A tale of 2 continents and 4 cities

about the influence of demographics and social constraints on ride-sharing

Nikolaos Laoutaris**

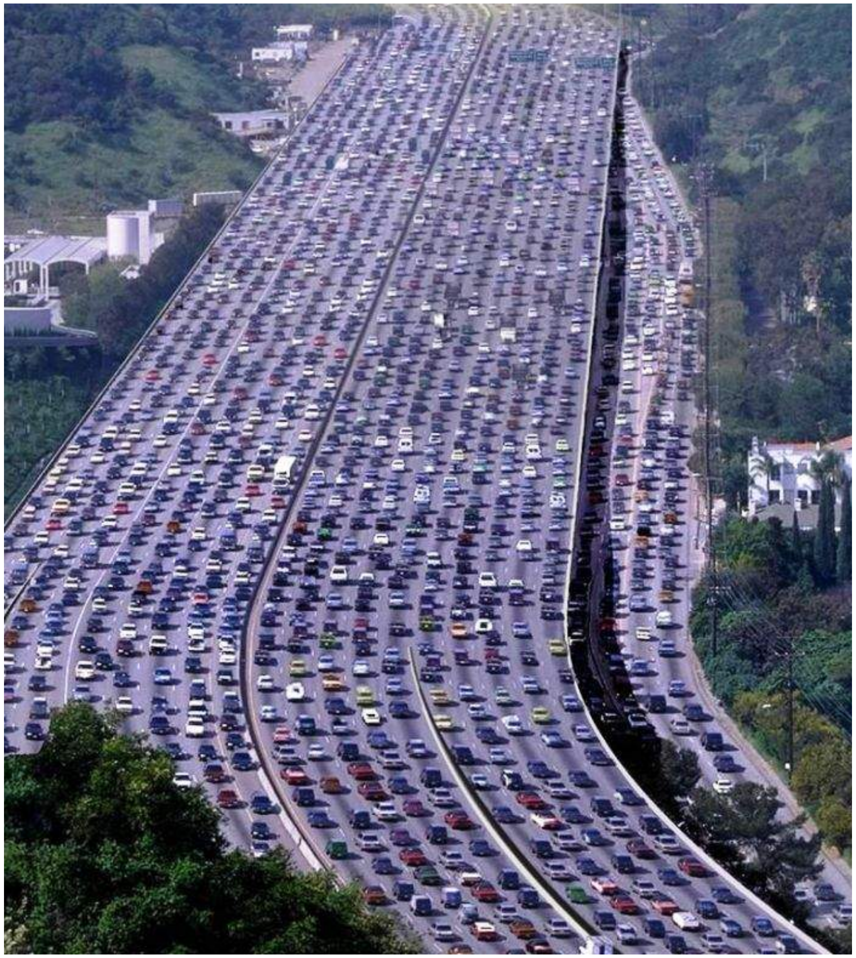
Blerim Cici*, Athina Markopoulou*, Enrique Frías-Martínez**,

*University of California, Irvine
**Telefonica Research

Outline

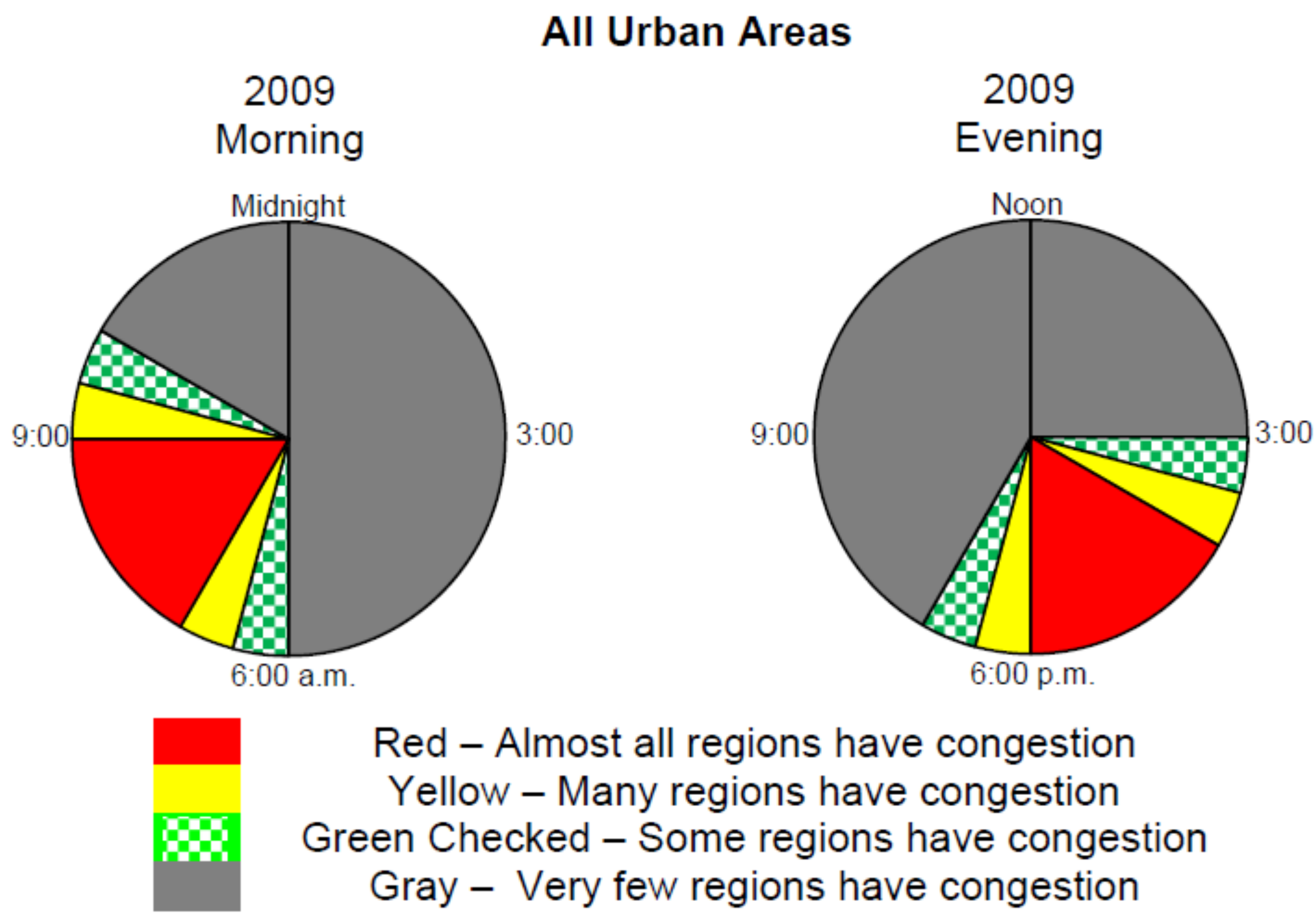
- Introduction
- Data
- Algorithms and Results
- Social Constraints
- System Design

Strong Car culture



- In US:
  - Commuters:128.3M
  - drive alone:75.7%
  - Bike:0.38%

Commuting


All Urban Areas
2009
Morning
Midnight
3:00
9:00
6:00 a.m.
2009
Evening
Noon
3:00
9:00
6:00 p.m.
Red – Almost all regions have congestion
Yellow – Many regions have congestion
Green Checked – Some regions have congestion
Gray – Very few regions have congestion

