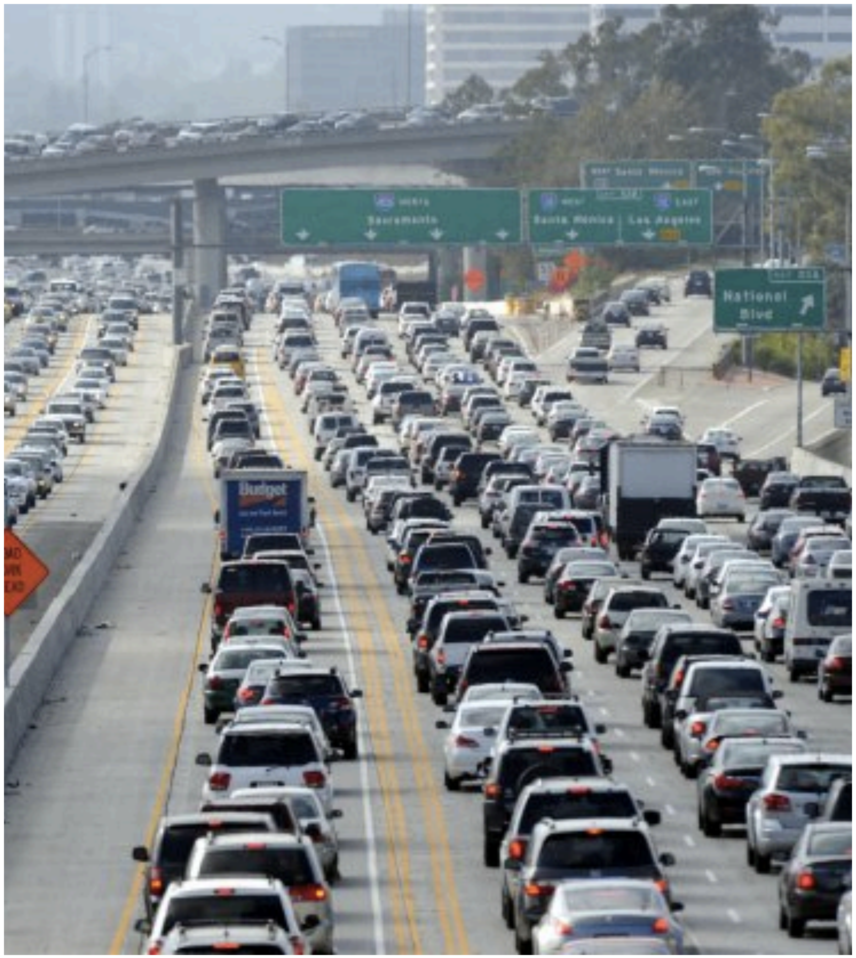
Assessing the Potential of Ride-Sharing Using Mobile and Social Data

