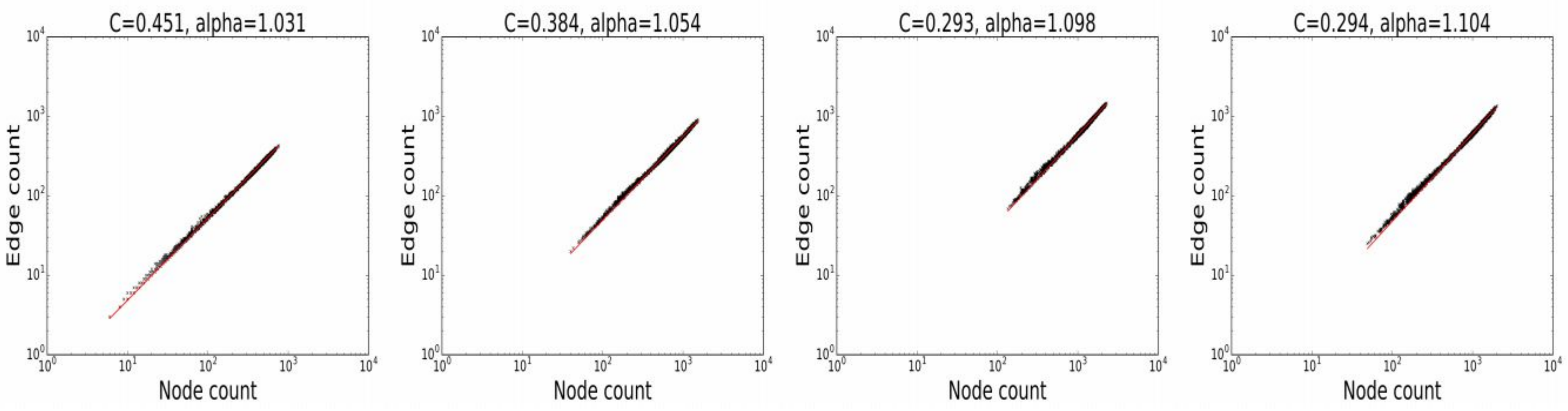
(a) arXiv (b) Patents

Image source: Chakrabarti, D., & Faloutsos, C. (2006). Graph mining: Laws, generators, and algorithms. *ACM computing surveys (CSUR)*, *38*(1), 2.

Densification Power Law property of RRG

Suppose, we have N(t) nodes, and E(t) edges at time t, and if at t+1, N(t+1) = 2*N(t), what will be E(t+1)?