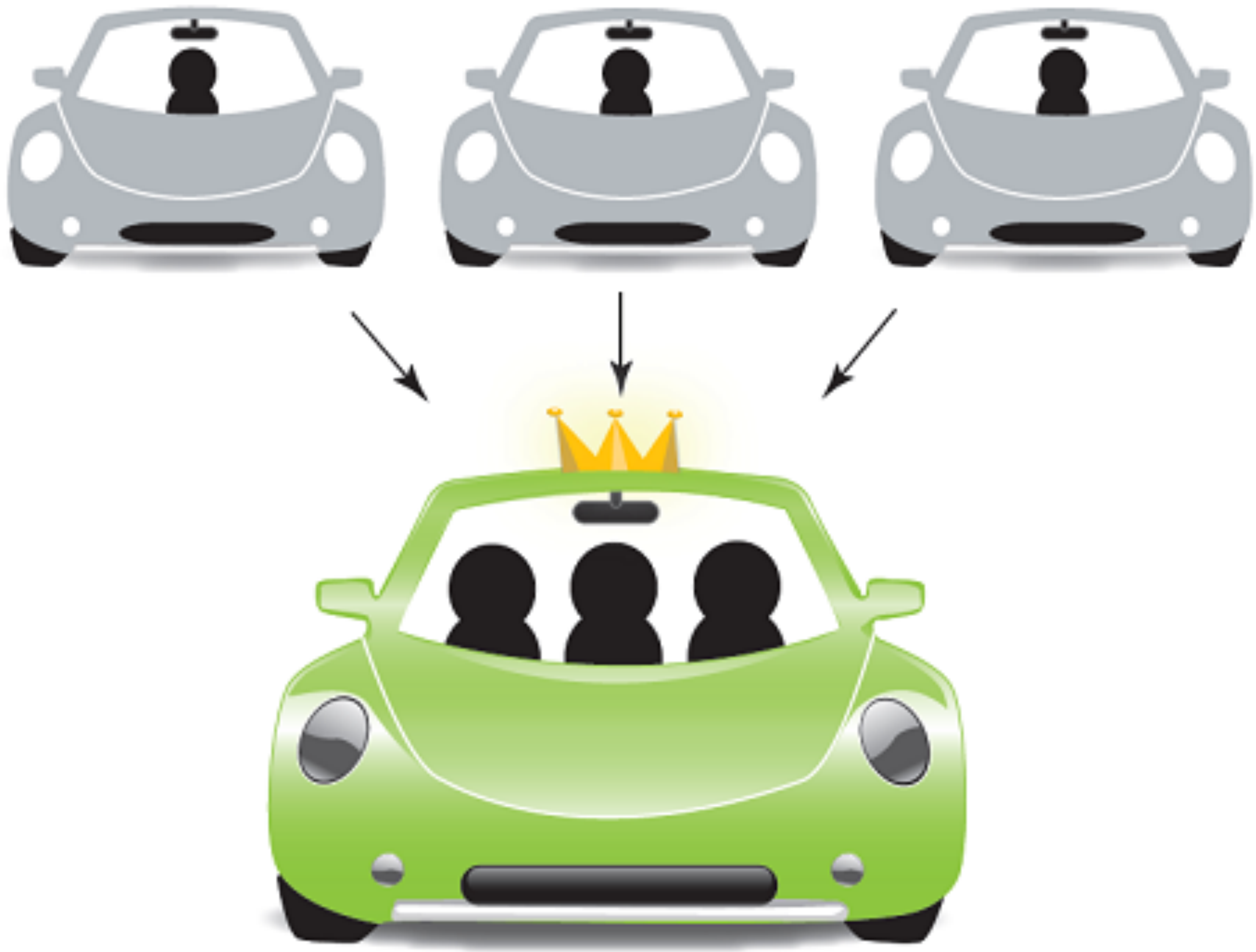
Annual cost of owing a 2010 VW Jetta

Example 2 - According to Edmunds.com, the cost to own a **2010 Volkswagen Jetta** 4 door Sedan over 5 years (assuming 15,000 miles a year) is approximately $35,500 broken down as follows:

	Year 1	Year 2	Year 3	Year 4	Year 5		5 Year Total
Depreciation	$3,732	$1,959	$1,725	$1,528	$1,371		$10,315
Taxes and Fees	$1,206	$80	$80	$80	$80		$1,526
Fuel	$1,644	$1,693	$1,744	$1,796	$1,850		$8,727
Maintenance	$29	$188	$540	$820	$1,328		$2,905
Repairs	$0	$0	$124	$297	$432		$853
Financing	$1,037	$833	$616	$385	$139		$3,010
Insurance	$1,531	$1,585	$1,640	$1,697	$1,757		$8,210
Yearly Totals	9,179	$6,338	$6,469	$6,603	$6,957		$35,546

Source: http://www.doughroller.net/smart-spending/true-cost-of-a-car-over-its-lifetime/

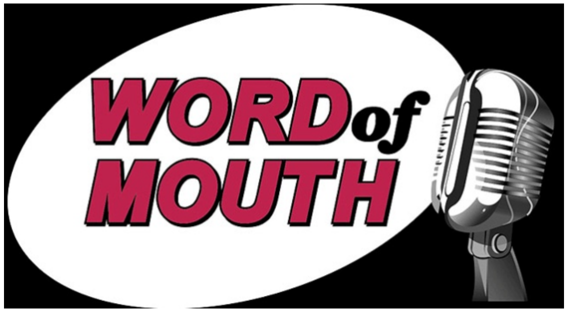
What is Ride-Sharing ?



