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

Quantifying the potential of ride-sharing using mobile phone data
Strong Car culture

• In US:
  – Commuters: 128.3M
  – drive alone: 75.7%
  – Bike: 0.38%
Commuting

All Urban Areas

2009 Morning

2009 Evening

Red – Almost all regions have congestion
Yellow – Many regions have congestion
Green Checked – Some regions have congestion
Gray – Very few regions have congestion

Quantifying the potential of ride-sharing using mobile phone data
Annual cost of owning 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:

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

Source: http://www.doughroller.net/smart-spending/true-cost-of-a-car-over-its-lifetime/
What is Ride-Sharing?
Ride-Sharing: An old idea

When you ride ALONE
you ride with Hitler!

Join a Car-Sharing Club
TODAY!

HI HO! HI HO!
IT'S OFF TO WORK
WE GO!

HELP WIN THE WAR
Squeeze in one more

6
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. Inflexible
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

• Quantification of Ride-sharing potential
Related Work

• CDR Analysis and Human Dynamics

• Call Description Record Analysis
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
  - \(<\text{lat}, \text{lng}>\) coordinates
  - Text
  - Links (e.g. YouTube)

NY: 5.2M tweets, 225K users
LA: 3.23M tweets, 155K users
Identifying Home/Work

- Small set of users with known Home/Work addresses

**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

- $\sigma$: standard deviation of Home/Work departure times
- $\tau$: time tolerance
Formulation

• Goal: minimize the number of cars given spatial and time constraints

• Capacitated Facility location with Unsplittable Demands:
  – Facilities: Drivers
  – Clients: Passengers

• Distance function:
  – \( d(u,v) = \max\{h_{\text{dist}}(u,v), w_{\text{dist}}(u,v)\} \)
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

Success of end-point ride-sharing

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

26% of the cars can be removed!

- Time indifferent
- $\tau = 10, \sigma = 10$
- $\tau = 10, \sigma = 20$

#users/4

(#users with >1 options)/4

Space constraints

Stricter time constraints

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

Success of en–route ride–sharing

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

<table>
<thead>
<tr>
<th>City</th>
<th>Friends only</th>
<th>Friends of friends</th>
<th>Anybody</th>
</tr>
</thead>
<tbody>
<tr>
<td>Madrid</td>
<td>0.2%</td>
<td>2.4%</td>
<td>47%</td>
</tr>
<tr>
<td>New York</td>
<td>1.5%</td>
<td>9.1%</td>
<td>52%</td>
</tr>
</tbody>
</table>
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

Matching

Preference Finder

Driver Monitoring
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

\[ \begin{align*} 
\text{p1:} & \quad \{\text{p1’s preferences}\} \\
\text{p2:} & \quad \{\text{p2’s preferences}\} \\
\vdots & \quad \vdots \\
\vdots & \quad \vdots \\
\text{pn:} & \quad \{\text{pn’s preferences}\} \\
\end{align*} \]
The matching algorithm:

• Match new request as soon as they arrive

• Refines existing (driver, passenger) pairs every 2 mins.

• Use distance function to model preferences.
Distance function

• $\text{dist}(d, p) =$

  \[
  \text{social\_weight} \times \text{social\_dist}(d, p) + \text{time\_weight} \times \text{time\_dist}(d, p) + \text{load\_balance\_weight} \times \text{empty}(d)
  \]
## Extreme cases

<table>
<thead>
<tr>
<th>Cases</th>
<th>Driver ratio (%)</th>
<th>Passenger ratio (%)</th>
<th>Social sharing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>social best</td>
<td>62</td>
<td>79</td>
<td>6</td>
</tr>
<tr>
<td>time best</td>
<td>48</td>
<td>79</td>
<td>0.4</td>
</tr>
<tr>
<td>load-balancing</td>
<td>73</td>
<td>79</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The diagram on the right shows the cumulative distribution function (CDF) of delay (in minutes) for different cases. The lines represent:
- **best load balancing**
- **best departure delay**
- **best social distance**

Delay (min) on the x-axis ranges from 0 to 10, and CDF on the y-axis ranges from 0 to 1.0.
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

\[d_1: (\text{lat}, \text{lng}) \ldots (\text{lat}, \text{lng})\]
\[d_2: (\text{lat}, \text{lng}) \ldots (\text{lat}, \text{lng})\]
\[
\vdots
\]
\[d_n: (\text{lat}, \text{lng}) \ldots (\text{lat}, \text{lng})\]

\[d_1: \text{loc. estimation}\]
\[
\vdots
\]
\[d_n: \text{loc. estimation}\]
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 of ride-sharing with using CDRs, and geo-tweets.

• Results:
  – The success 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 Лаутарис