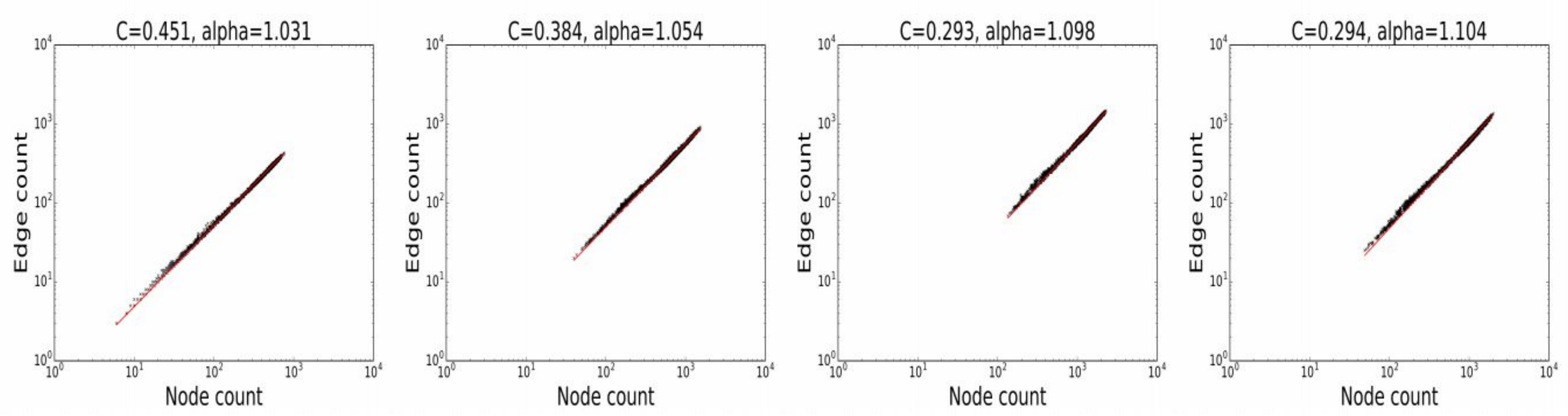
$$E(t) = C * N(t)^{\alpha}$$



Hyderabad Paris New York San Francisco

Question 2: What is the correlation, if any, between the **ride request patterns** of a city and the **ride poolability profiles** of that city?

DPL and Ride Poolability



Alpha is correlated with ride poolability due to the **"human community"** effect.

City	*Mean*	*Minimum*	*Maximum*
Hyderabad	2.23	0.84	7.41
Paris	2.39	0.79	4.22
New York	4.48	1.70	7.84
San Francisco	5.48	2.50	9.16

Auto generation property of DPL graphs

- → Graphs exhibiting DPL can be automatically generated
- → Auto generation is based on a small number of parameters

Desiderata for auto generation of RRGs

- → Generated graphs should match graphs extracted from real data
- → Should exhibit the same **human community** effects
- → Use as few parameters as possible to do auto generation
- → Should be efficient and scalable

Question 3: Is it possible to auto generate RRGs that exhibit the same DPL metrics as those RRGs extracted from actual real data?

