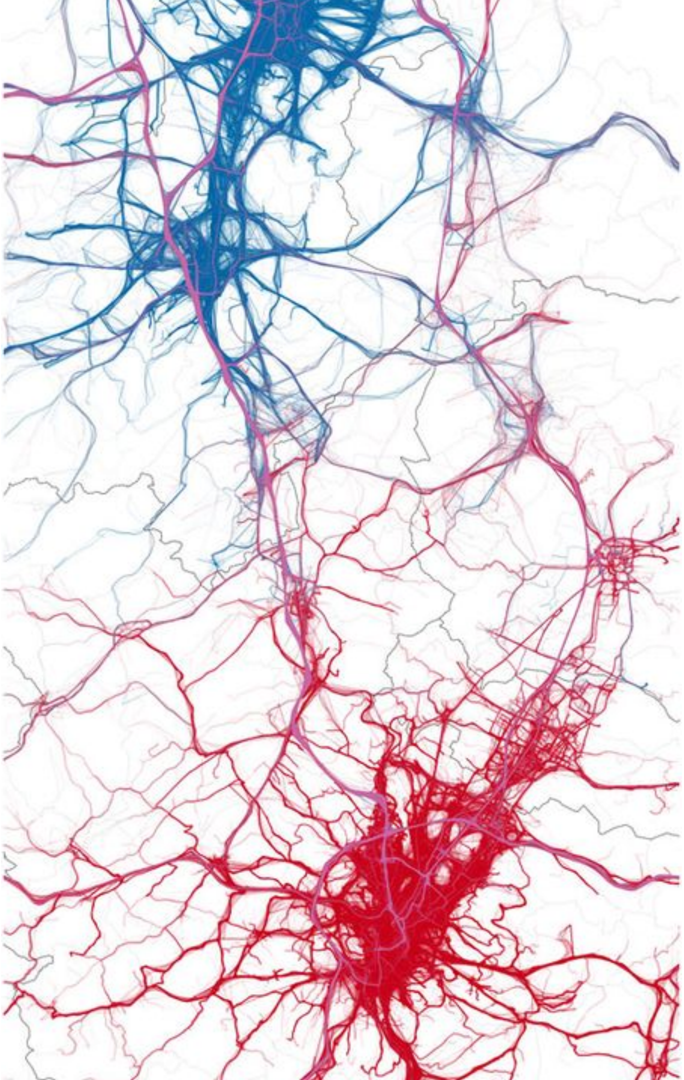


Consiglio Nazionale
delle Ricerche

Collective mobility laws and models

Predictive vs Generative

- **predictive models**
  predict future trips/flows given past history of individuals
  - *machine learning, deep learning*

- **generative models**
  generate synthetic trajs or flows with realistic mobility patterns
  - *mechanistic modelling, machine learning, deep learning*

Individual vs Collective

- **individual models**
  generate/predict the trajectory of a single agent
  - *EPR and its variants*

- **collective models**
  generate/predict flows between locations
  - *Gravity, Radiation, Deep Gravity*

