INTELLIGENT TRANSPORTATION SYSTEMS
Lecture 11 (17 Nov 2021)
Human Mobility and Mobile Network Data
Mozhgan Pourmoradnasseri, PhD
We need to know how people move!

- History
- Global Changes
- Local Developments
- Spread of Biological Viruses
- Utility Management
- Traffic Forecasting
- Location-based Services
- Social Networks
- Business

Spatial, Temporal, and Social Aspects.
Example: Study of Cultural History

Reconstructed aggregate intellectual mobility over two millennia through the birth and death locations of more than 150,000 notable individuals.

Data Sources for Studying Human Mobility

- Census Data
- Travel Surveys
- Traffic Sensors
- Bank Notes
- Transit Smart Cards
- Social Networks (Twitter, Flicker, ...)
- GPS Trajectories of Prob Vehicles
- Cellular Network Data
Data Sources for Studying Human Mobility

- Census Data
- Travel Surveys
- Traffic Sensors
- Bank Notes
- Transit Smart Cards
- Social Networks (Twitter, Flicker, …)
- GPS Trajectories of Prob Vehicles
- Cellular Network Data

**Mobile Phones as Sensors**

Most people have mobile phones
It allows massive data collection.
Data Sources for Studying Human Mobility

- Census Data
- Travel Surveys
- Traffic Sensors
- Bank Notes
- Transit Smart Cards
- Social Networks (Twitter, Flicker, ...)
- GPS Trajectories of Prob Vehicles
- Cellular Network Data

Most people have mobile phones.
It allows massive data collection.

Difficult to access.
Mobile Network Data

1. From the mobile device
   - Collecting information from developed APIs.
   - Limited sample size, based on volunteers.

2. From network
   - Covers almost all populations.
   - Data is already available in the network provides' databases, for billing purposes.
   - Difficult to access, due to privacy concerns and business secrets.
Mobile Network Data

1. From the mobile device
   - Collecting information from developed APIs.
   - Limited sample, based on volunteers.

2. From network
   - Covers almost all populations.
   - Data is already available in the network provides' databases, for billing purposes.
   - Difficult to access, due to privacy concerns and business secrets.
Mobile Phone Network Data
(cellular network data)

- **Events-driven mobile phone network data**
  - Data is generated by the interaction between the device and the network.
- **Network-driven mobile phone network data**
  - Data generated by the interaction between the network and the device
Identifiers of a Mobile Subscriber

- **IMEI**: (International Mobile Equipment Identity): Usually unique number to identify the mobile phone. GSM networks use the IMEI number to identify valid devices.
- **IMSI**: (International mobile subscriber identity): Uniquely identifies every user of a cellular network, stored as a 64-bit field, and is sent by the mobile device to the network.
- **MSISDN**: The actual phone number.
The network can be divided into cells, services and connected to each other by a collection of transceivers, controllers, switches, routers, and registers.

- **MS (Mobile Station)**: The physical equipment used for connecting to the network; Mobile Phone+SIM
- **BTS (Base Transceiver Station)**: The antenna installed on top of the tower. Mobile Phone's access point to the network and responsible for carrying out radio communications between the network and the mobile phone.
- **BSC (Base Station Controller)**: controls multiple BTSs. Handles allocation of radio channels, frequency administration, power and signal measurements from the MS, and handovers from one BTS to another.
- **MSC (Mobile Switching Center)**: The heart of the GSM network. It handles call routing, call setup, and basic switching functions.

Mobile phones are radio transceivers: transmitter + receiver connected to each other via cell towers.

Source: https://www.coai.com
The target area is chopped into small cells such that each cell is covered by a cell tower.

A cell-plan consisting of hexagon shaped cells
Cell Plan

The target area is chopped into small cells such that each cell is covered by a cell tower.

Advantage:

1. Cell phones only require low-power transmitters. So, small batteries and lightweight devices.
2. Communication frequencies can be re-used without disturbance of nearby cell towers.
The target area is chopped into small cells such that each cell is covered by a cell tower.

**Advantage:**

1. Cell phones only require low-power transmitters. So, small batteries and lightweight devices.
2. Communication frequencies can be re-used without disturbance of nearby cell towers.

**Math exercise :)**

How many colors are needed to color the regions of any map so that no two adjacent regions have the same color?
Cell Plan

The target area is chopped into small cells such that each cell is covered by a cell tower.