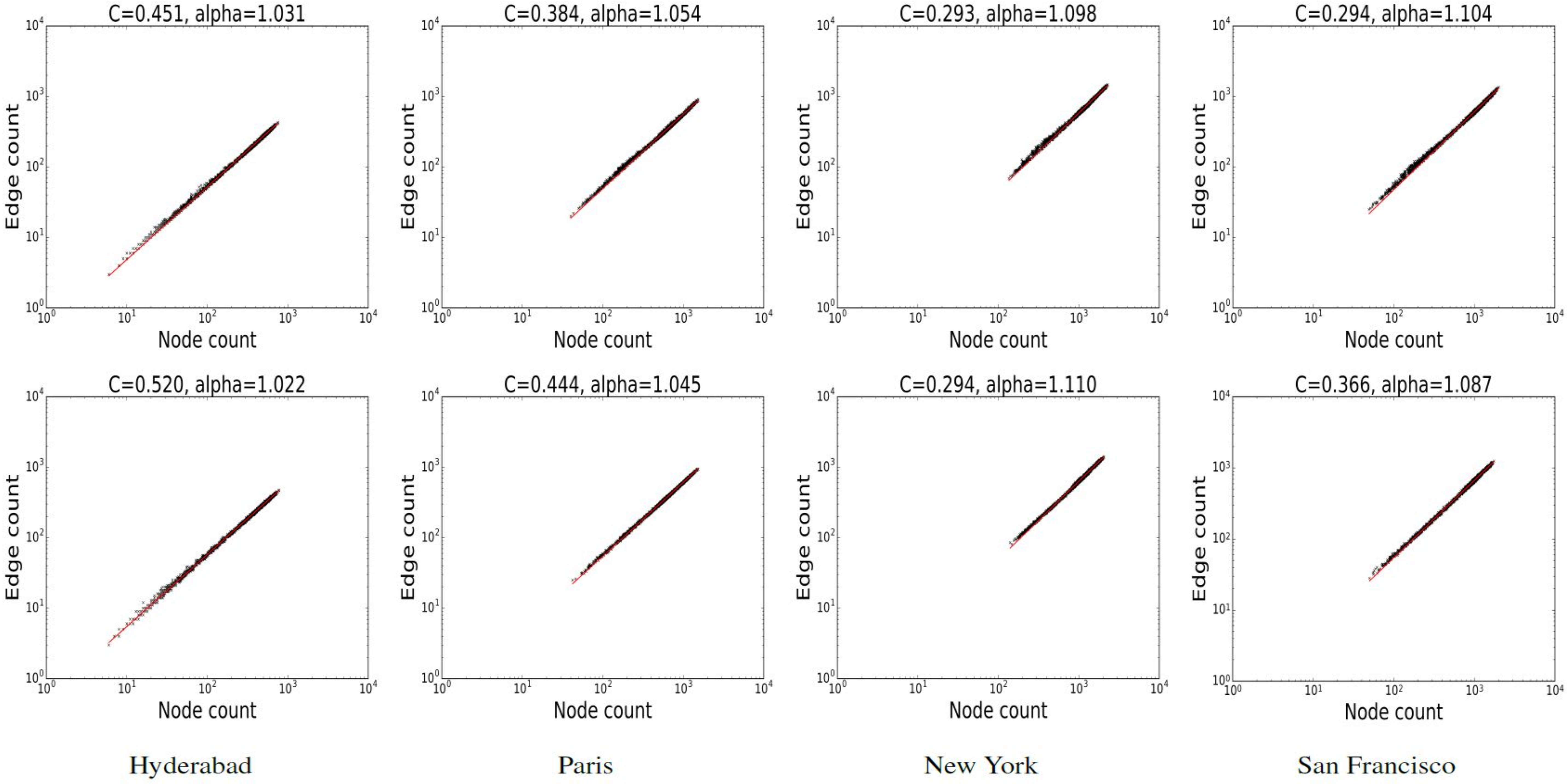


Figure 4: DPL plot from real data (top row) and synthetic data (bottom row) for four cities. The red line is the least square fit of the form $y = Cx^{\alpha}$, where y and x are number of edges and nodes respectively. $R^2 \approx 1.00$ for all of them.

Question 4: Is it possible to create an accurate **ride-request predictive system** based on the analysis of extensive historical data?

Question 5: Can such a ride-request predictive system lead to optimal **ride-pooling algorithm and service** at the city scale?

Joint Venture SILICON VALLEY NETWORK

CITY OF PALO ALTO

Bay Area Fair Value Commuting: Project Summary

An FTA MoD Sandbox Demonstration Project

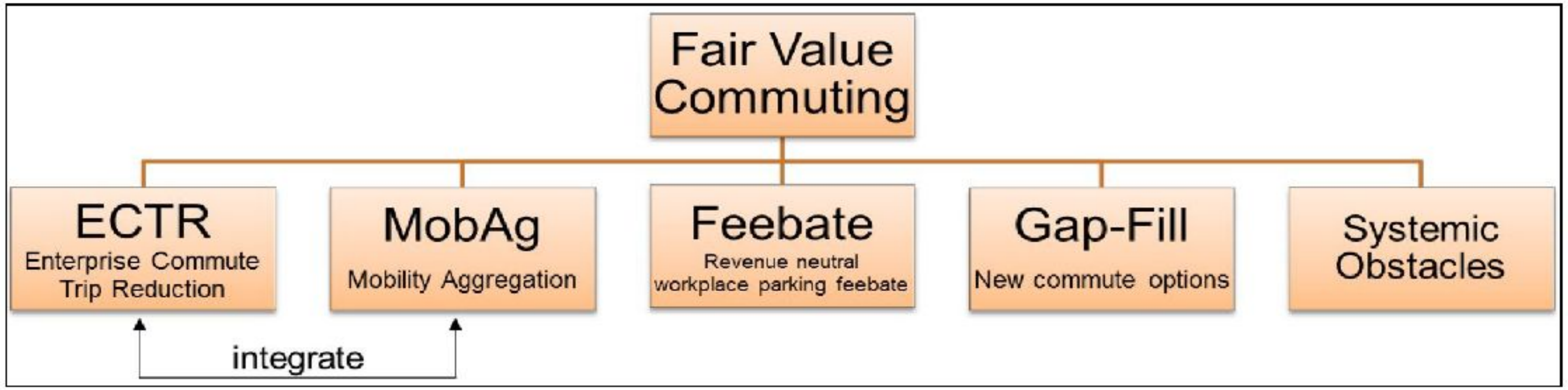
MOD MOBILITY ON DEMAND

Federal Transit Administration

Link to this google doc: http://bit.ly/FVCsummary

WHY: In pursuit of climate protection and traffic congestion relief, state/regional/local objectives have converged for 15% per-capita VMT reduction and 2X transit/biking.