Ride-Sharing: An old idea

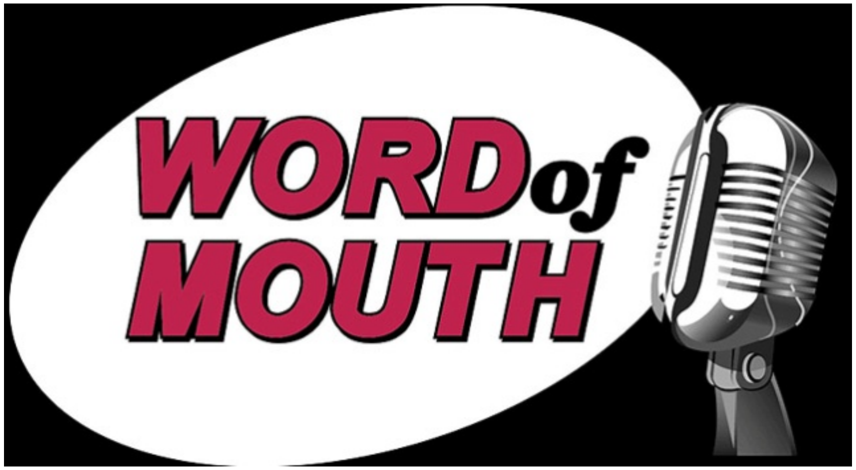
That never really made it to mainstream

Ride-Sharing in the past

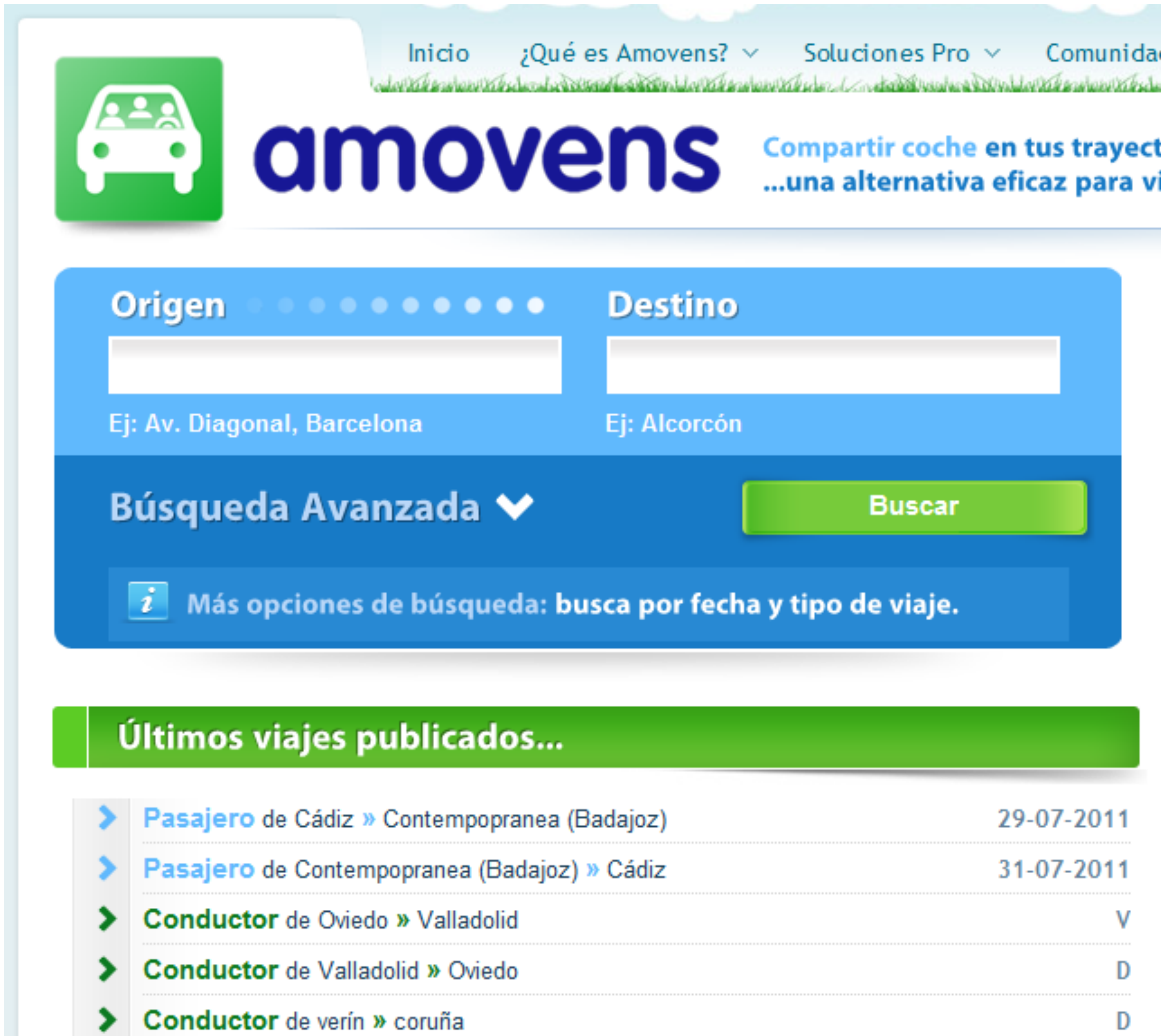


Ride-Sharing in the past

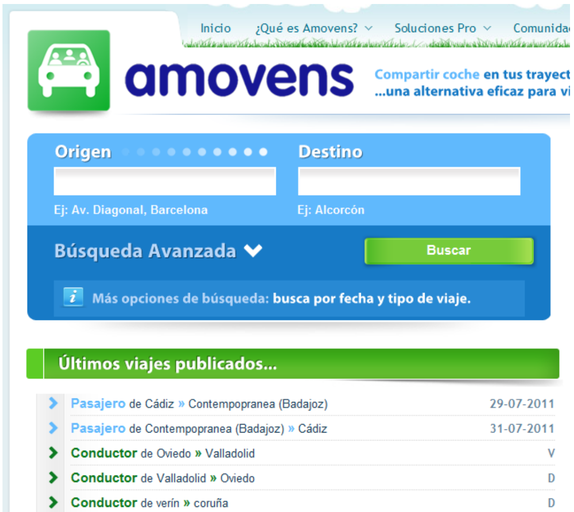
1. Few opportunities
2. Inflexible
3. Difficult to set up

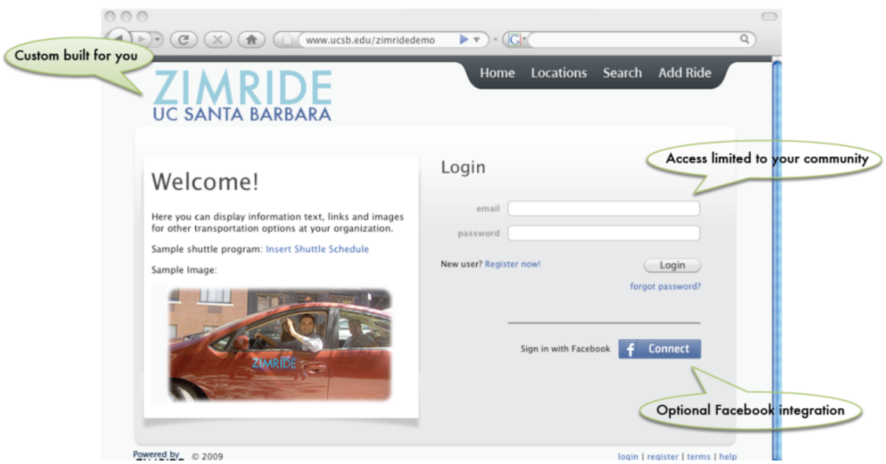


2nd gen Ride-Sharing: web based



2nd gen Ride-Sharing: web based





~~Difficult to set up~~

~~Few opportunities~~

Inflexible

Ride-Sharing Now

1. ~~Few opportunities~~

2. ~~Inflex~~

3. ~~Difficult to set up~~

But, why it's not mainstream yet ?

What affects ride-sharing?

- Mobility patterns:
  - Trajectories
  - Distribution of departure times
- User's tolerance:
  - Distance tolerance
  - Time tolerance
- Stranger danger: fear of sharing a ride with strangers.

Contributions

- We use large scale mobility data to derive bounds on the potential of ride-sharing.
- Formulate ride-sharing as a facility location problem, and developed efficient solutions
- Use social graph to study the effect of "stranger danger"
- Building a scalable Ride-Sharing system

Related Work

- Analysis of Ride-sharing
  - R.F. Teal. “Carpooling: Who, how and why.”, Transportation, Research, 1987.
  - W. He, D. Li, T. Zhang, L. An, M. Guo, and G. Chen. “Mining regular routes from gps data for ridesharing recommendations”, In UrbComp. ACM, 2012.
  - R. Trasarti, F. Pinelli, M. Nanni, and F. Giannotti. “Mining mobility user proles for car pooling”. In Proc. UrbComp., ACM, 2011.

- Quantification of Ride-sharing potential
  - H.-S. J. Tsao and D. Lin, “Spatial and temporal factors in estimating the potential of ride-sharing for demand reduction”, California PATH Research Report, UCBITS-PRR-99-2, 1999.
  - A. M. Amey, J. P. Attanucci, “Real-Time Ridesharing: Exploring the Opportunities and Challenges of Designing a Technology-based Rideshare Trial for the MIT Community”

