Paraisopolis slum mobility diagnosis based on smartphone app, smartcard and mobile phone data

Mariana Giannotti

São Paulo
31st October 2017
poverty and inequality

experience in improving infrastructure in informal settlement contributes to discussion on urbanization strategies in dense low-income communities
12 times denser than the city itself

approximately 70,000 residents

second largest by population
Greater access to credit resulted in an increase of car ownership.

Low-income residents’ main transport mode is walking, but in spite of that, pedestrians are often forced to walk between cars due to the presence of vendors and parked cars on the sidewalk.
To develop a mobility diagnosis and identify the main bottlenecks that the community of Paraisopolis faces in accessing daily activities to support public policies.
Mobility Diagnosis

- Multiday X One day only
- More trips then reported?
- Low pay Jobs
- Hard to reach group
- Walking Trips
- Less expensive survey? Shorten intervals

ESCOLA POLÍTÉCNICA DA UNIVERSIDADE DE SÃO PAULO
Current and foreseen 60 minutes accessibility to Paraisópolis
Smartphone App and Field Survey

Pilot Sample Survey
- 21 residents
- Pen and paper survey

Sample Planning
- Two stages
- Geographical quotas
- Individuals quotas
- Two lists of households for substitutions

Smartphone Application
- FollowMee
- Interval of data collection
- Period for free use
- Background
- Remote access to data

Field survey and data collection
- Target 381 residentes
- Household travel survey and smartphones app
- Collected up to 14 days of GPS data
Hard to reach group

- Historically ill-represented: Census data x Habisp
- Sampling: lack of data (formal address...)
- Target person: schedule respecting organized crime time Windows
- Technology barriers: less than expected
Data collection

To achieve 381: 1,585 visits at 963 households

<table>
<thead>
<tr>
<th>Final survey status</th>
<th>Quantity</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview and GPS data collection</td>
<td>360</td>
<td>37.4%</td>
</tr>
<tr>
<td>Interview only</td>
<td>21</td>
<td>2.2%</td>
</tr>
<tr>
<td>Refused to participate</td>
<td>271</td>
<td>28.1%</td>
</tr>
<tr>
<td>Do not own a smartphone</td>
<td>147</td>
<td>15.3%</td>
</tr>
<tr>
<td>No one at home</td>
<td>92</td>
<td>9.6%</td>
</tr>
<tr>
<td>Address non-existent</td>
<td>72</td>
<td>7.5%</td>
</tr>
</tbody>
</table>

Total households contacted 963 100.0%

<table>
<thead>
<tr>
<th>Attempt</th>
<th>Number of visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st visit</td>
<td>1,052</td>
</tr>
<tr>
<td>2nd visit</td>
<td>368</td>
</tr>
<tr>
<td>3rd visit</td>
<td>165</td>
</tr>
<tr>
<td>Total</td>
<td>1,585</td>
</tr>
</tbody>
</table>
Data collection

<table>
<thead>
<tr>
<th>Final survey status</th>
<th>Quantity</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview and GPS data collection</td>
<td>360</td>
<td>37.4%</td>
</tr>
<tr>
<td>Interview only</td>
<td>21</td>
<td>2.2%</td>
</tr>
<tr>
<td>Refused to participate</td>
<td>271</td>
<td>28.1%</td>
</tr>
<tr>
<td>Do not own a smartphone</td>
<td>147</td>
<td>15.3%</td>
</tr>
<tr>
<td>No one at home</td>
<td>92</td>
<td>9.6%</td>
</tr>
<tr>
<td>Address nonexistent</td>
<td>72</td>
<td>7.5%</td>
</tr>
<tr>
<td><strong>Total households contacted</strong></td>
<td><strong>963</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attempt</th>
<th>Number of visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st visit</td>
<td>1,052</td>
</tr>
<tr>
<td>2nd visit</td>
<td>368</td>
</tr>
<tr>
<td>3rd visit</td>
<td>165</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1,585</strong></td>
</tr>
</tbody>
</table>

28.1% refused
15.1% did not own a smartphone
9.6% not contacted after 3 visits
Data cleaning

- Desired sample
- Problems with the application
- Long time gaps between GPS records (> 30 minutes)
- 7 complete continuous days of GPS data
17 residents
Field survey and data collection