WHAT: **In pursuit of regional objectives, our solution has the potential to reduce Bay Area SOV commute share from 75% to 50%.** Our technology/policy solution is called Fair Value Commuting (FVC) and consists of five components:



Promising avenues for future research. Some research questions:

1. If the ride pooling proximity constraints, both temporal and spatial, can be relaxed, is it possible to significantly improve ride poolability?
2. Can the pattern of temporal and spatial variation of ride poolability be leveraged to create intelligent predictive ride pooling algorithms?
3. Can we rigorously characterize the relationship between ride poolability and ride request graph densification power law factor?
4. Is it possible to use the space-time graph model extracted from historical data to perform real-time traffic congestion prediction and alleviation, as well as real-time accurate travel time prediction?

History of CMU in Silicon Valley

Epoch 1: 2002 – 2008 (School of Computer Science)

- Full-time & Part-time MS degrees in SE & SM

Epoch 2: 2008 – 2014 (College of Engineering)

- MS degrees in SE & SM; **INI** MS bi-costal degrees; **ECE** MS & PhD degrees

Epoch 3: 2014 – 2020 (College of Engineering)

- **Electrical & Computer Engineering (ECE)**: MS, MS-SE, & PhD
- **Information Networking Institute (INI)**: MSIT-IS, MSIT-MOB, MSIT-SM
- **Integrated Innovation Institute (III)**: MS-SM, MS-TV

Current CMU-SV Total: ~300 MS; ~40 PhD; ~20 Faculty; ~20 Staff

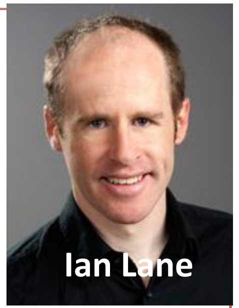
HUMANS Industry Partnership Program

- Membership categories:
  - ✓ **Founding Members**: $500K/year (4-year commitment)
  - ✓ **Associate Members**: $250K/year (2-year commitment)
- Represented industries:
  - ✓ Telecomm Infrastructure
  - ✓ Wireless Carrier
  - ✓ Mobile Hardware
  - ✓ Cloud Infrastructure
  - ✓ Transportation Systems
- Membership benefits:
  1. Direct access to faculty and PhD students (including internships & consulting)
  2. Access to pre-published research results (quarterly updates & annual reviews)
  3. Access to prototype software and tools (internal evaluation and commercialization)
  4. Non-exclusive royalty-free licenses to all IP generated through this initiative