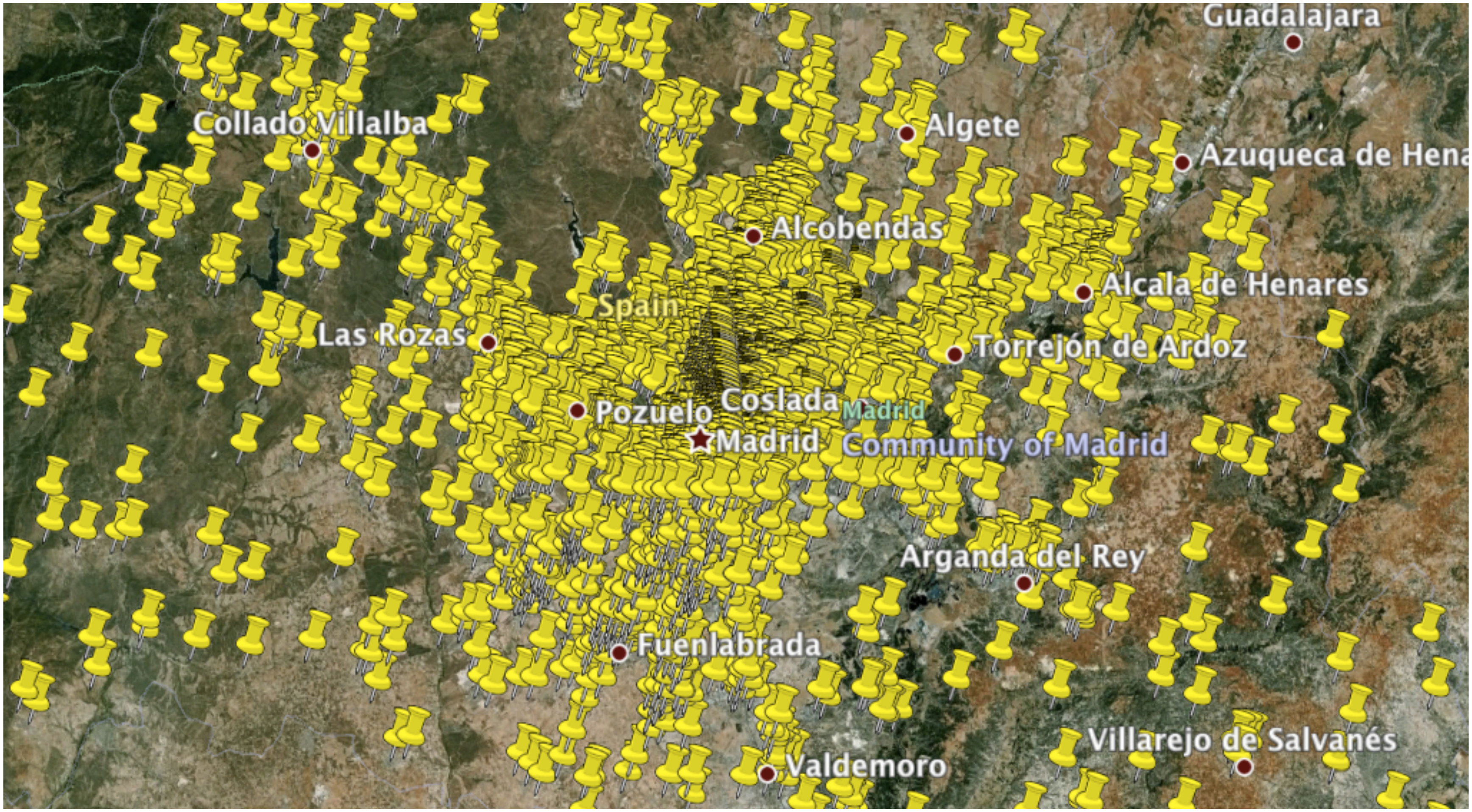
Related Work

- CDR Analysis and Human Dynamics
  - M. C. Gonzalez, C. A. Hidalgo, A. L. Barabasi, "Understanding individual human mobility patterns", Nature 2008
  - S. Isaacman, R. Becker, R. Caceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky, "Identifying Important Places in People's Lives from Cellular Network Data", Pervasive 2011
  - E. Cho, S. A. Myers, J. Leskovec, "Friendship and Mobility: User Movement In Location-Based Social Networks", KDD 2011
  - F. Calabrese, F. C. Pereira, G. Di Lorenzo, L. Liu, and C. Ratti, "The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events", Pervasive 2010
- Call Description Record Analysis
  - V. Frias-Martinez, J. Virseda, A. Rubio, E. Frias-Martinez, "Towards Large Scale Technology Impact Analyses: Automatic Residential Localization from Mobile Phone-Call Data", ICTD, 2010
  - V. Soto, E. Friaz-Martinez. "Automated Land Use Identification using Cell-Phone Records", HotPlanet 2011

Outline

- Introduction
- Data
- Algorithms and Results
- Social Constraints
- System Design

Mobile phone data & location info.



CDR Data

- Call Description Records (CDRs):
  - Every phone call: caller#, callee#, timestamp, cell-tower coordinates …
  - Maintained for billing purposes

- Our CDR dataset:
  - September – December 2009
  - 5M users in Madrid (820M calls)
  - 2M users in Barcelona (465M calls)

Geo-tagged Tweets

Geo-tagged Tweets

- JSON Format
- Contains:
  - User id
  - Timestamp
  - <lat, lng> coordinates
  - Text

NY: 5.2M tweets, 225K users
LA: 3.23M tweets, 155K users

Identifying Home/Work

- Small set of users with known Home/ Work addresses … the rest.

Home/Work locations:

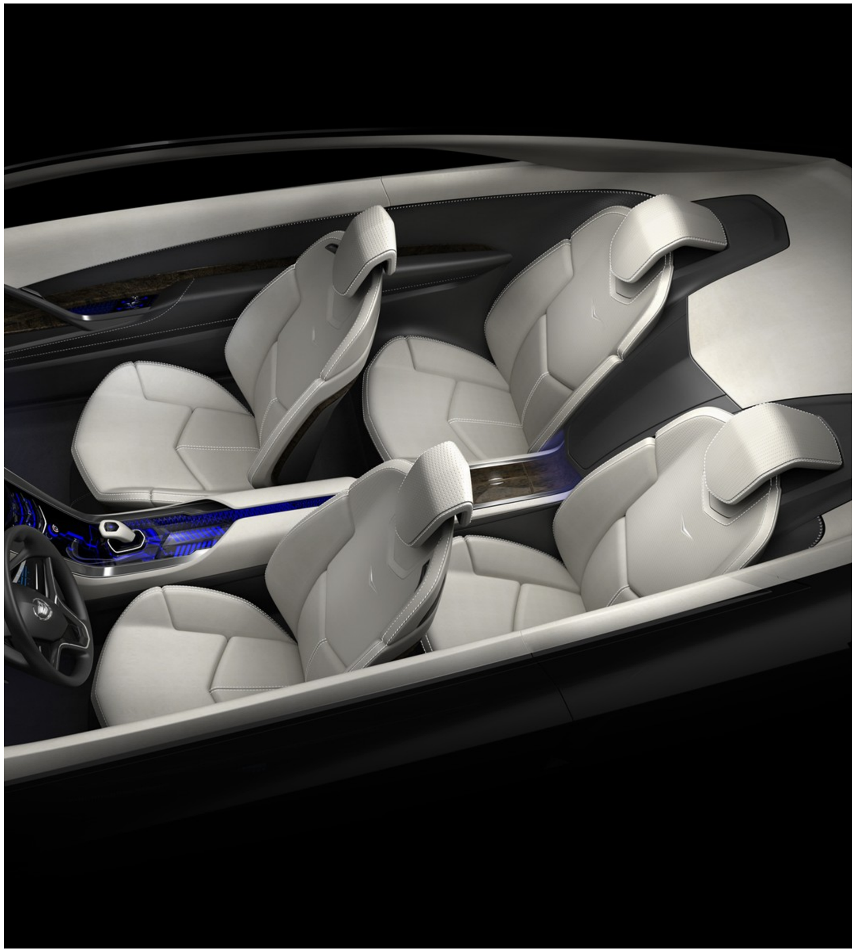
Madrid (CDRs) : 272, 479

New York (Twitter): 71,977

Outline

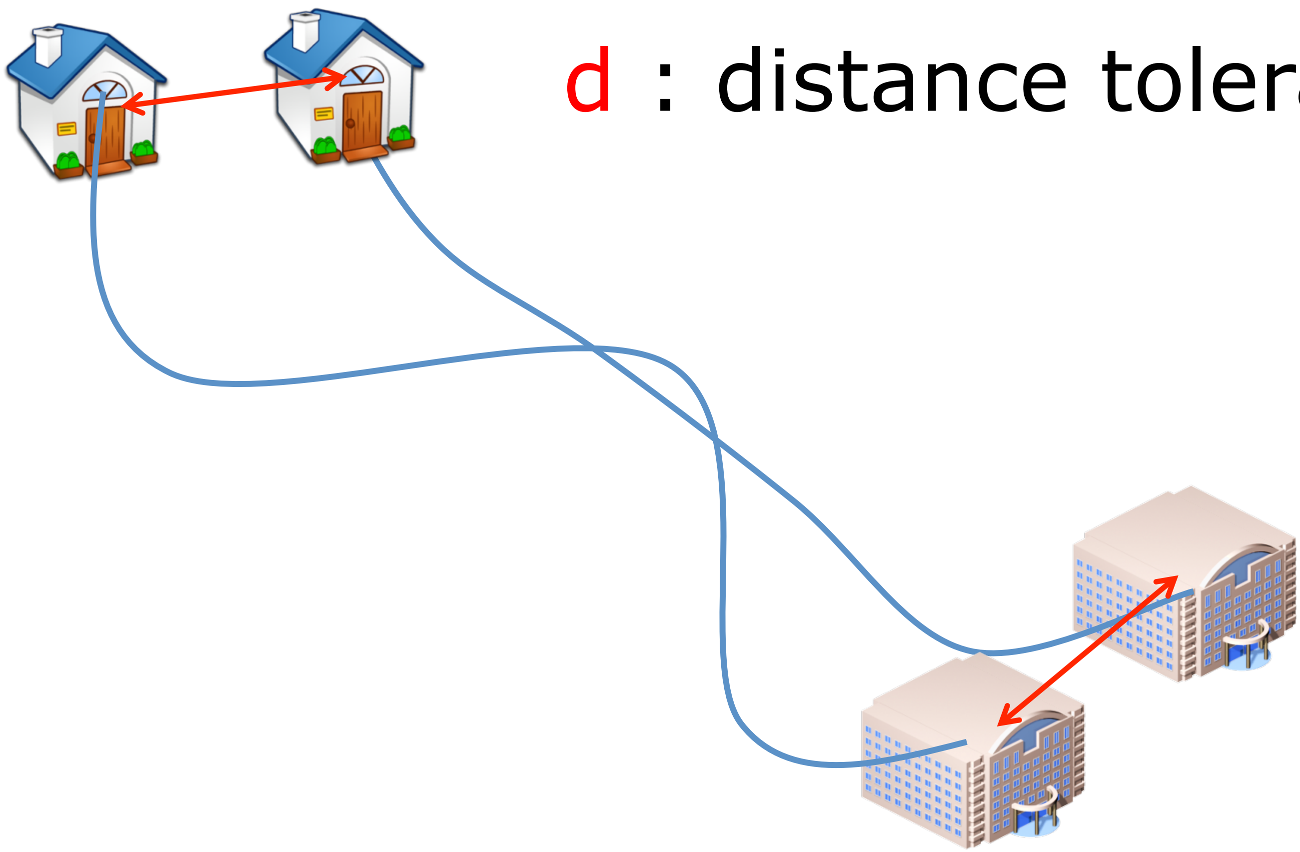
- Introduction
- Data
- Algorithms and Results
- Social Constraints
- System Design