Blerim Cici,
Athina Markopoulou

Nikolaos Laoutaris,
Enrique Frias-Martinez

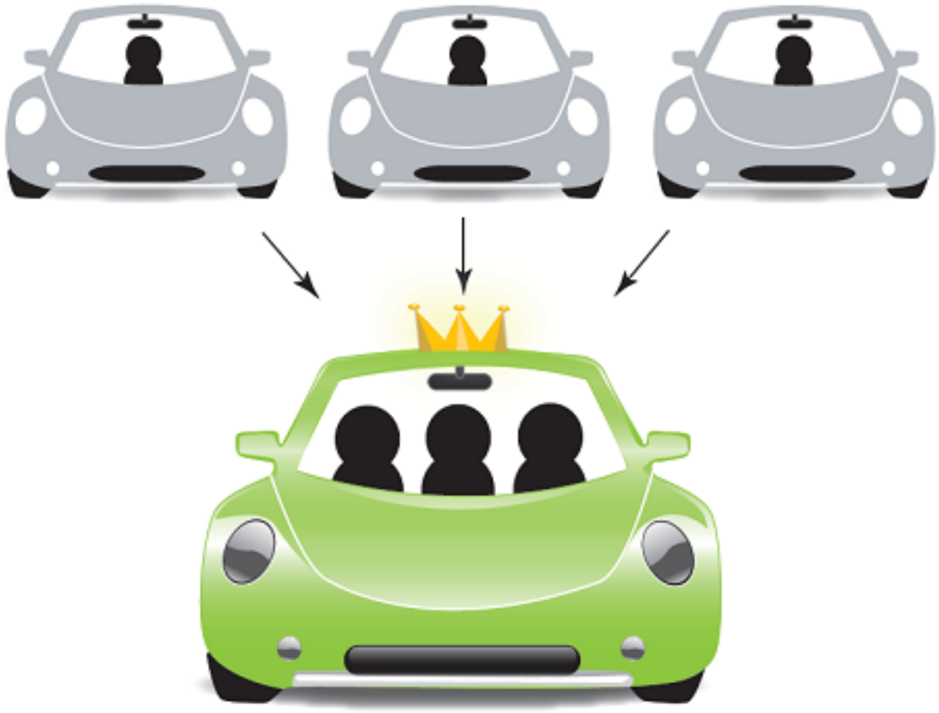
Car Usage and Impact



- In USA*:
  - Commuters: 132.3M
  - Driving alone: 79.9%
- Impact
  - Pollution
  - Lost productivity
  - High car expenses

*Brian McKenzie, "Out of state and long commutes:2011", American Community Survey Reports, 2011

Introducing Ride-Sharing



An Old Idea, yet …

- Challenges:
  - Live/Work close by
  - Similar schedules
  - Avoid strangers
- Opportunities:
  - Smartphones
  - Social media

An Old Idea, yet …

- Challenges:
  - Live/Work close by
  - Similar schedules
  - Avoid strangers
- Opportunities:
  - Smartphones
  - Social media

Full potential of ride-sharing is still unknown

Related Work