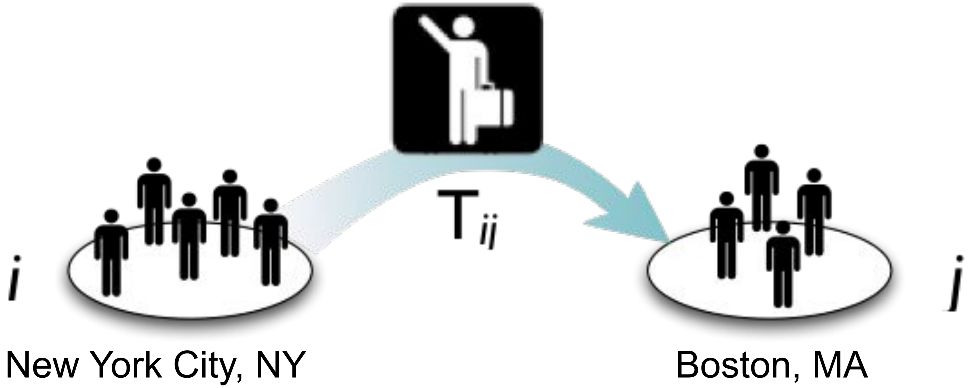
Collective models

generate mobility flows between origins and destinations

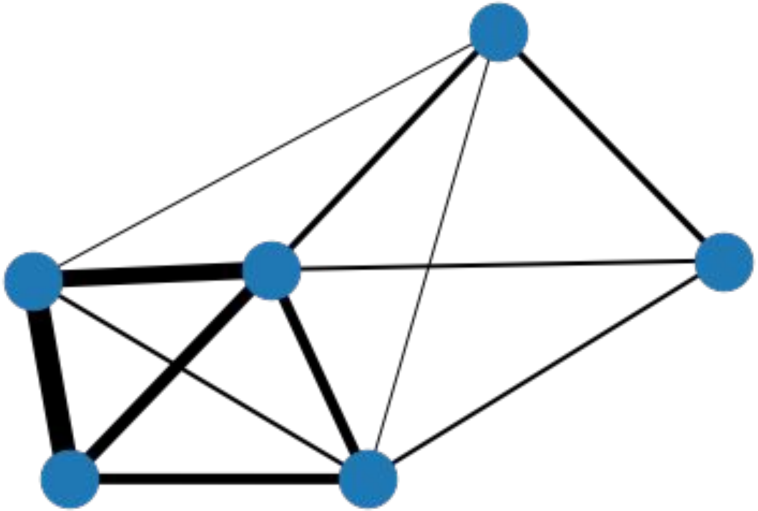
Spatial flows

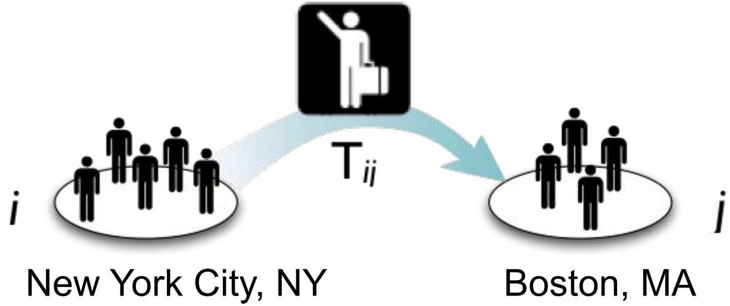
are mathematically represented as an OD matrix T

1. Define locations discretizing space (tessellation)
   e.g., counties, municipalities
2. T_{ij} is *the number of trips from i to j per unit time*.



Spatial Flows





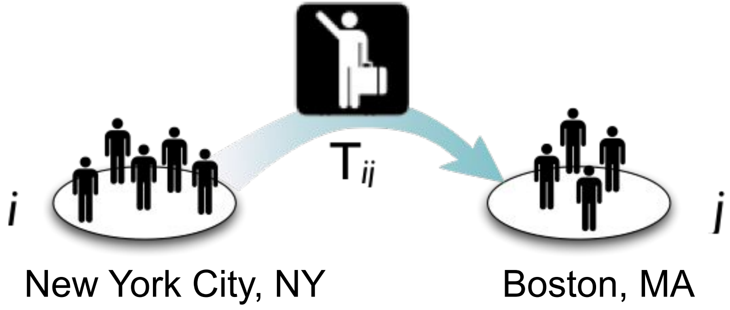
Spatial Flows

destination

origin

	a	b	c	d	e	f	
a	-	3	27	2	1	0	33
b	1	-	4	0	0	5	10
c	8	3	-	1	13	6	31
d	2	1	5	-	0	2	10
e	11	0	6	5	-	1	23
f	0	3	2	2	0	-	7
	22	10	44	10	14	14	114