Human Mobility Analytics and Services Initiative

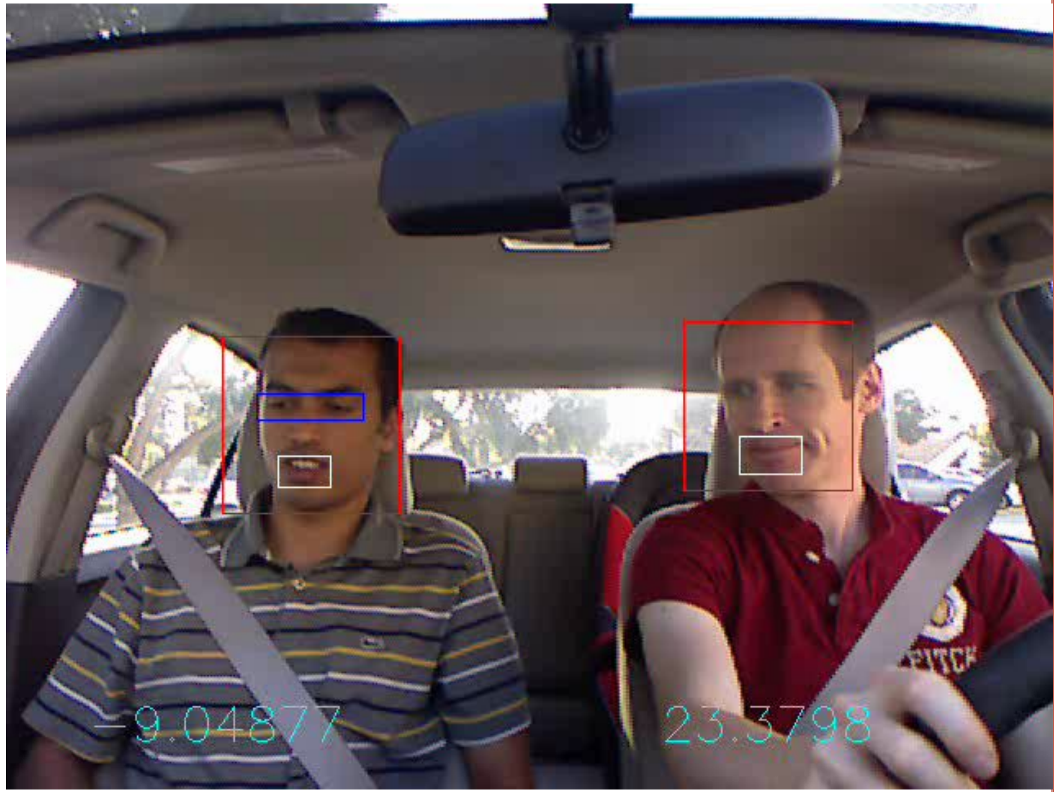
- Leadership:
  - Executive Director: Ole Mengshoel
  - Research Director: John Paul Shen
  - CMU Advisors: Raj Rajkumar, Dan Siewiorek, Jonathan Cagan, Burcu Akinci, Matthew Sanfilippo
- Founding ECE Faculty Members:
  - **Ian Lane**: Spoken Dialog systems and Embedded Speech Technologies
  - **Ole Mengshoel**: Data Analytics, Recommendation Engines, Inferencing
  - **John Paul Shen**: Mobile & Cloud-Edge Computing, Connected Vehicles
  - **Pei Zhang**: Sensor Networks, Cyber Physical Systems, Mesh Networks
- Affiliated CEE Faculty Members:
  - **Hae Young Noh**: Context-Aware Smart Structures, Statistical Signal Processing
  - **Zhen (Sean) Qian**: Intelligent Transportation Systems, Travel Behavior

Expertise: Machine Learning for Smart HCI



Speech and Interaction Technologies Group:

Inter-disciplinary research group looking at the intersection of….



Machine Learning

Distributed Machine Learning, Continuous Learning, Signal Processing, Speech Recognition, Image Processing, Automatic Model Optimization

Heterogeneous Computing

Machine Learning on Heterogeneous Computing: CPU, Multicore CPU, Manycore GPU, fPGA, DSP

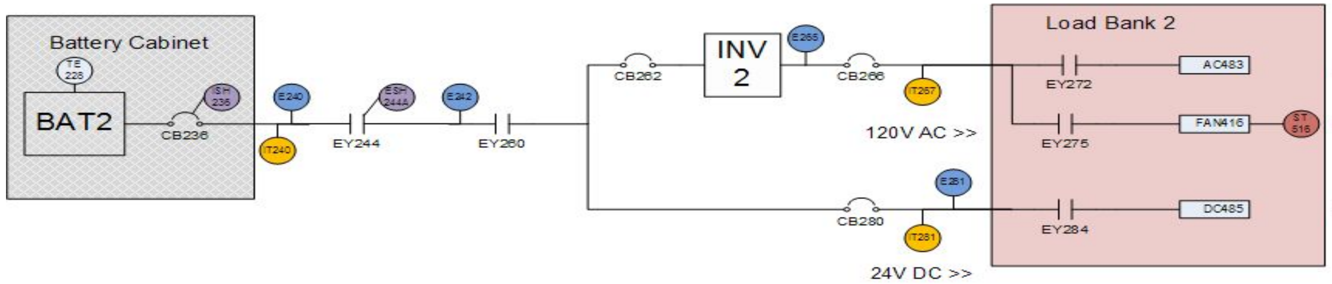
Human-Computer-Interaction

Speech and Multimodal Interaction, Context-Modeling, Cognitive Load Estimation, Spoken Language Understanding in Situated Environments