Math exercise :)

How many colors are needed to color the regions of any map so that no two adjacent regions have the same color?

Advantage:

1. Cell phones only require low-power transmitters. So, small batteries and lightweight devices.
2. Communication frequencies can be re-used without disturbance of nearby cell towers.
The target area is chopped into small cells such that each cell is covered by a cell tower.

Advantage:

1. Cell phones only require low-power transmitters. So, small batteries and lightweight devices.
2. Communication frequencies can be re-used without disturbance of nearby cell towers.

Math exercise :) How many colors are needed to color the regions of any map so that no two adjacent regions have the same color?

Credit: Martijn Tennekes (Mobile Phone technology)
Base Transceiver Stations (BTS)

Cell tower
- 3 cells, each covering 120°
- Coverage up to 40 km

Rooftop cell site
- Coverage up to 40 km

Small cell
- Coverage up to 2 km

Indoor cell
- Coverage up to 200 m

Credit: Martijn Tennekes (Mobile Phone technology)
Cell Plan

More realistically:

- Cells overlap each other.
- MS is not always connected to the nearest antenna.
Cell Plan

More realistically:

- Cells overlap each other.
- MS is not always connected to the nearest antenna.
Cell Plan

More realistically:

- Cells overlap each other.
- MS is not always connected to the nearest antenna.

Credit: Martijn Tennekes (Mobile Phone technology)
Cellular Network Data

- Tabular data format
- Each data record (row) corresponds to an event, i.e. active mobile phone use, or triggered by the network
- Main attributes:
  - event type (call, data, etc)
  - timestamp of the event
  - IMSI
  - IMEI
  - cell-ID of the event

Active user: Call Detail Records (CDR)

Triggered by network: Visitor Location Register (VLR)
Cellular Network Data

- Tabular data format
- Each data record (row) corresponds to an event, i.e. active mobile phone use, or triggered by the network
- Main attributes:
  - event type (call, data, etc)
  - timestamp of the event
  - IMSI
  - IMEI
  - cell-ID of the event

Active user: Call Detail Records (CDR)

Triggered by network: Visitor Location Register (VLR)

How the Data Looks Like

<table>
<thead>
<tr>
<th>Anonymized user ID</th>
<th>cell-ID</th>
<th>Timestamp</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>G3Z03R8269</td>
<td>W8575</td>
<td>1,525,757,266</td>
<td>par-c</td>
</tr>
<tr>
<td>G2H99K9882</td>
<td>F3268</td>
<td>1,531,421,050</td>
<td>detach-c</td>
</tr>
<tr>
<td>G8J84W8462</td>
<td>I7520</td>
<td>1,526,482,568</td>
<td>loc-up-c</td>
</tr>
</tbody>
</table>
We need the cell plan to merge with spatial aspects.

An example of CDR and VLR dataset attributes.

<table>
<thead>
<tr>
<th>Anonymized user ID</th>
<th>cell-ID</th>
<th>Timestamp</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>G3Z03R8269</td>
<td>W8575</td>
<td>1,525,757,266</td>
<td>par-c</td>
</tr>
<tr>
<td>G2H99K9882</td>
<td>F3268</td>
<td>1,531,421,050</td>
<td>detach-c</td>
</tr>
<tr>
<td>G8J84W8462</td>
<td>I7520</td>
<td>1,526,482,568</td>
<td>loc-up-c</td>
</tr>
</tbody>
</table>

Header =
```
{ "CGI", "Coverage Area", "Tower ID Longitude", "Tower ID Latitude", "Site ID" }
```
Record =
```
{ "MCC-MNC-LAC-CI", "POLYGON (((x1,y1),(x2,y2), ...,(xn,yn)))", "X.xxxx", "Y.yyyy", "Site-N" }
```
Mobility Extraction: First Step

User-ID and Event + Cell-ID = Cell-based Trajectory

Lots of filtering and data cleaning is needed for removing ping-pong effects, etc.
Spatial Perspective: Reality vs Data
Issues of Cellular Network Data

- The goal has been network connectivity and management, not mobility analysis.
- The data is already available, without adjustment for research purposes.
- Hard to access.