(self-loops excluded)



total out-flow from i

$$\sum_j T_{ij} = O_i$$

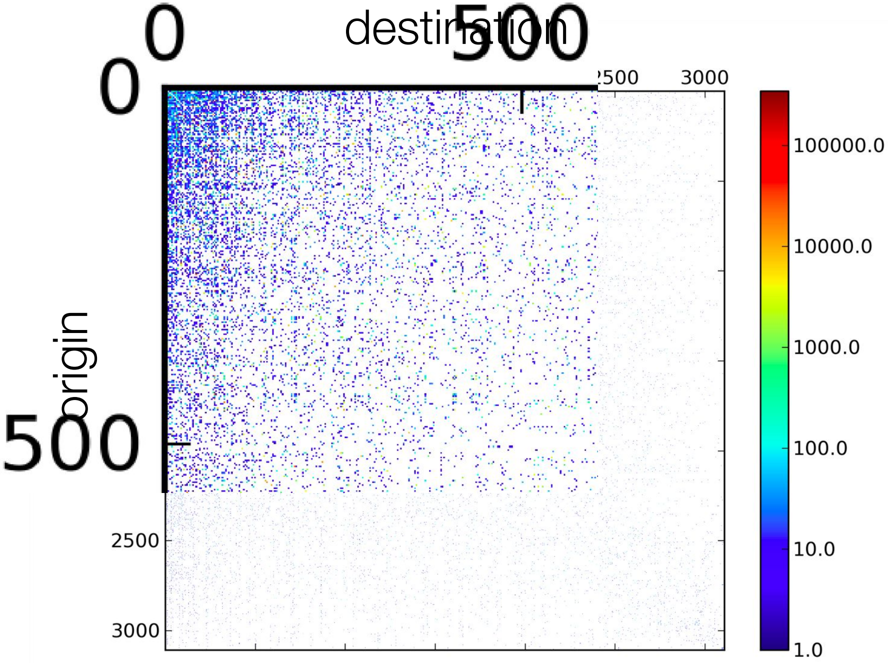
total in-flow to j

$$\sum_i T_{ij} = D_j$$

total flow

$$\sum_{ij} T_{ij} = N$$

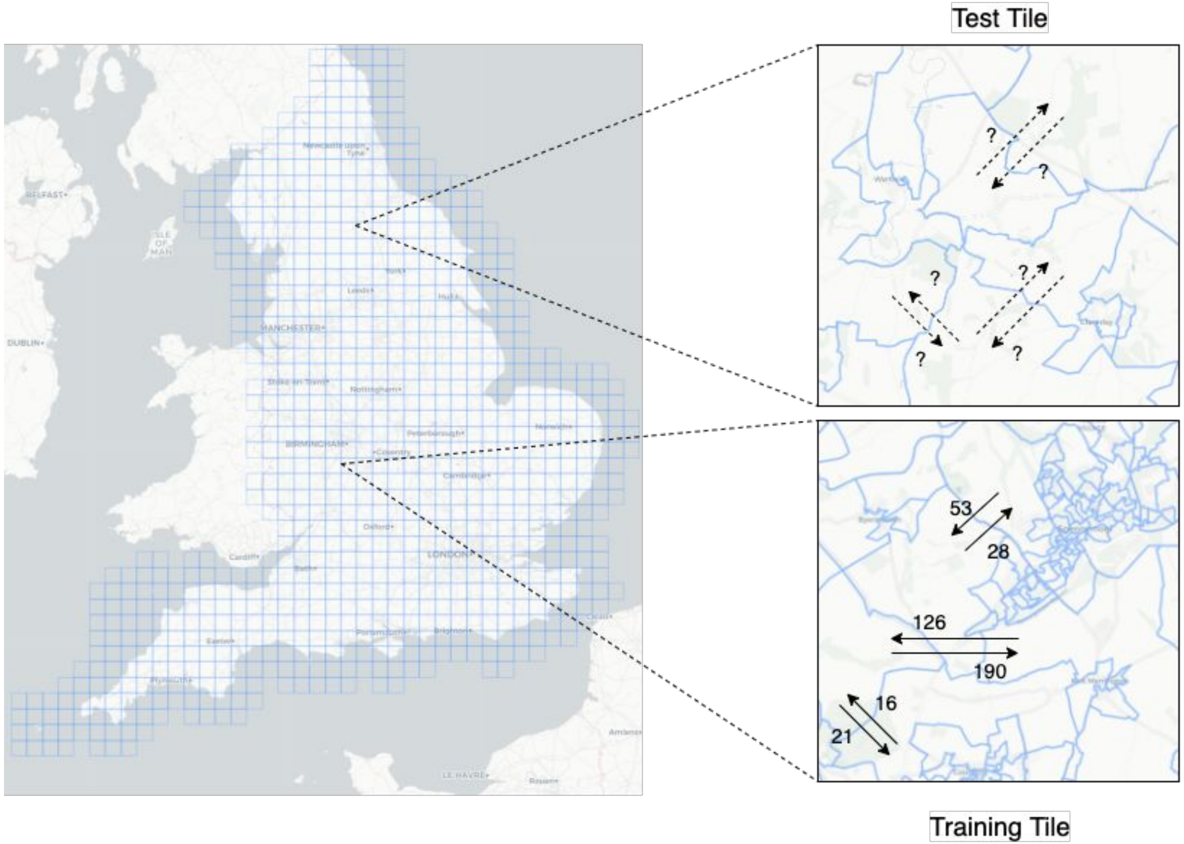
Spatial flows



US county to county commuting flows

Flow generation problem

generate realistic mobility flows among locations given their properties

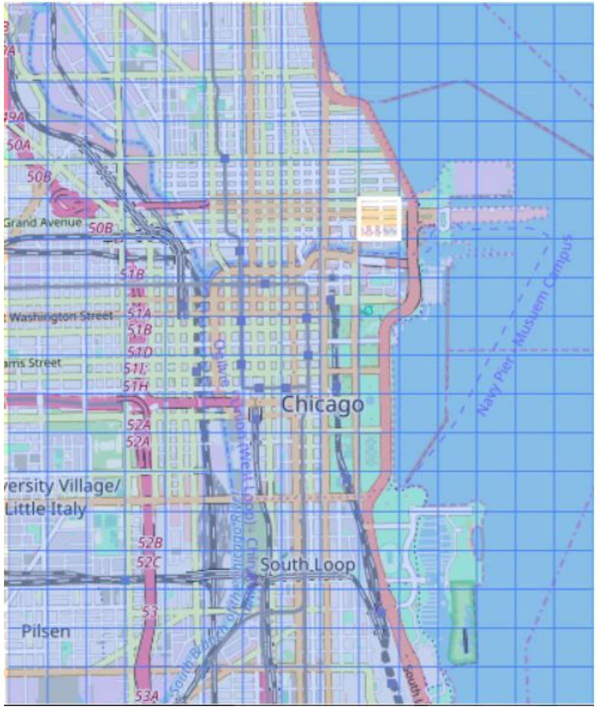


Flow generation problem

Interpret the problem as a classification task

classes = locations

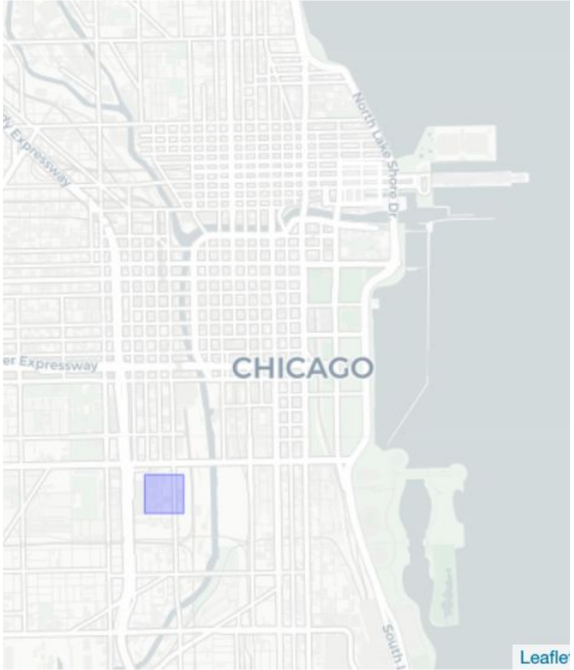
Probabilistic models



Interpret the problem as
a classification task

given a trip's origin location,
predict the destination

Probabilistic models



Goal: find the correct class (i.e., location of destination)