- 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.
- 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 profiles for car pooling". In Proc. UrbComp., ACM, 2011.
- 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"

Goal: Assess Ride-Sharing Potential

- Q: How many cars can be removed ?

- Ideal Data:
  - For all people in a city
  - Full commuting trajectories
  - Willingness to share a ride

- Available Mobile and Social Datasets:
  - Large (but not entire) population
  - Samples of trajectories
  - (Parts of) social media graphs

Goal: Assess Ride-Sharing Potential

- Q: How many cars can be removed ?

- Ideal Data:
  - For all people in a city

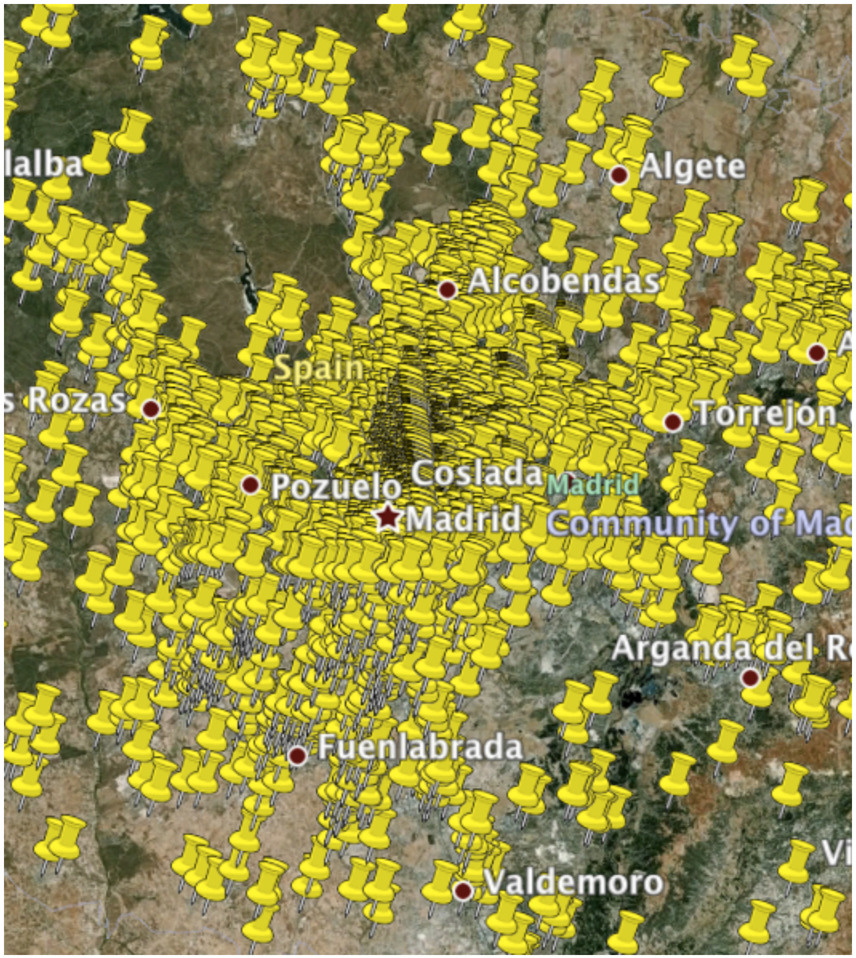
Find an upper bound to the ride-sharing potential