Initial Assumptions



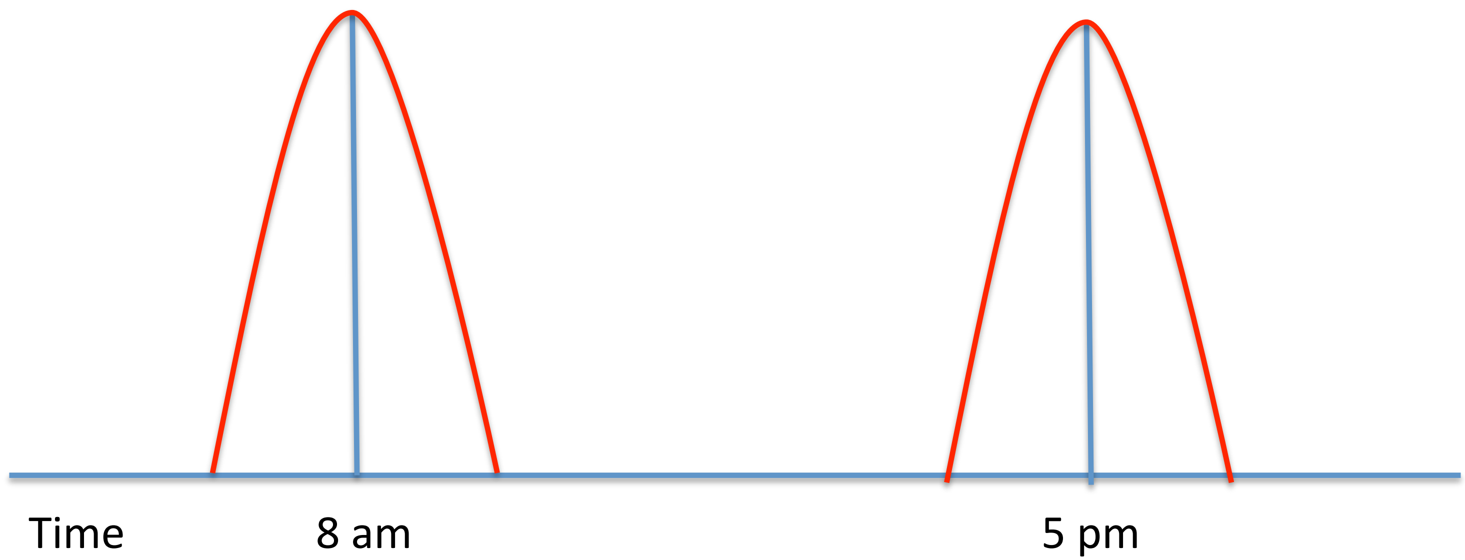
- Stranger danger is not a problem
- All cars have a capacity of 4

Space proximity



d : distance tolerance

Time proximity



- σ : standard deviation of Home/Work departure times
- τ : time tolerance

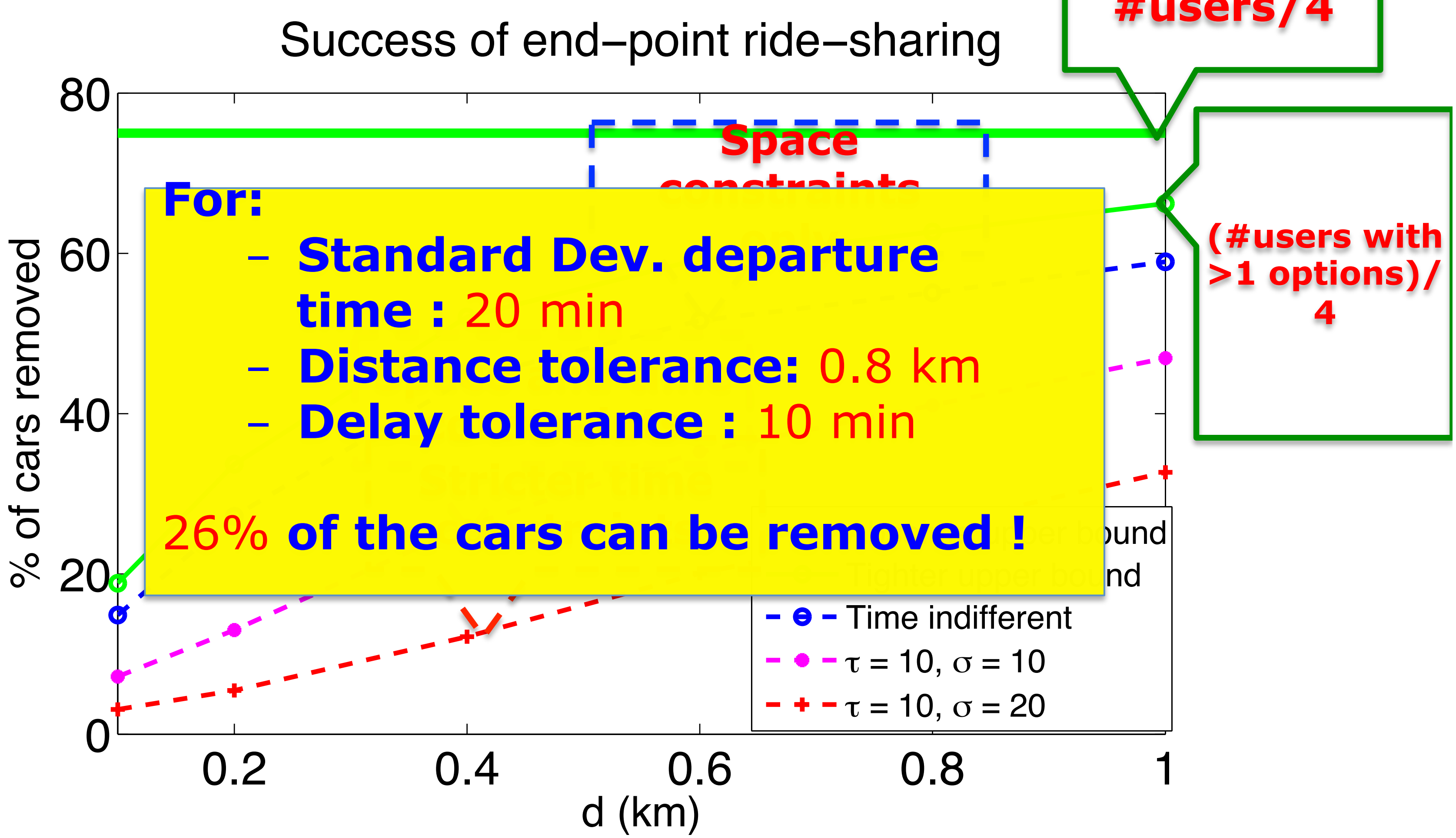
Formulation

- Goal: minimize the number of cars give spatial and time constraints
- Capacitated Facility location with Unsplittable Demands:
  - Facilities : Drivers
  - Clients : Passengers
- Distance function:
  - d(u,v) = max{h_dist(u,v), w_dist(u,v)}

EndPoints RS

- Since our problem is NP-hard, we use an efficient and scalable heuristic
- EndPoints RS:
  - Start with an initial “smart” solution
  - Iterative improvements by local search in solution space
- Scalability
  - Fixed local search steps
  - Fix numbers of iterations

Results for EndPoint RS



For:

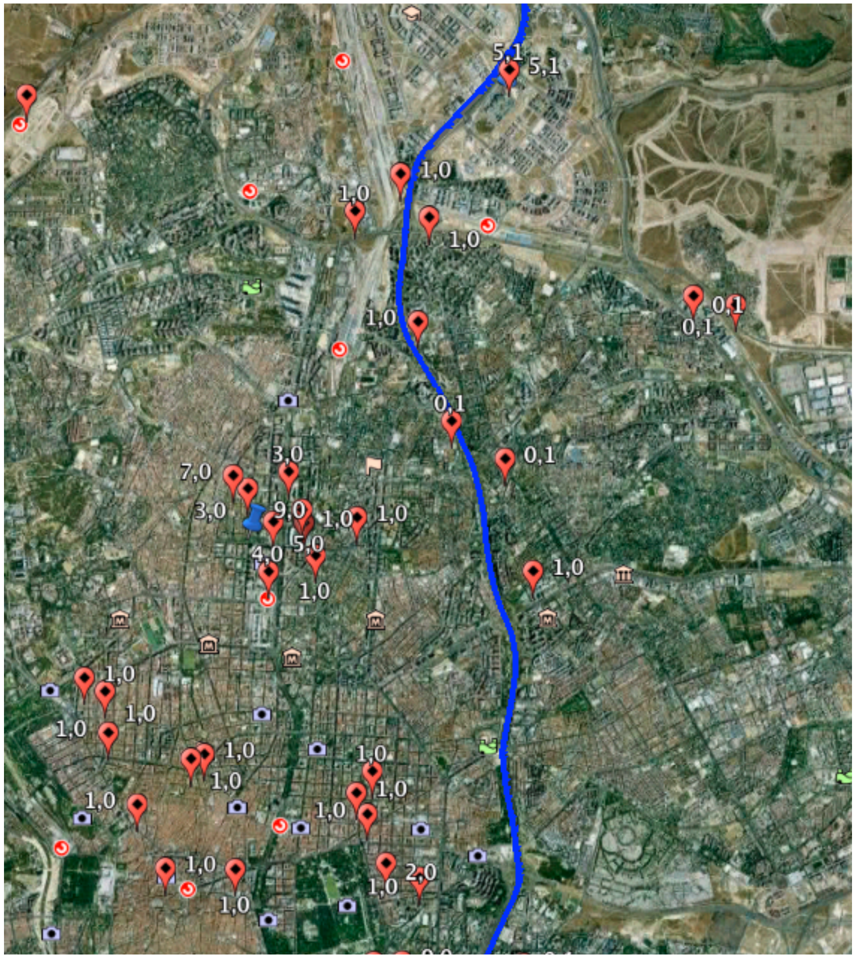
- **Standard Dev. departure time :** 20 min
- **Distance tolerance:** 0.8 km
- **Delay tolerance :** 10 min

26% **of the cars can be removed !**

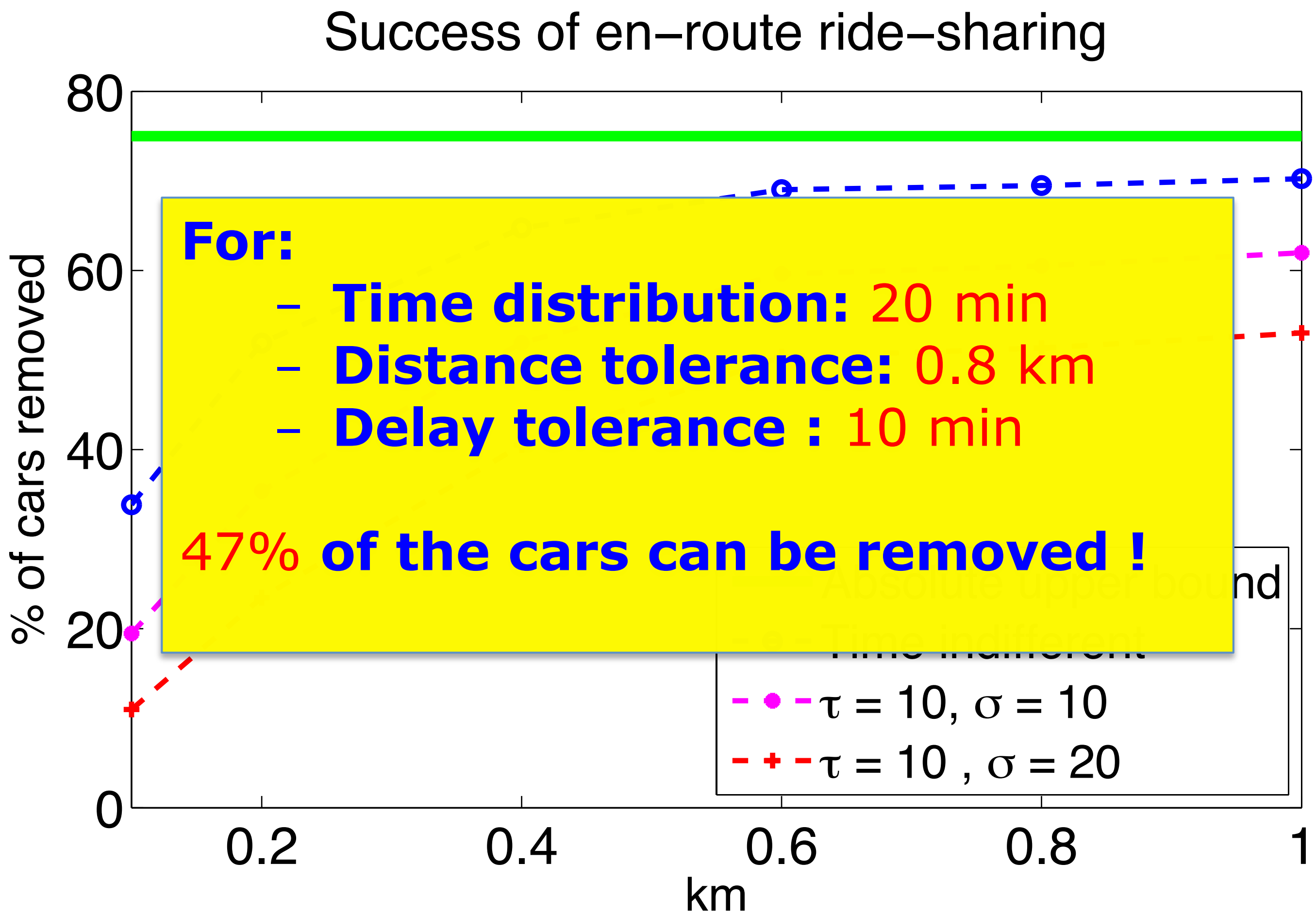
EnRoute RS

- Find Home/Work path through Google Maps

- EnRoute RS:
  - Get the solution of EndPoints RS
  - Iterative improvements
  - Fill empty seats by pick-ups



Results for EnRoute RS



For:

- **Time distribution:** 20 min
- **Distance tolerance:** 0.8 km
- **Delay tolerance :** 10 min

47% **of the cars can be removed !**

Outline

- Introduction
- Data
- Algorithms and Results
- Social Constraints
- System Design

Reducing “Stranger Danger”

- Assume users are willing to share a ride only with:
  - friends
  - friends of friends

- Social graphs:
  - CDRs : call graph
  - Twitter : mutual declared friendship

Filtering with social constr.

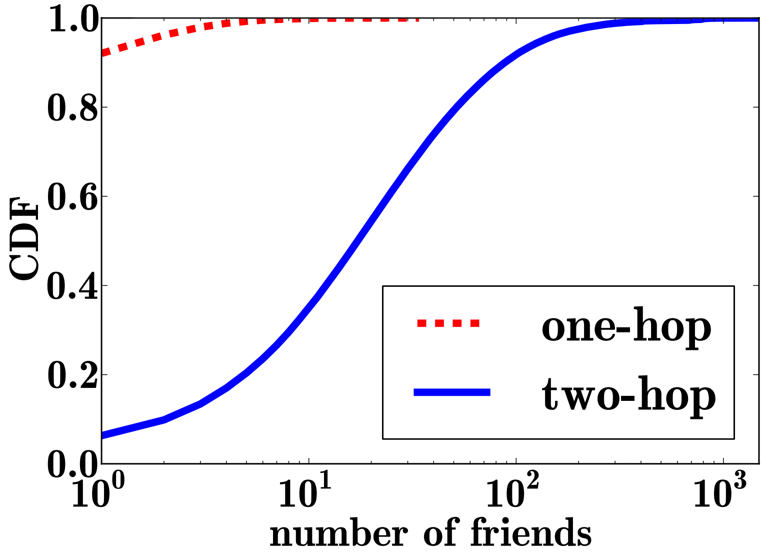
Ride-sharing parameters:

- Time distribution: 20 min
- Distance tolerance : 0.8 km
- Delay tolerance : 10 min

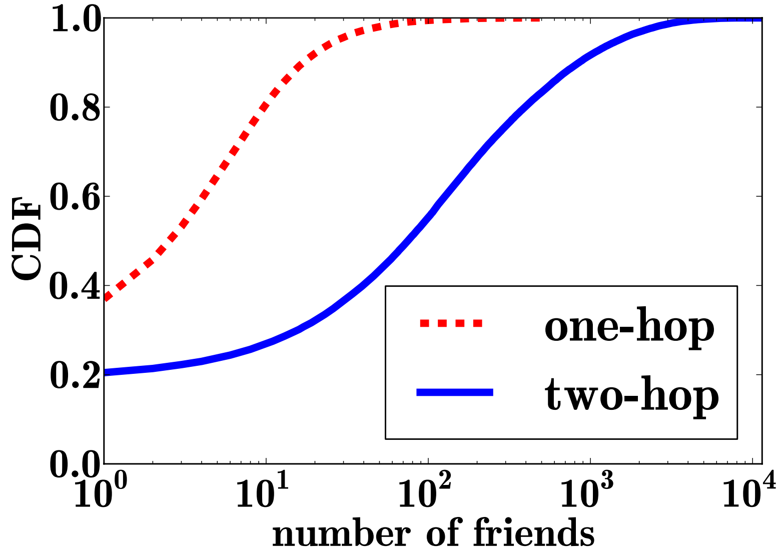
City	Friends only	Friends of friends	Anybody
Madrid	0.2%	2.4%	47%
New York	1.5%	9.1%	52%

Social graph properties

CDRs - Madrid



Twitter graph – New York



The next big question !!!

How to design an efficient Ride-Sharing application ?

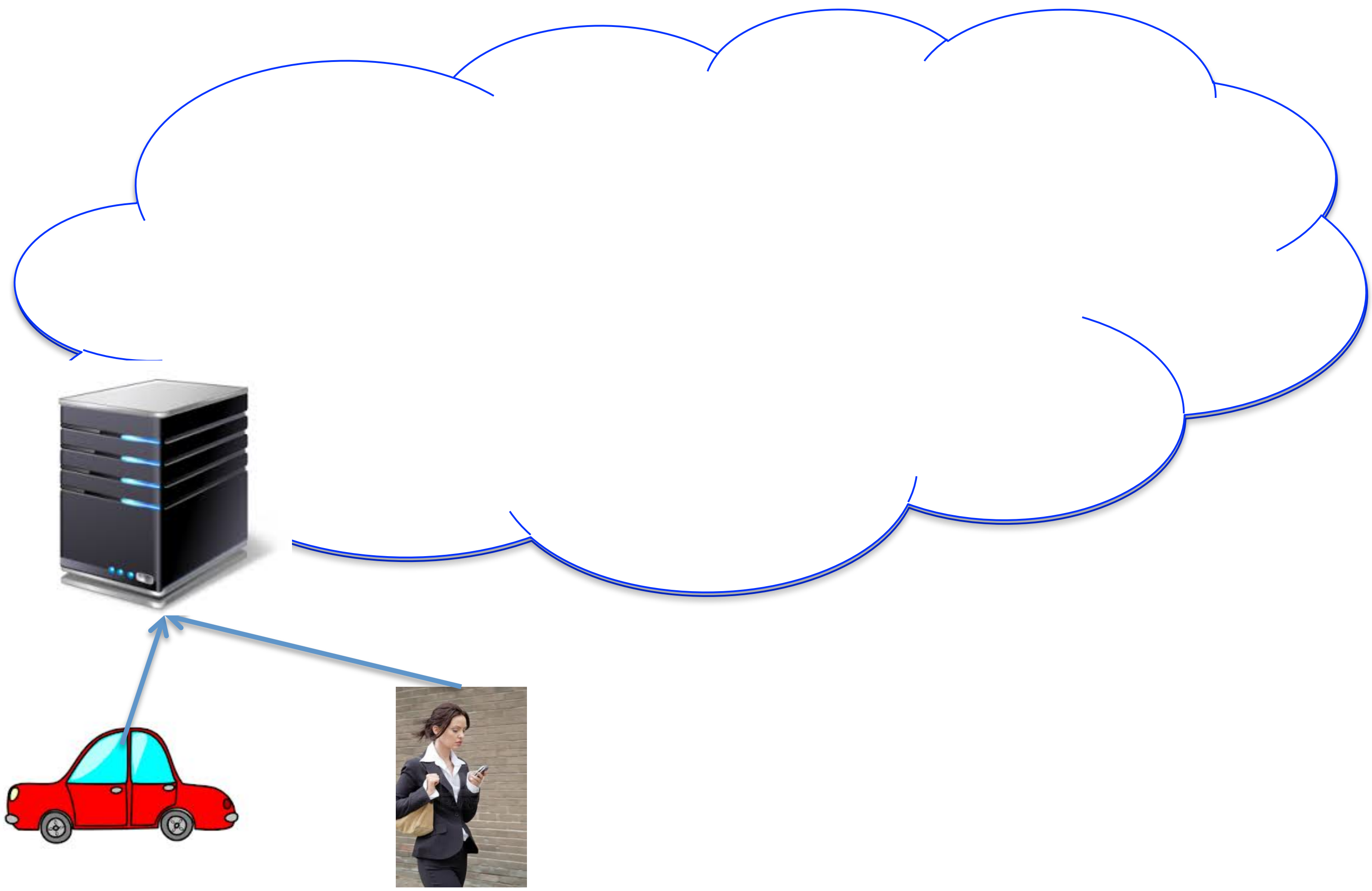
Outline

- Introduction
- Data
- Algorithms and Results
- Social Constraints
- System design