Each location has some probability to be the destination

How do we estimate these probabilities?

Probabilistic models

- assign a probability to each possible OD-matrix T

- fit model's parameters
    - maximizing the likelihood of observed T^*
    - minimizing the distance from observed

Constrained models

○ *globally* constrained
(aka unconstrained)

$$\sum_{ij} T_{ij} = N$$

singly

○ *origin* constrained

$$\sum_{j} T_{ij} = O_i \quad \forall i$$

○ *destination* constrained

$$\sum_{i} T_{ij} = D_j \quad \forall j$$

○ *doubly* constrained

$$\sum_{j} T_{ij} = O_i \quad \sum_{i} T_{ij} = D_j$$

Properties of spatial flows

- Flows **decay** with distance

- Flows **grow** with population

- Flows **grow** with opportunities

Two main modelling approaches

1. Gravity (*G*) models
2. Intervening opportunities (*IO*) models

Similarities

Individual trips are independent. A trip’s probability depends on:

- *weight*, an attribute of each individual location
  e.g., population, number of opportunities
- *distance*, a quantity relating a pair of locations

Differences

- different distance variables considered:
  - distance (G) vs # of intervening opportunities (IO)

Gravity model

Gravity model

Analogy with Newton’s law of gravitation:

$$T_{ij} \propto \frac{P_i P_j}{r_{ij}} \longrightarrow T_{ij} = K m_i m_j f(r_{ij})$$

Gravity model

Analogy with Newton’s law of gravitation:

$$T_{ij} \propto \frac{P_i P_j}{r_{ij}} \longrightarrow T_{ij} = K m_i^{\alpha} m_j^{\beta} \boxed{f(r_{ij})}$$