Expertise: Data Analytics & Machine Learning

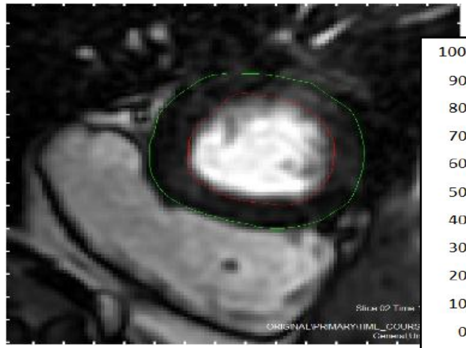
Algorithms

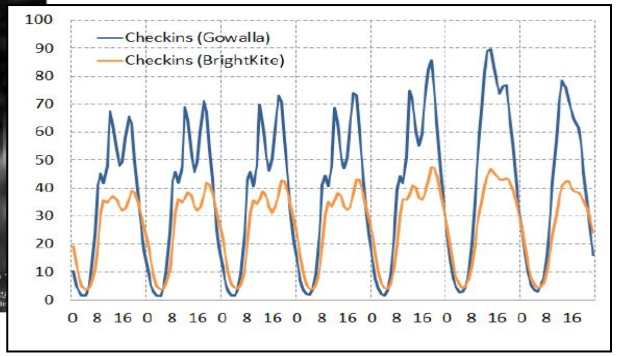
- Machine learning
- Stochastic optimization
- Inference, compilation



Models

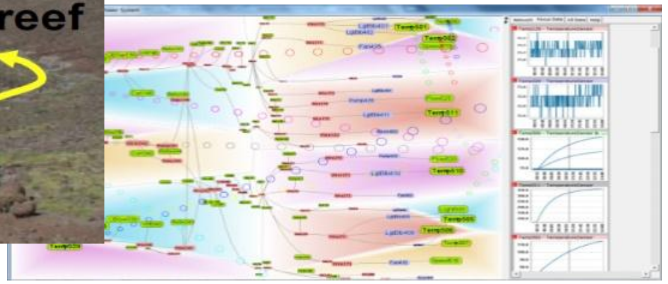
- Probabilistic graphical models
- Bayesian networks
- Markov chains
- Matrix factorization





Applications (and Experiments)

- Recommender systems
- Networks: computer, telecom, social, …
- Mobility : Vehicles, devices, …
- Science: Earth sciences , medical, …



Expertise: Mobile & Cloud-Edge Computing

Live Maps and Mixed Reality

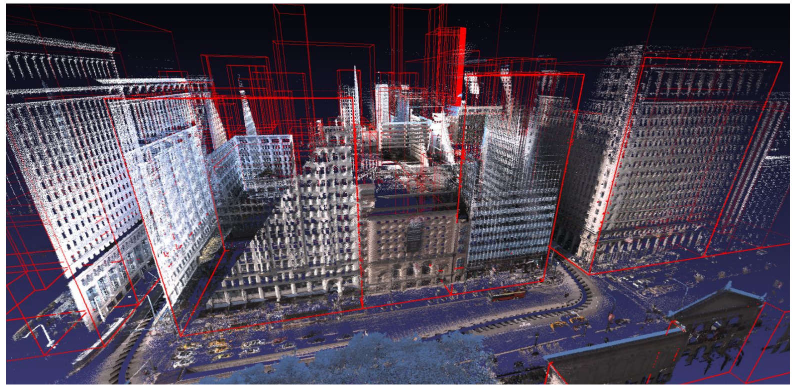
Next Generation 3D Interactive Maps
Visual, Local, and RT Event Searches
Live maps as a Platform for MR Apps

