Human Mobility Models in Location-based Social Networks

Anastasios Noulas
Computer Laboratory, University of Cambridge

@ University of Namur, March 2014
Today ....

we will learn about 
**human** urban mobility.

we will learn and assess various **models** in this setting

what’s the affect of **geography** in human movement?

Monday, 24 March 14
Why human mobility?

Urban planning: understand the city and optimise services

Mobile applications and recommendations: study the user and offer services
Why human mobility

Animals have the capacity to search and navigate across space

We can understand how our brain is wired

Have we really left the monkey?
History of modern human mobility studies

Ernst Georg Ravenstein

WTF is this dude talkin’ about?

Human migration follows no definitive law ...

William Farr ... or Dark Vader

the main man ...
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.

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

\[
T_{i,j} = k \frac{O_i D_j}{d_{i,j}^2}
\]

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...

Data on human movement...

Mobile Social VS Cellular

GPS accuracy ~ 10 meters
Global Coverage
Publicly Available

BTS Tower Accuracy ~ 1KM
Country Coverage
Private / Corporate
Power-law tales ...

**Mobile Social Network Data**

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

exponent \( \beta = 1.50 \)

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

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

exponent \( \beta = 1.75 \)
Power-law tales ...

Mobile Social Network Data

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

exponent \( \beta = 1.50 \)

Nature 453, 779-782(5 June 2008)

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

exponent \( \beta = 1.75 \)
Power-law tales ...

Mobile Social Network Data

Nature 453, 779-782 (5 June 2008)

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

exponent \( \beta = 1.50 \)

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

exponent \( \beta = 1.75 \)
Mobile users are the stars

https://foursquare.com/infographics/500million
The Data Crawling Combo ...
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).
Urban vs Global mobility

PDF

urban

10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 

10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 

Distance [km]

10^{-2} 10^{-1} 10^{0} 10^{1} 10^{2} 

Distance [km]

HOUSTON

SAN FRANCISCO

SINGAPORE

Monday, 24 March 14
Urban vs Global mobility

PDF

Distance [km]

PDF

Distance [km]

urban

global

Data

(Δr + Δr₀)⁻β

Monday, 24 March 14
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

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

We have measured a power law exponent $\alpha = 0.84 \pm 0.07$
The rank of all cities collapse to a single line.

We have measured a power law exponent $\alpha = 0.84 \pm 0.07$
Rank universality

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!

\[ Pr[u \rightarrow v] \propto \frac{1}{\text{rank}_u(v)^a} \]
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 is simpler and achieves better quality fits for all cities.

Gravity overestimates short transitions ...
The importance of Geography

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.
Shuffling Urban Geography
Is this the ultimate truth about urban mobility?

- No. We had not explored all possible parameterizations of the gravity model. The two models (rank and gravity) have been shown to be statistically equivalent under certain assumptions: A Wilson. A statistical theory of spatial distribution models. 1967.


- The latter has been criticised due to its “hidden” relationships with the intervening opportunities model. When applied in the city saying it has failed to capture the statistical properties of movements there. [http://perso.uclouvain.be/vincent.blondel/netmob/2013/NetMob2013-abstracts.pdf](http://perso.uclouvain.be/vincent.blondel/netmob/2013/NetMob2013-abstracts.pdf) page 16

- See Occam’s Razor wikipedia article to gain a philosophical perspective on modeling.
Questions and Answers

Email: Anastasios.Noulas@cl.cam.ac.uk
Exercise:

Plot the probability density function of the distances traveled by Foursquare users in Brussels.
Exercise:

If user were moving in the city picking their next place in a uniformly number fashion, how would the probability density function of distances would look like?

Compare this with the previous distribution which described the movements of users in Brussels empirically.
Exercise:

Pick a random sample of 3000 places in Brussels. Then calculate a place distances matrix of dimensionality 3000x3000
Exercise:

Convert the place distances matrix to a place rank distance matrix.

hint: do import scipy.stats as ss and use ss.rankdata
Exercise:

For each movement in Brussels, calculate the rank distance. Then use bin_data (or the histogram of your choice) to plot the rank distances distribution of movements.

Can you try to fit a power-law on these data?
Rank Model (0): Initialise utility functions

```python
MIN_D = -2.0
MAX_D = 2.0

def scale_distance(s,N):
    l = math.log10(s)
    i = (l-MIN_D)/(MAX_D-MIN_D)
    v = int((N-1)*i)
    return v


def compute_probability_power_law(distance,alphaExp):
    alpha = alphaExp
    return distance**(-alpha)
```
def get_place_rank_probabilities_matrix(listOfPlaces, placeDistancesMatrix, compute_probability, powerLawExp):
    listLength = len(listOfPlaces)
    placeTransitionProbabilitiesMatrix = numpy.zeros((listLength, listLength))

    #converting distances to rank-distances
    placesRankedMatrix = numpy.zeros((listLength, listLength))
    for i in xrange(0, listLength):
        placesRankedMatrix[i] = ss.rankdata(placeDistancesMatrix[i])

    #for every place rank all other places according to distance
    for i in xrange(0, listLength):
        sum_rij = 0.0
        for j in xrange(0, listLength):
            if i == j:
                continue
            sum_rij += compute_probability(placesRankedMatrix[i,j], powerLawExp)

        for j in xrange(0, listLength):
            if i == j:
                continue
            p_rij = compute_probability(placesRankedMatrix[i,j], powerLawExp) / (sum_rij*1.0)

            placeTransitionProbabilitiesMatrix[i,j] = p_rij

    ### returns matrix P_rij that corresponds to transition probability to location j, when at location i.
    return placeTransitionProbabilitiesMatrix
Rank Model (2): Compute transition probabilities from a place \( i \) to a distance \( k \).

```python
def get_place_distance_transitions_probabilities(listOfPlaces, placeDistancesMatrix, placeTransitionProbabilitiesMatrix):
    listLength = len(listOfPlaces)
    numOfBins = 100 + 1

    placeDistanceProbabilitiesMatrix = numpy.zeros((listLength, numOfBins))

    for i in xrange(0, listLength):
        for j in xrange(0, listLength):
            if i == j:
                continue
            d_ij = placeDistancesMatrix[i, j]
            binNumber = scale_distance(d_ij, numOfBins)
            placeDistanceProbabilitiesMatrix[i, binNumber] += placeTransitionProbabilitiesMatrix

    # return matrix \( Z_{ik} \) that corresponds to transition from location \( i \) to a distance
    return placeDistanceProbabilitiesMatrix
```
Rank Model (3): Average the results from step 2 and reduce the output to a histogram that describes frequencies at a distance $k$.

def get_general_place_distance_transition_matrix(locationLinks, placeDistanceProbabilitiesMatrix):
    listLength = len(locationLinks)
    numOfBins = 100+1
    generalPlaceDistanceProbabilitiesMatrix = numpy.zeros((1, numOfBins))

    for i in xrange(0, listLength):
        for j in xrange(0, numOfBins):  
            generalPlaceDistanceProbabilitiesMatrix[0, j] += placeDistanceProbabilitiesMatrix[i, j]

    generalPlaceDistanceProbabilitiesMatrix = generalPlaceDistanceProbabilitiesMatrix/(listLength*1.0)

    return generalPlaceDistanceProbabilitiesMatrix
Exercise for home:

i) Plot the rank-model’s output against the empirically observed data and the case of random uniform movements.

ii) Give a try to a gravity model and compare the results to the above.