$$f(r_{ij}) = r_{ij}^{\gamma}$$

power law

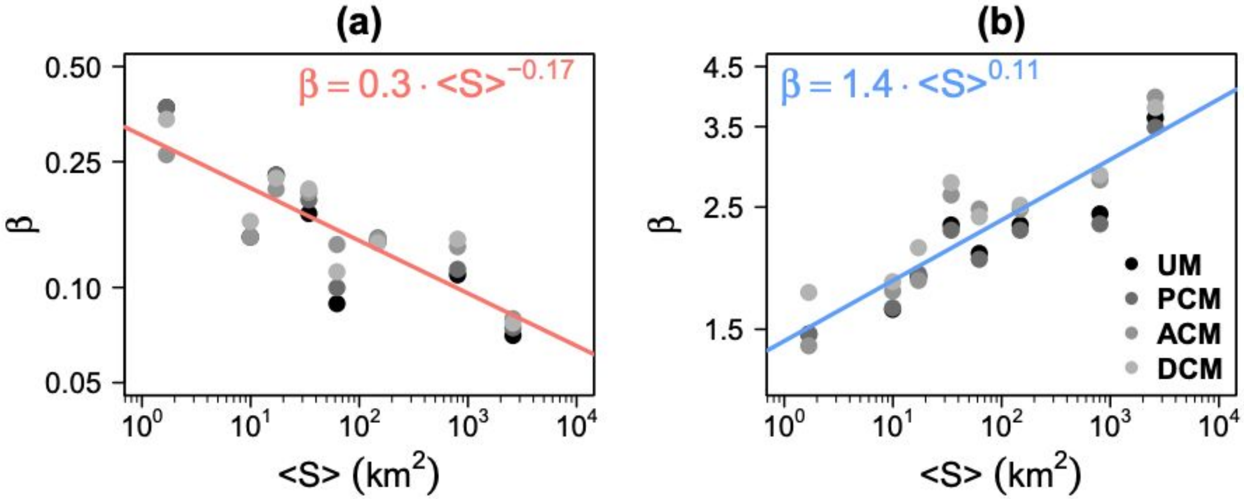
$$f(r_{ij}) = e^{\gamma r_{ij}}$$

exponential

$$f(r_{ij}) = \alpha r_{ij}^{\beta} e^{\gamma r_{ij}}$$

combination

the function’s optimal form may change according to:
the purpose of the trips, the spatial granularity, and the transportation mode



The distance exponent β as a function of average unit surface area. (a) Normalized gravity laws with an exponential distance decay function. (b) Normalized gravity laws with a power distance decay function. Figure 25 from https://arxiv.org/pdf/1710.00004.pdf

Constrained gravity models

The number of people originating from a location, or arriving to, are constrained to be a known quantity, and the gravity model is then used to estimate the destination:

proportionality constant

Singly constrained

$$T_{ij} = \boxed{K_i} O_i m_j f(r_{ij}) = O_i \frac{m_j f(r_{ij})}{\sum_k m_k f(r_{ik})} \qquad O_i = \sum_j T_{ij}$$

Globally constrained

$$T_{ij} = K_i O_i L_j D_j f(r_{ij}) \qquad D_j = \sum_i T_{ij}$$

$$K_i = \frac{1}{\sum_j L_j D_j f(r_{ij})} \qquad L_j = \frac{1}{\sum_i K_i O_i f(r_{ij})}$$

Constrained gravity models

The number of people originating from a location, or arriving to, are constrained to be a known quantity, and the gravity model is then used to estimate the destination:

proportionality constant

Singly constrained

$$T_{ij} = \boxed{K_i} O_i m_j f(r_{ij}) = O_i \frac{m_j f(r_{ij})}{\sum_k m_k f(r_{ik})}$$

$$O_i = \sum_j T_{ij}$$

Globally constrained

$$T_{ij} = O_i D_j \frac{f(r_{ij})}{\sum_j L_j D_j f(r_{ij}) \sum_i K_i O_i f(r_{ij})}$$

$$D_j = \sum_i T_{ij}$$

Choosing the right gravity model

The use of singly-, doubly- or non-constrained models depends on the amount of information available and on the pursued objective:

- If the aim is to approximate the mobility flows and transport demand from indirect socio-economic variables of different geographical areas, then one employs non-constrained models

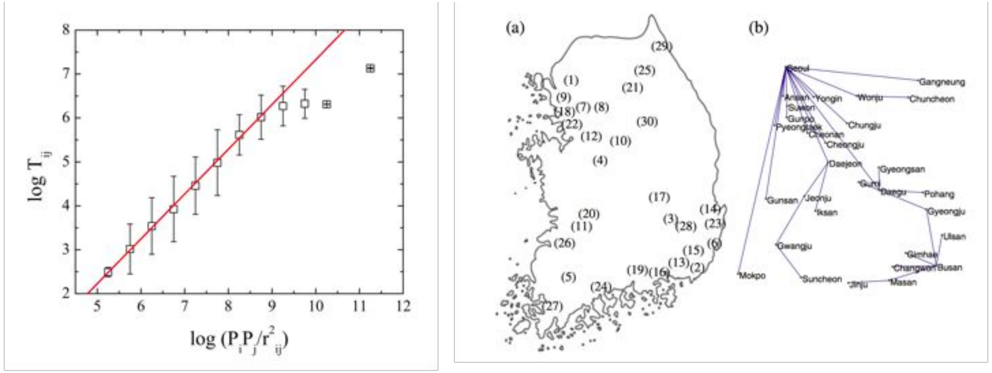
- if out-going or in-going flows are empirically measured quantities, and the objective is to estimate the elements of the OD matrix, then one employs constrained models