Cellular data Issues:
- Sparsity in time
- Low spatial resolution
- Localization errors due to ping-pong effect and load balancing

Cell-plan issues:
- Polygons do not represent the reality
- Fluctuations, due to weather, technical reasons, etc.
- Business secrets
Overcoming the Challenges

- Some problems don't have solutions :
- Data filtering and cleaning methods.
- The data with the current resolution is still suitable for large-scale studies.
- Data enrichment methods for filling the temporal gaps in the data.
- Methods for refining position accuracy and moving towards GPS-like trajectories.
- More realistic models for cell coverage areas.
Overcoming the Challenges

- Some problems don't have solutions :(
- Data filtering and cleaning methods.
- The data with the current resolution is still suitable for large-scale studies.
- Data enrichment methods for filling the temporal gaps in the data.
- Methods for refining position accuracy and moving towards GPS-like trajectories.
- More realistic models for cell coverage areas.

Red dots indicate the recorded GPS locations and the circles are the associated CDR-based trajectory. Red circles indicate the cells that disappeared from the trajectory due to filtering the events and black circles are the ones remaining after filtering.

Overcoming the Challenges

- Some problems don't have solutions :(  
- Data filtering and cleaning methods.  
- The data with the current resolution is still suitable for large-scale studies.  
- Data enrichment methods for filling the temporal gaps in the data.  
- Methods for refining position accuracy and moving towards GPS-like trajectories.  
- More realistic models for cell coverage areas.

https://tarktartu.telia.ee/en-GB/people
Overcoming the Challenges

- Some problems don't have solutions :(
- Data filtering and cleaning methods.
- The data with the current resolution is still suitable for large-scale studies.
- Data enrichment methods for filling the temporal gaps in the data.
- Methods for refining position accuracy and moving towards GPS-like trajectories.
- More realistic models for cell coverage areas.


A reconstructed trajectory of a random user. Each circle represents a coverage area associated with a BTS. Red circles are observed by the mobile network, and the blue circles are the hidden parts of the trajectory discovered by our algorithm.
Overcoming the Challenges

- Some problems don't have solutions :(
- Data filtering and cleaning methods.
- The data with the current resolution is still suitable for large-scale studies.
- Data enrichment methods for filling the temporal gaps in the data.
- **Methods for refining position accuracy and moving towards GPS-like trajectories.**
- More realistic models for cell coverage areas.


![Map showing GPS trace and cellular trajectory](image)
Overcoming the Challenges

- Some problems don't have solutions 😞
- Data filtering and cleaning methods.
- The data with the current resolution is still suitable for large-scale studies.
- Data enrichment methods for filling the temporal gaps in the data.
- Methods for refining position accuracy and moving towards GPS-like trajectories.
- More realistic models for cell coverage areas.

Coverage area approximation of single cell represented by bold-line polygon; approximated to circle (a) and sector (b); and corresponding normal distribution single (c) and n-components (d)

Application Areas and Case Studies
Daily Mobility Pattern Mechanisms

Only 17 unique patterns are present in daily mobility and they follow simple rules. These patterns are sufficient to capture up to 90% of the population.

Decomposition of the mobility profile over 10 days into daily mobility patterns for two anonymous mobile phone users.

Daily Mobility Pattern Mechanisms

Only 17 unique patterns are present in daily mobility and they follow simple rules. These patterns are sufficient to capture up to 90% of the population.

Daily Mobility Pattern Mechanisms

Only 17 unique patterns are present in daily mobility and they follow simple rules. These patterns are sufficient to capture up to 90% of the population.

Daily human mobility patterns are stable over several months. The values show how more or less likely a motif is found during the observation period of six months under the condition that the individual has a given motif on another day. Positive values indicate that these motifs are more likely than expected and negative values that these motifs are suppressed.

Discovering Points of Interest

Distribution/Frequency of visited towers in a day:

- Home/work location (individual perspective) with the assumption of a normal pattern
- High-resolution perspective
Discovering Points of Interest

Distribution/Frequency of visited towers in a day:

- Home/work location (individual perspective) with the assumption of a normal pattern
- High-resolution perspective

Discovering Points of Interest

Distribution/Frequency of visited towers in a day:

- Home/work location (individual perspective) with the assumption of a normal pattern
- High-resolution perspective

Understanding Social Challenges

- Data for Development Challenge
- Data for Refugees Challenge
- Gender Gap in Urban Mobility
- Segregation Studies

Cellular Data and COVID

This paper presents a comparative analysis between mobility and the infection reproduction number $R_t$ over time.