Social and Technological Network Analysis: Spatial Networks, Mobility and Applications

Anastasios Noulas
Data Science Institute
School of Computing & Communications
Lancaster University
History of modern human mobility studies

Ernst Georg Ravenstein

What is he even talkin’ about?

William Farr ... or
Dark Vader

Human migration follows no definitive law ...
The laws of human migration

The following was a standard list after Ravenstein's (1834-1913) proposal in the 1880s. The theories are as follows:

1. *every migration flow generates a return or countermigration.*
2. *the majority of migrants move a short distance.*
3. *migrants who move longer distances tend to choose big-city destinations.*
4. *urban residents are often less migratory than inhabitants of rural areas.*
5. *families are less likely to make international moves than young adults.*
6. *most migrants are adults.*
7. *large towns grow by migration rather than natural increase.*

Ravenstein exploited census data from the United Kingdom to support empirically his findings ...

Gravity Models

Inspired by Newtonian physics, gravity models suggest that two places attraction is proportional to their mass and inversely proportional to their geographic distance.

\[ F_{ij} = \gamma \frac{m_i m_j}{d_{ij}^2} \]
Data in urban transport modeling has been based primarily on surveys...

Cellular Datasets

One of the first large scale studies of human movement with modern mobile datasets...

Where’s George?

One of the most creative ways to study human movement that has used the displacement of dollar bills as a proxy to human mobility...

Mobile users are the stars

https://foursquare.com/infographics/500million
Dataset Statistics

925,030 users around the globe over a period of 6 months in 2010.

34 Cities that span 4 continents and 11 countries.

For the first time human mobility is analyzed in light of 5 million recorded settlements (places).
Power-law tales ... 

*Mobile Social Network Data*

\[(\Delta r + \Delta r_0)^{-\beta}\]

exponent $\beta = 1.50$

*Nature 453, 779-782 (5 June 2008)*

\[(\Delta r + \Delta r_0)^{-\beta}e^{-\Delta r/k}\]

exponent $\beta = 1.75$
Urban vs Global mobility

Power law kicks in at 18.42km!!!
Samuel A. Stouffer

Stouffer's law of intervening opportunities states, "The number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities." *

- Empirically proven using data for migrating families in the city of Cleveland.

- We investigate the plausibility of the theory for urban movements in Foursquare.

* S. Stouffer (1940) Intervening opportunities: A theory relating mobility and distance, American Sociological Review 5, 845-867
Samuel A. Stouffer was a big data pioneer!
- Stouffer's Theory of Intervening Opportunities motivated us to inspect the impact of places (=opportunities) in human mobility.

- Place density by far more important than the city area size with respect to mean length of human movements ($R^2 = 0.59$ and 0.19 respectively).
Defining Rank-Distance

\[ \text{rank}_u(v) = \left| \{w : d(u, w) < d(u, v)\} \right| \]
The rank of all cities collapse to a single line.

We have measured a power law exponent $\alpha = 0.84 \pm 0.07$
A new model for urban mobility

soil...

and mind!

\[
Pr[u \rightarrow v] \propto \frac{1}{rank_u(v)^a}
\]
Set ... and go!

for all cities we have used the average value $\alpha = 0.84$ for the rank exponent.

all places in the city used as potential starting points for our agents.

the rank element is universal, only the set of places differs from city to city.
Simulation Results ...
“Zero” Gravity

We have also built a gravity model in the urban context!

\[ P_g [u \rightarrow v] \propto \frac{m_u \cdot m_v}{d(u, v)^b} \]

Issue #1: how do we define “mass” in the urban context.

Issue #2: how do we set its parameters?
Rank vs Gravity

Rank is simpler and achieves better quality fits for all cities.

Gravity overestimates short transitions ...
Heterogeneities observed in human mobility is due to geographic variations. Cultural, organisational or other factors do not appear to play a role in urban movements.

The rank model, although simple, can cope with the complex spatial variations in densities observed in urban environments.
Computer Science at your Service

User Specific features
- historic visits
- friend check-ins
- preferred venue types

Geographic
- distance and rank-distance
  - 0.7km
  - 1.2km

Temporal
- trending places (hour/day)
- trending place types (eg. cinema at nights)

Place Network
Feature Performance #2

![Graphs showing feature performance for various categories and days, with thumbs up and thumbs down indicating success or failure. The graphs display average rank of visited venue with 1: best and 0: bad.]
Supervised Training: teaching the good and the bad!

**Key Insight:** Every time little Amy checks-in she expresses a direct preference at a place and implicitly ignores all the rest!

**learning:** supervised classifier trains on millions of check-ins generated by populations of users.

[visited place] +1

[randomly selected NOT-visited place] -1
human mobility

network form

Link Probability

Distance

Popularity

Common Neighbors

\[ \sum_{z \in \Gamma_i \cap \Gamma_j} \frac{1}{\log(|\Gamma_z|)} \]
static gravity

\[
\frac{C_i C_j}{d(i, j)^\beta}
\]

dynamic gravity

\[
\frac{1}{T} \sum_{\tau=1}^{T} \frac{c_i(\tau)^+ c_j(\tau)^-}{d(i, j)^\beta}
\]

**awesome fact:**
when \( T=1 \) and \( a_{ij} = 1 \) we fall back to the static gravity model.
Connecting the Fractal City.

http://zeta.math.utsa.edu/~yxk833/connecting.html

http://urbagram.stdio-london.com/v1/show/Network
https://nextcity.org/daily/entry/city-street-grid-maps-visualize-density
Extracting land patches