Fitting the gravity model

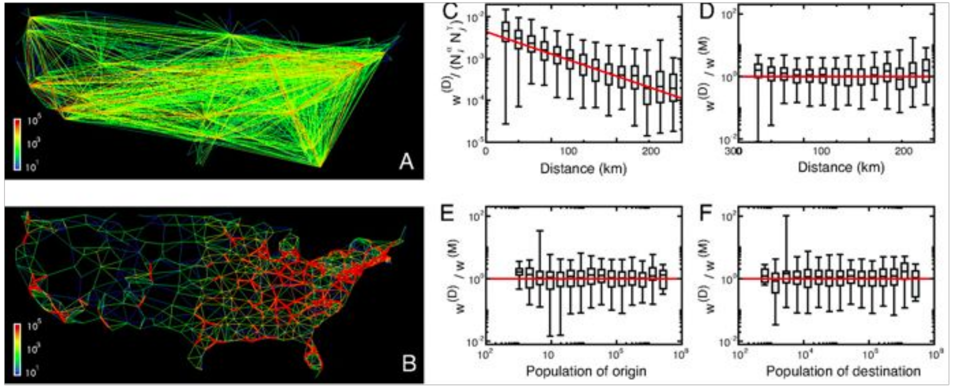
1. The set of independent variables (e.g., population size, gdp, distance) and the functions for these variables and the distance are established
    - power laws for populations and exponential or power laws for the distance dependence are common choices

2. The parameter values are selected to maximize the fit between the estimated flows and the empirical flows:
    - The best fit values of the parameters are determined using an optimization algorithm that minimizes some error function or maximizes the likelihood function of the observed data given the model’s parameters
    - Generalized Linear Models (GLM) (generalization of linear regression) are usually applied to fit the parameters of globally and singly constrained gravity models

Gravity model: applications



Jung, W. S., Wang, F., & Stanley, H. E. (2008). Gravity model in the Korean highway. *EPL (Europhysics Letters)*, *81* (4), 48005.



Balcan, D., et al. "Multiscale mobility networks and the spatial spreading of infectious diseases." *PNAS* 106.51 (2009): 21484-21489.

Gravity model

PROs

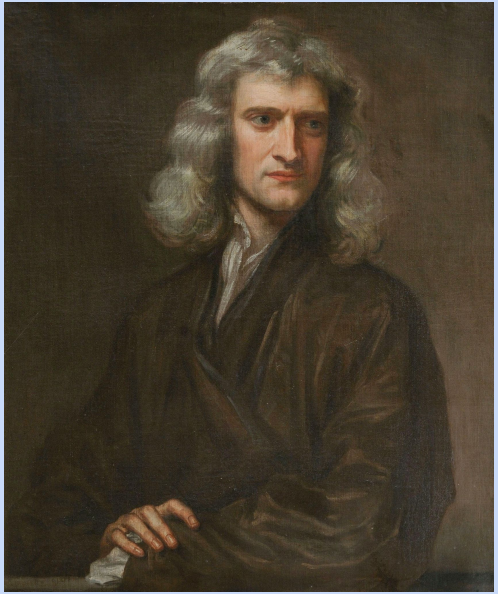
- parameters are easy to fit
- state-of-the-art performance
- versatility and wide applicability