- Available Mobile and Social Datasets:
  - Large (but not entire) population
  - Samples of trajectories
  - (Parts of) social media graphs

Outline

- Introduction
- Datasets
- Algorithms for Matching Users
- Results

Call Description Records (CDRs)



- Spatio-temporal:
  - Cell tower coordinates
  - Timestamps
- Social:
  - Calls among users
- Details:
  - Sept – Dec 2009
  - Madrid: 820M calls, 5M users
  - Barcelona:465M calls, 2M users

Geo-tagged Tweets

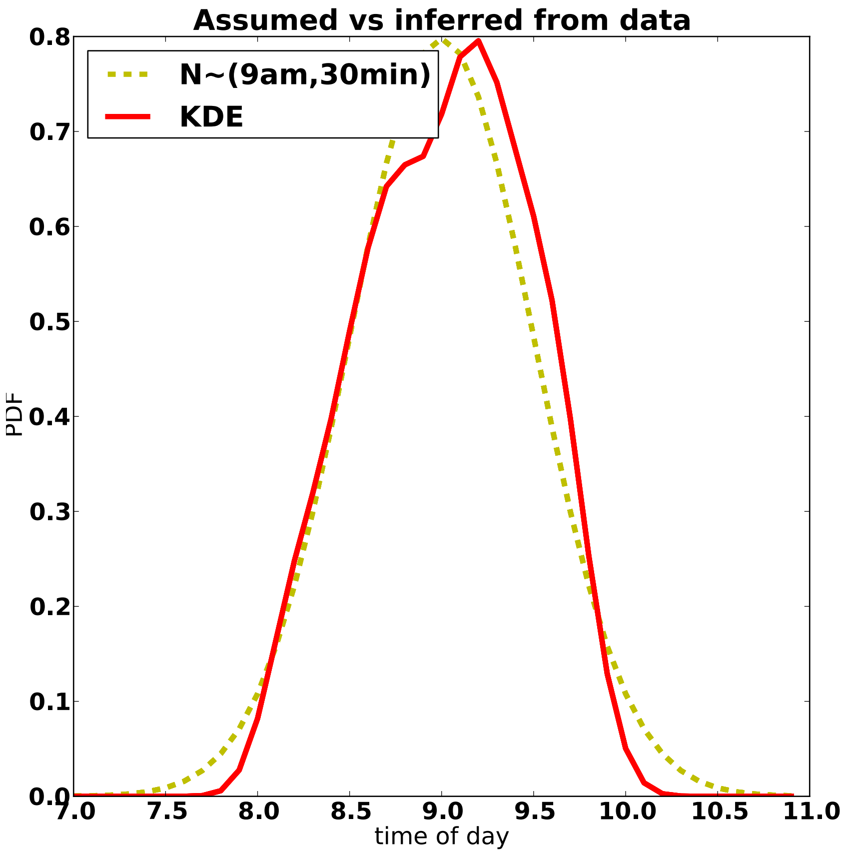
- Spatio-temporal:
  - (lat,lng) coordinates
  - Timestamps
- Social:
  - Twitter Graph
- Details:
  - Nov ‘12 – Feb ‘13
  - New York: 5.20M geo-tweets, 225K users
  - Los Angeles: 3.23M geo-tweets,155K users