Development of algorithm for data collected treatment

Analysis, discussion and dissemination of results
% Users

Number of days of collected data

2.5 trips/day
71% stays can be accessed within 60 minutes.
Time sensitive effects
Multiday analysis
Main destinations from Paraisópolis residents

Mobile App
Main destinations from Paraisópolis residents

Smart Card

Trip Chaining (Renato Arbex)
Main difficulties in daily transport

- Crowded buses: 45%
- Waiting time: 18%
- Other: 15%
- Distance to the bus stop: 4%
- Disrespect and lack of education: 2%
- Lack of accessibility for People with...: 1%
Mobile phone data

10 ERBs 25th april 2016 until 25th may 2016

• ~ 250 K unique users
  • ~ 10 K frequent users (10 days during period) período)

~30 million CDR 25th april 2016 until 9th august 2016

• ~3,6 K unique users with enough data to infer stay points
Resultados das distâncias médias de casa nos dias úteis

<table>
<thead>
<tr>
<th>Faixa</th>
<th>Hora</th>
<th>Distância em Quilômetros</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
Legenda

- Aglomerados Subnormais

DISTÂNCIA

- Até 500 m
- De 500 m a 2 km
- Acima de 2 km

5 – 18h
20 – 0h

Legenda
- Aglomerados Subnormais

DISTÂNCIA
- Até 500 m
- De 500 m a 2 km
- Acima de 2 km

Quilômetros
Dataset Description

- **Smart Card Dataset**
  - Data provided by SPTrans – São Paulo Transporte S.A.,
  - Raw data:
    - 25th April 2016 until 9th August 2016 (recently, not yet for this presentation)
    - 31 days of analysis between March 1st to March 31st, 2015;
    - ~8M card holders, with ~320M validations during this period.
  - Database main attributes:
    - Card ID (hash_bilhete);
    - Date / Hour of validation;
    - Type of card;
    - Route.

![Bar chart showing frequency of validations by type of card](image)
Cluster 2: Lunch-Afternoon

Cluster 3: Morning-Afternoon

Percentage of users by cluster - Paraisópolis

- Week 1:
  - Morning-Night: 9.7%
  - Lunch-Afternoon: 9.8%
  - Morning-Afternoon: 10.1%
  - Morning-Lunch: 31.8%
  - Lunch-Night: 31.8%

- Week 2:
  - Morning-Night: 11.3%
  - Lunch-Afternoon: 11.1%
  - Morning-Afternoon: 11.1%
  - Morning-Lunch: 34.5%
  - Lunch-Night: 34.3%

- Week 3:
  - Morning-Night: 12.8%
  - Lunch-Afternoon: 13.4%
  - Morning-Afternoon: 12.8%
  - Morning-Lunch: 34.2%
  - Lunch-Night: 35.1%

- Week 4:
  - Morning-Night: 12.7%
  - Lunch-Afternoon: 12.7%
  - Morning-Afternoon: 11.0%
  - Morning-Lunch: 31.8%
  - Lunch-Night: 31.4%
Multiday patterns
Add semantic context (e.g. land use)

• Caio Pieroni (Smart Card)
• Bruna Pizzol (Smartphone App)
• André Leite (Mobile phone)

Activity Spaces

• Rebeca Bianco
Visualising accessibility using an agent-based model

- Greater London Authority (GLA) – not metropolitan area yet
- Main road network from OpenStreetMap
1. Managers, directors and senior officials
2. Professional occupations
3. Associate professional and technical
4. Administrative and secretarial
5. Skilled trades occupations
6. Caring, leisure and other services
7. Sales and customer service
8. Process, plant and machine operatives
9. Elementary occupations
1 diretores e gerentes
2 profissionais das ciências e intelectuais
3 técnicos e profissionais de nível médio
4 trabalhadores de apoio administrativo
5 trabalhadores dos serviços, vendedores
7 trabalhadores qualificados, operários e artesãos
9 ocupações elementares

All groups combined
Mobility x Safety
Final Considerations

- Big data: new methods for sampling strategies?
- Overlapping inequalities
- Big data: big opportunities for research!
Thanks!!!

mariana.giannotti@usp.br