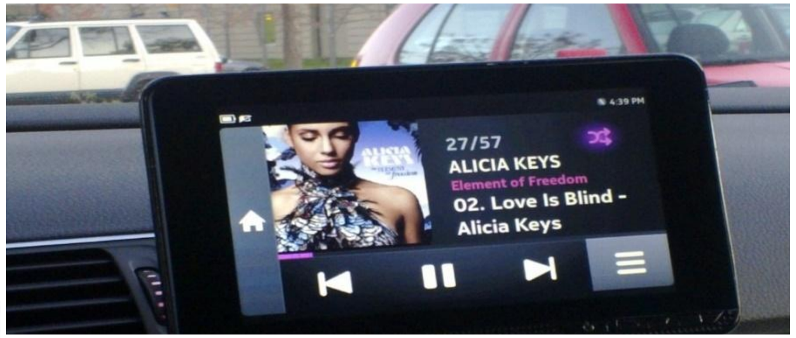
Connected Vehicles

Vehicle/Mobile Device Interoperability
Vehicles As Sensing/Cloud Platforms
Leverage ride-hailing service as platform

Federated Personal Computing

Multi/Cross-Device Seamless FPC UX
Mobile/Cloud Computing Convergence
Connected car as personal mobile server





Expertise: Sensing & Cyber Physical Systems

Internet of Things (drones, wearable, and devices)

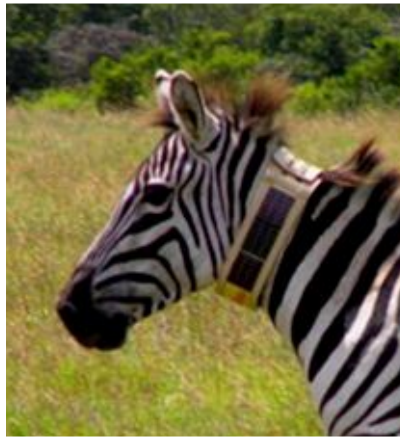
- Device localization
- Wearable muscle fatigue/activity inferencing
- Automatic device configuration and deployment

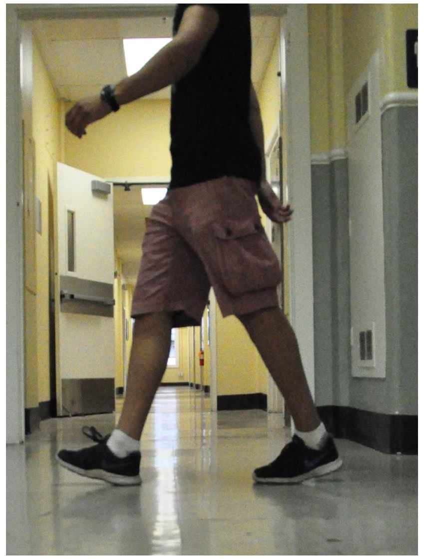
Building as Sensors

- Infer people’s characteristics (weight/height, etc.)
- Occupant health status (depression, happiness)
- Location/status/movement of people, and machines
- Hybrid model

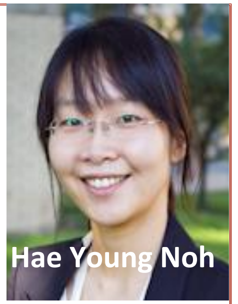
Applications (and Experiments)

- 10,000, taxi-based pollution monitoring in Shenzhen
- Elderly health monitoring in Vincentian Homes
- Baby Monitoring at Colorado Medical Center
- Consumer tracking at major retailers





Expertise: Context-Aware Smart Structures



Buildings as Sensors

within: occupant identity/location/health status (walking ability, stroke, etc.)
self: structural health monitoring (damage diagnostics & prognostics), energy management
around: surrounding traffic, earthquake, etc.

Vehicles as Sensors

within: driver pose/movement/breathing/heartbeat
around: road/railway/bridge/air pollution monitoring

Applications (and Experiments)

Elderly health monitoring in Vincentian Homes
Seismic damage diagnosis and risk analysis
Train based Pittsburgh light rail asset monitoring
10,000, taxi-based pollution monitoring in Shenzhen

Expertise: Intelligent Transportation Systems

Travel Behavior

- Data mining
- Game theory
- Network flow modeling

System Optimization

- Non-linear optimization
- Stochastic control
- Large-scale system simulation
- Transportation economics

Applications (and Experiments)

- Multi-modal: roadway, transit, parking, …
- Agencies: Incident, traffic management, air quality, pricing, road closure, HOV/HOT…
- Private sector: routing, sharing, facility allocation…