Learning from Data 1: Home/Work Locations

- Methodology
  - Based on:
    - S. Isaacman, et. al., “Identifying Important Places in People’s Lives from Cellular Network Data”, Pervasive 2011
  - Ground truth (known home/work):
    - CDRs: Known industrial and residential areas
    - Geo-tweets: Foursquare
  - Train classifiers to identify home/work
- Home and Work locations inferred:
  - Madrid (CDRs): 272,479
  - NY (Twitter): 71,977

- Home and Work distribution is NOT uniform
  - In contrast to related work:
    - H.-S. J. Tsao and D. Lin et al., … 1999.

Learning from Data 2: Departure Times



- Exploit consecutive Home-Work calls
- Home-Work travel
  - Time: Online maps
- Similar for work departure times

Distance Function

$$d(v,u) = \begin{cases} h(v,u) + w(v,u), \\ \text{IF} \quad \max(h(v,u), w(v,u)) \leq \delta \\ \text{AND} \quad \max(|LH(u) - LH(v)|, |LW(u) - LW(v)|) \leq \tau \\ \\ \infty, \quad \text{otherwise} \end{cases}$$

distance tolerance

time tolerance

Problem Formulation

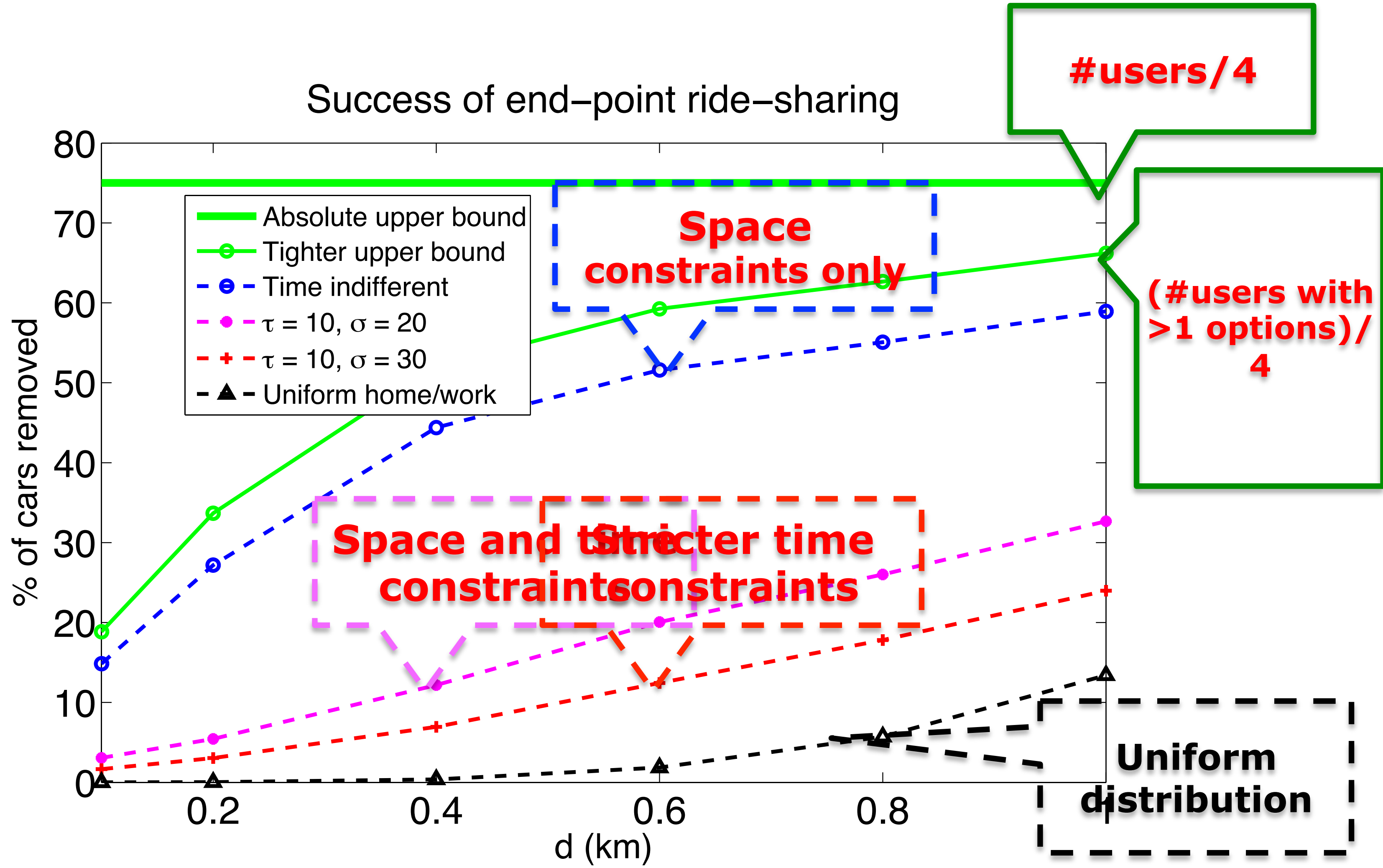
- Capacitated Facility Location with Unsplittable Demands:
  - Users : V
    - Drivers: S $\subseteq$ V
    - Passengers: V – S
  - Capacity: 4 users/car
  - Find:
    - Assignment a: (V – S) → S
    - Minimize:

$$\underbrace{\sum_{u \in V} d(a(u),u)}_{\text{driver-passenger distances}} + \underbrace{\sum_{v \in S} p(v)}_{\text{driver penalties}}$$

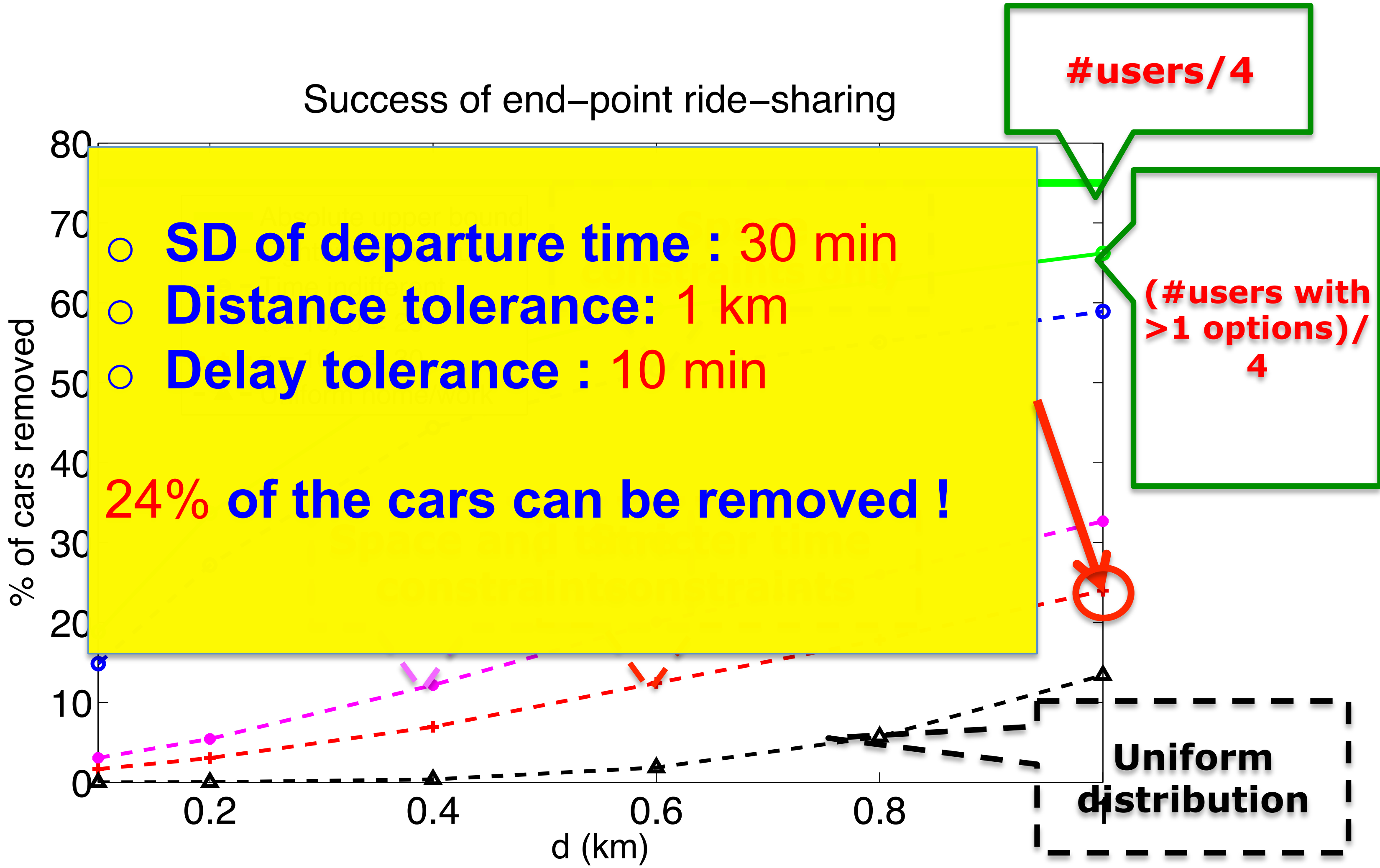
Algorithm: EndPoints RS

- Heuristic solution:
  - Based on:
    - M. R. Korupolu et. al, “Analysis of a local search heuristic for facility location problems,” Journal of Algorithms, 2000.
  - Initial solution:
    - b-matching
  - Iterative improvements
    - Scalability
      - Fixed local search steps
      - Fixed numbers of iterations
  - Polynomial complexity
    - O(nlogn)+O(n) for initial solution
    - O(n) to evaluate solution