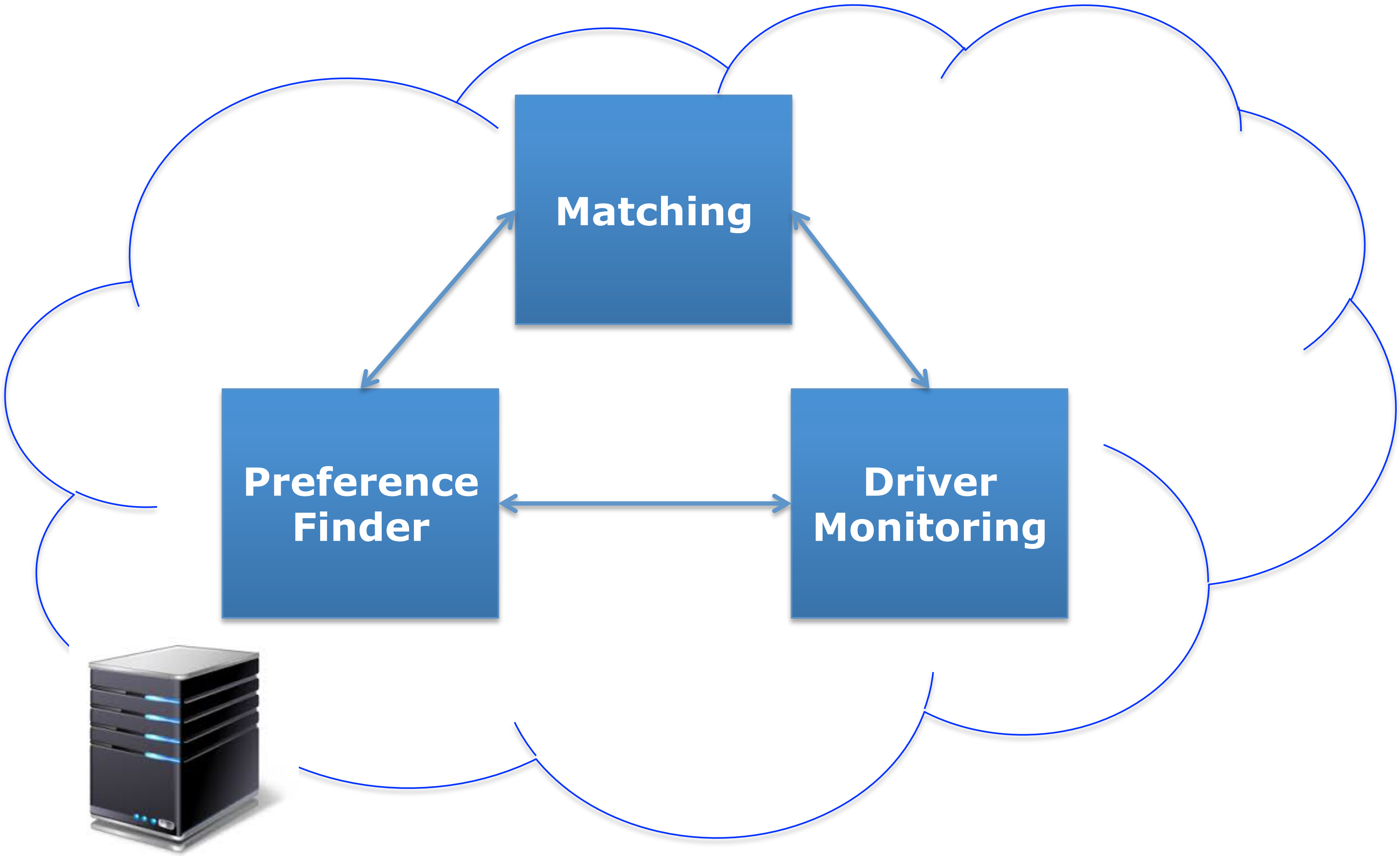
Requirements

- Immediate response to request
- Spatial-temporal constraints:
  - Max dist. : 0.8 km
  - Max deviation from time routine: 10 min
- Matching ratio is crucial !

Online Ride Sharing



Online Ride Sharing



Online Ride Sharing

Preference Finder

Role: Find **<passenger, drivers>** meeting spatio-temporal constraints !
Challenge: Scalable, real-time spatio-temporal queries !

Matching

Role: Assign passengers to drivers, based on their preferences.
Challenge: High matching ratios, small departure delay, social proximity between drivers and passengers.

Driver Monitoring

Role: Monitor drivers and estimate pick-up times.
Challenge: Generate accurate estimations, deal with delays.

Preference Finder

- Current implementation based on KDTrees
- For 272K users can run on a single machine

The heart of the system

p1: {p1's preferences}
p2: {p2's preferences}
…
…
…
pn: {pn's preferences}

→ **Matching** →

<p1, d1>
<p2, d2>
…
…
…
<pn, dn>

The matching algorithm:

- Match new request as soon as they arrive
- Refines existing **(driver, passng)** pairs every 2 mins.
- Use distance function to model preferences.

Distance function

- dist(d, p) =

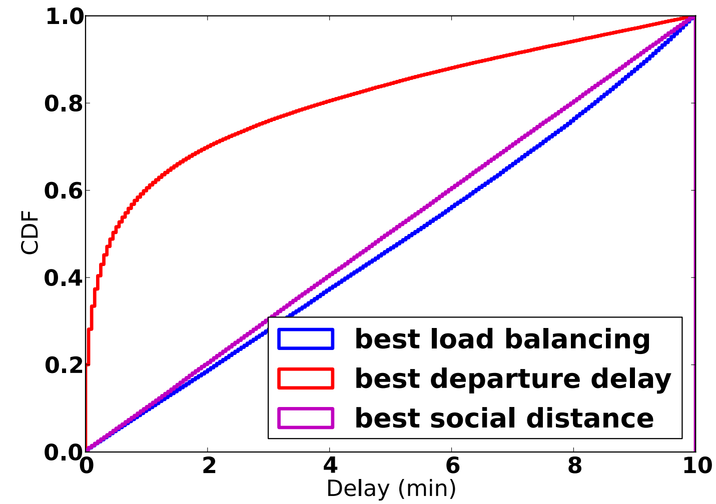
  social_weight * social_dist (d, p)

  + **time_weight** * time_dist(d, p)

  + **load_balance_weight** * empty(d)

Extreme cases

Cases	Driver ratio (%)	Passenger ratio (%)	Social sharing (%)
social best	62	79	6
time best	48	79	0.4
load-balancing best	73	79	0.5



Prediction Errors

- On line ride sharing depends on en-route pick-ups
- Predicting driver arrival time of users is very important
- How is ride-sharing affected by prediction errors

Driver Monitoring

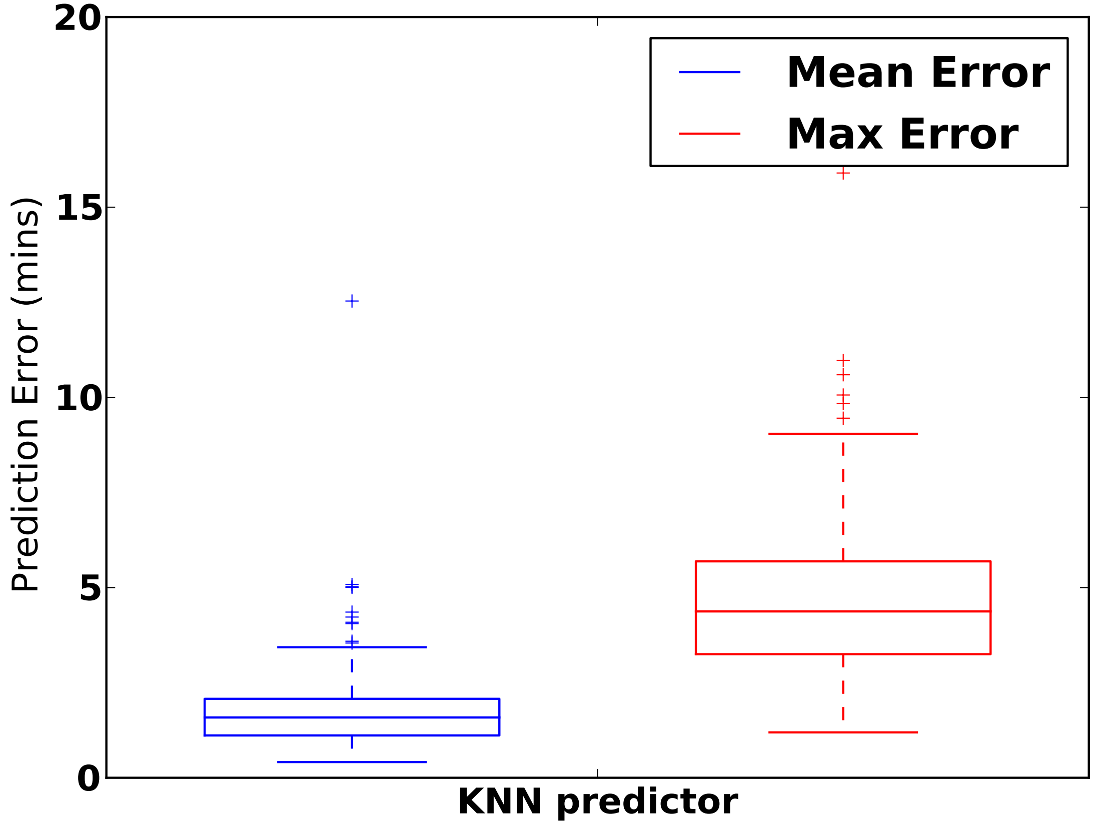
d1:(lat, lng) … (lat, lng)
d2:(lat, lng) … (lat, lng)
…
…
…
dn:(lat, lng) … (lat, lng)

d1: loc. estimation
…
…
…
dn: loc. estimation

Modeling prediction error

- We implemented a state-of-the-art arrival prediction algorithm
- We used **GPS** a dataset of 500 taxi drivers in Silicon Valley.
- Modeled the error & plugged it in our simulations.

Modeling prediction error



Extreme cases – Matching Ratio

Best Load balancing

- **Weights:**
  - **social_weight** = 0
  - **time_weight** = 0
  - **load_balance_weight** = **1**

- Canceled pairs: 5%

Lowest time deviation

- **Weights:**
  - **social_weight** = 0
  - **time_weight** = **1**
  - **load_balance_weight** = 0

- Canceled pairs: 3%

Summary

- We evaluated the potential or ride-sharing with using CDRs, and geo-tweets.

- Results:
  - The success of of ride-sharing can be as high as 47%, if we don’t consider *“stranger danger”*
  - ONLY with friends is too restrictive
  - Sharing rides with friends of friends, can lead to a success up to 9.1%, depending on the density of the social graph

Summary

- Ride-sharing application with a real-world feeling

- Highlighted trade-offs in the design of ride-sharing system.

- Showed the impact of arrival predictions such a system

Beyond Ride-Sharing

Thank You

More at: http://people.tid.es/Nikolaos.Laoutaris