A typology of street patterns
R Louf, M Barthelemy
Journal of The Royal Society Interface 11 (101), 20140924
Tracking the birth of places [London 2010-2014]

existing places

less than expected

new places

more than expected

Stratford: Olympic Village

\[ n_{i,j}^{\text{null}} = \frac{n_{i,j}^{\text{existing}}}{n_{i,j}^{\text{existing}}} \]

\[ v_{i,j} = n_{i,j}^{\text{null}} - n_{i,j}^{\text{new}} \]
A typology of street patterns
R Louf, M Barthelemy
Journal of The Royal Society Interface 11 (101), 20140924
<table>
<thead>
<tr>
<th>no.</th>
<th>MPD (km)</th>
<th>cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10996</td>
<td>Dubai, Borough of Queens</td>
</tr>
<tr>
<td>2</td>
<td>5800</td>
<td>Athens, Brooklyn, Bucharest, Portland, Sofia</td>
</tr>
<tr>
<td>3</td>
<td>5250</td>
<td>Belo Horizonte, Coyoacán, Curitiba, Fortaleza, Gent, Manaus, Porto Alegre</td>
</tr>
<tr>
<td>4</td>
<td>4924</td>
<td>Adana, Ankara, Bursa, Denizli, Eskişehir, İstanbul, İzmir, Lima, Santiago, Trabzon</td>
</tr>
<tr>
<td>5</td>
<td>5887</td>
<td>Charlotte, Chiba, Columbus, Houston, Indianapolis, Jacksonville, Kiev, Moscow, Nashville, Orlando, Osaka, Phoenix, Raleigh, Saint Petersburg, San Antonio, San Jose, Yokohama</td>
</tr>
<tr>
<td>6</td>
<td>3537</td>
<td>Bandung, Bangkok, Chiang Mai, George Town, Hong Kong, Jakarta, Kuala Lumpur, Makati City, Medan, Petaling Jaya, Pineda, Quezon City, Seoul, Shah Alam, Singapore, Surabaya, Tokyo, Toronto, Yogyakarta</td>
</tr>
<tr>
<td>8</td>
<td>4276</td>
<td>Antwerpen, Atlanta, Austin, Brasília, Chicago, Dallas, Denver, Las Vegas, Los Angeles, Mexico City, Milwaukee, Minneapolis, New York, Philadelphia, Rio de Janeiro, San Diego, San Francisco, São Paulo, Seattle, Washington DC</td>
</tr>
</tbody>
</table>
The “gig” economy

Mobile web and digital mapping technologies have brought a revolution on private resource utilization....

Rent out your extra space to travelers or use your car to drive them around... this is just the beginning of a big revolution that brings together the physical and digital space.
TLC Trip Record Data

The yellow and green taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The data used in the attached datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxi & Livery Passenger Enhancement Programs (TPEP/LPEP). The trip data was not created by the TLC, and TLC makes no representations as to the accuracy of these data.

The For-Hire Vehicle ("FHV") trip records include fields capturing the dispatching base license number and the pick-up date, time, and taxi zone location ID (shape file below). These records are generated from the FHV Trip Record submissions made by bases. Note: The TLC publishes base trip record data as submitted by the bases, and we cannot guarantee or confirm their accuracy or completeness. Therefore, this may not represent the total amount of trips dispatched by all TLC-licensed bases. The TLC performs routine reviews of the records and takes enforcement actions when necessary to ensure, to the extent possible, complete and accurate information.

For trip record data including TLC taxi zone location IDs, location names and corresponding boroughs for each ID can be found here. A shapefile containing the boundaries for the taxi zones can be found here.

Trip Sheet Data (CSV Format)

<table>
<thead>
<tr>
<th>Year</th>
<th>Yellow</th>
<th>Green</th>
<th>FHV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>January</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>February</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>April</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>July</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>August</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>October</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>November</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>December</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
THE NEW YORK CITY TAXI DATASET

FOILing NYC’s Taxi Trip Data

Freedom of Information Law

2013 Trip Data, 11GB, zipped!

2013 Fare Data, 7.7GB

Idea: Uber Vs Yellow Taxi Price Comparison.
HOW THESE PRICES COMPARE

Williamsburg to East Village

- uberX: $15
- old uberX: $19
- taxi: $16

KEEP IN MIND

These prices are only in effect for a limited time. The more you ride, the more likely we can keep them this low!

We know you may be asking yourself how this affects our partner drivers. What we’ve seen in cities across the country is that lower fares mean greater demand, lower pickup times and more trips per hour - increasing earning potential and creating better economics for drivers. What does that mean in the long run? They’ll be making more than ever!
THE EXPERIMENT

1. For every trip in NYC taxi dataset.
2. Record origin & destination coordinates.
3. Retrieve total fare paid.
4. Query Uber API price for the same trip.
5. Compare yellow taxi VS uber prices.
Figure 2: Distribution of prices per journey for Uber X and Yellow Taxis in New York City.
uber more expensive for short trips
Most taxi movements are within a short distance range with longer movements occurring less frequently in the data.
Figure 9: User Query Frequency in terms of weekly temporal evolution patterns.
> 5K only in New York (~8K downloaded the app)

> 14000 search queries generated
Total Money Saved: $64,468.18
Number of Queries: 9,372
Save on avg per trip: $7.0
In ~8 months
Urban Complexity & Performance

- Uber faster
- Black Cab faster
- Journey duration tie

\[
\text{Trip Density} = \frac{1}{|T|} \sum_{i=1}^{|T|} P(x = \text{lng}_i, y = \text{lat}_i, r = 200m) \frac{\pi r^2}{\pi r^2}
\]

![Graph 1: Price Difference vs. Time Difference](image)

![Graph 2: Fraction of Black Cab Wins vs. Average Place Density](image)
Drivers: Black Cab vs Yellow vs Uber

uses his (big) brain

Does not know where is Brooklyn!

blindly follows the GPS

Thanks!

Questions?

email: a.noulas@lancaster.ac.uk