EndPoint RS for Madrid-CDRs



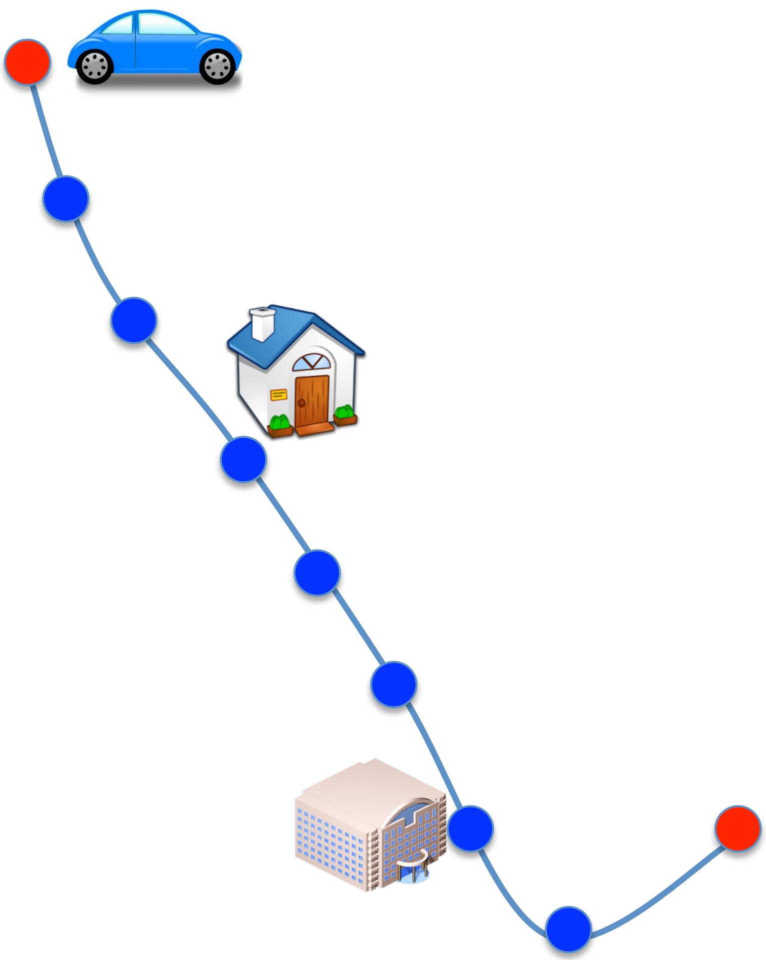
EndPoint RS for Madrid-CDRs



- SD of departure time : 30 min
- Distance tolerance: 1 km
- Delay tolerance : 10 min

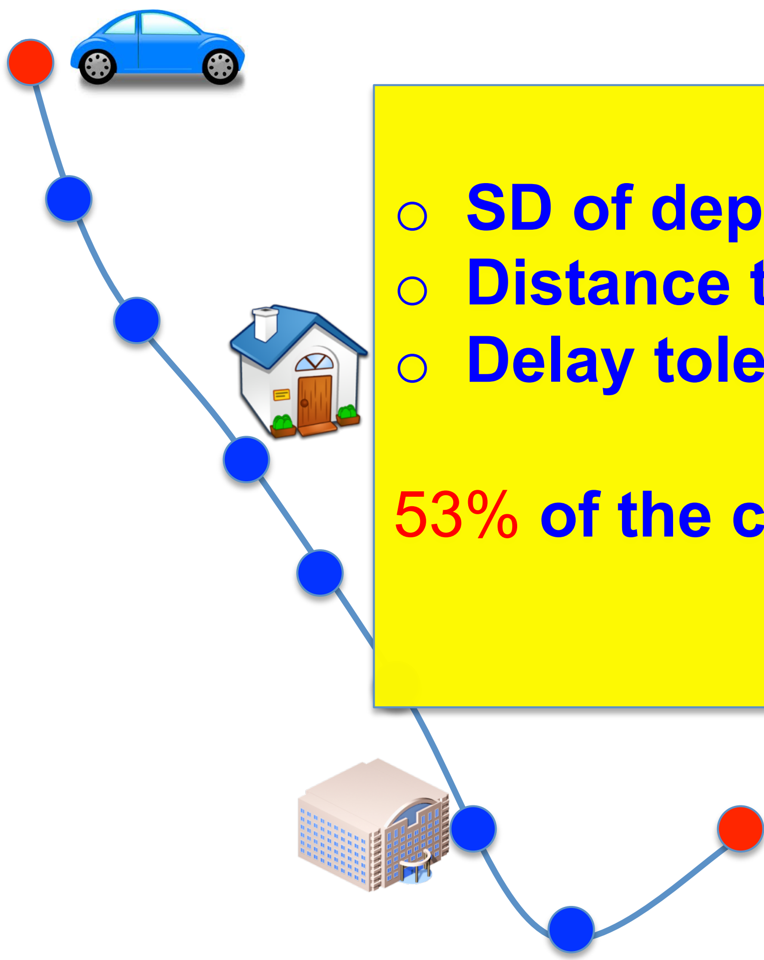
24% of the cars can be removed !

Algorithm: EnRoute RS



- Home/Work paths:
  - Popular Online Maps
- EnRoute RS:
  - Get the solution of EndPoints RS
  - Iterative improvements
  - Fill empty seats by pick-ups
- Spatio-temporal constr. intermediate points:
  - Same and point constraints

Algorithm: EnRoute RS



- Home/Work paths:
  – Popular Online Maps
- EnRoute RS:
  – On the solution of EndPoints RS
  – Improvements
  – Fill empty seats by pick-ups
- Spatio-temporal constr. intermediate points:
  – Same and point constraints

- **SD of departure time : 30 min**
- **Distance tolerance: 1 km**
- **Delay tolerance : 10 min**

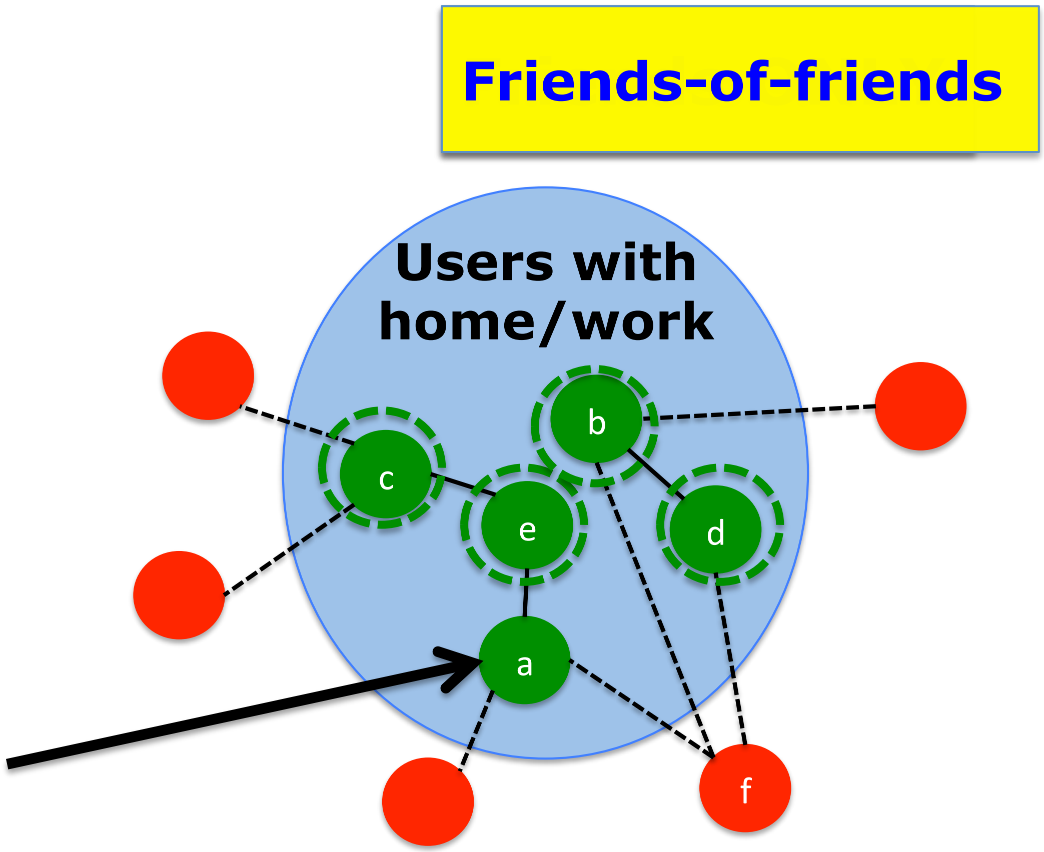
53% of the cars can be removed !

Learning from Data 3: Social Ties

- CDRs graph:
  - Nodes: Users
  - Edges: ≥ 1 call
- Geo-Tweets graph:
  - Nodes: Twitter ids
  - Edges: mutually declared friendship

Social Filtering

- Friends:
  - Graph neighbors
- Sharing rides with:
  - Friends
  - Friend-of-friends



Results

- Ride-sharing parameters:
  - Time distribution: 30 min
  - Distance tolerance : 1 km
  - Delay tolerance : 10 min

City	Friends only	Friends of friends	Anybody
Madrid - CDR	1.1%	19% (31%)	53% (65%)
NY - Tweets	1.2%	8.2% (26%)	44% (68%)

Green numbers show potential of ride-sharing projected to commuters' population.

Conclusion

- High potential based on route overlap:
  - E.g. 53% for Madrid-CDR
- Bottleneck:
  - Willingness to ride-share
  - Riding ONLY with friends is too restrictive
- Technology and building trust:
  - Riding with friends of friends: up to 31% potential.
- Other lessons:
  - Lessons from data sets
  - Spatio-temporal constraints
  - Comparisons between cities

Thank You

Blerim Cici,
Athina Markopoulou

Nikolaos Laoutaris,
Enrique Frias-Martinez