CONs

- underfitting
- low generalisation power

INTERVALLO

Newton and the apple accident



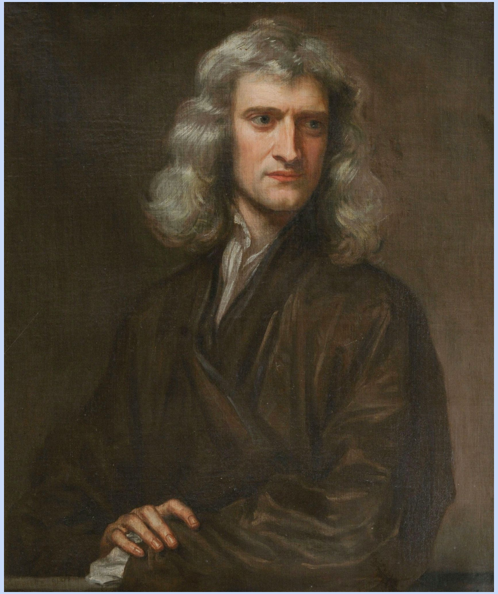
Newton came up with his theory of universal gravitation as a result of an apple falling on his head.

Is this story true?

YES!

Newton himself told the story many times and claimed that the incident had inspired him.

Newton and the apple accident



In his “Memoirs of Sir Isaac Newton’s Life” (1752), William Stukeley mentions a conversation in which Newton described pondering the nature of gravity while watching an apple fall:

> *“…we went into the garden, & drank thea under the shade of some apple trees; only he, & my self. amidst other discourse, he told me, he was just in the same situation, as when formerly, the notion of gravitation came into his mind. “why should that apple always descend perpendicularly to the ground,” thought he to himself; occasion’d by the fall of an apple…”*

INTERVALLO

Where is Newton’s apple tree?

Various trees are claimed to be “the” apple tree:

- The King’s School in Grantham claims they purchased the original tree, uprooted it, and transported it to the headmaster’s garden some years later;
- The National Trust, which holds the Woolsthorpe Manor (where Newton grew up) in trust, claims that the tree still resides in their garden.
- A descendant of the original tree can be seen growing outside the main gate of Trinity College, Cambridge, below the room Newton lived in when he studied there.

Intervening opportunities

Intervening opportunities (IO)

> “ 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 ”

Stouffer, 1940

Stouffer, American Sociological Review 1940, Intervening opportunities: a theory relating mobility and distance

Intervening opportunities (IO)

Stouffer proposed a conceptual framework in which distance and mobility are not directly related:

- what plays the key role in determining migration is the **number of intervening opportunities** between the origin and the destination
- Stouffer does not provide a precise definition for “opportunities”, leaving it to be defined depending on the social phenomena under investigation

Stouffer, American Sociological Review 1940, Intervening opportunities: a theory relating mobility and distance

Intervening opportunities (IO)

The decision to make a trip is explicitly related to the relative accessibility of opportunities for satisfying the objective of the trip:

- an opportunity is a destination that a trip-maker considers as a possible termination point for their journey

- an intervening opportunity is a location that is closer to the trip maker than the final destination but is rejected by the trip maker

Stouffer, American Sociological Review 1940, Intervening opportunities: a theory relating mobility and distance

Intervening opportunities (IO)

> “The probability that a trip ends in a given location is equal to the probability that this location offers an acceptable opportunity times the probability that an acceptable opportunity in another location closer to the origin of the trips has not been chosen.”

Schneider, 1959

M. Schneider, Gravity models and trip distribution theory, Papers of the regional science association 5 (1959) 51–58.

Intervening opportunities (IO)

cumulative number of
opportunities up to
the j-th location
ranked by travel cost
from origin location

$$T_{ij} = O_i \frac{e^{-LV_{i,j-1}} - e^{-L\boxed{V_{i,j}}}}{\boxed{1 - e^{-LV_{i,n}}}}$$

normalization factor

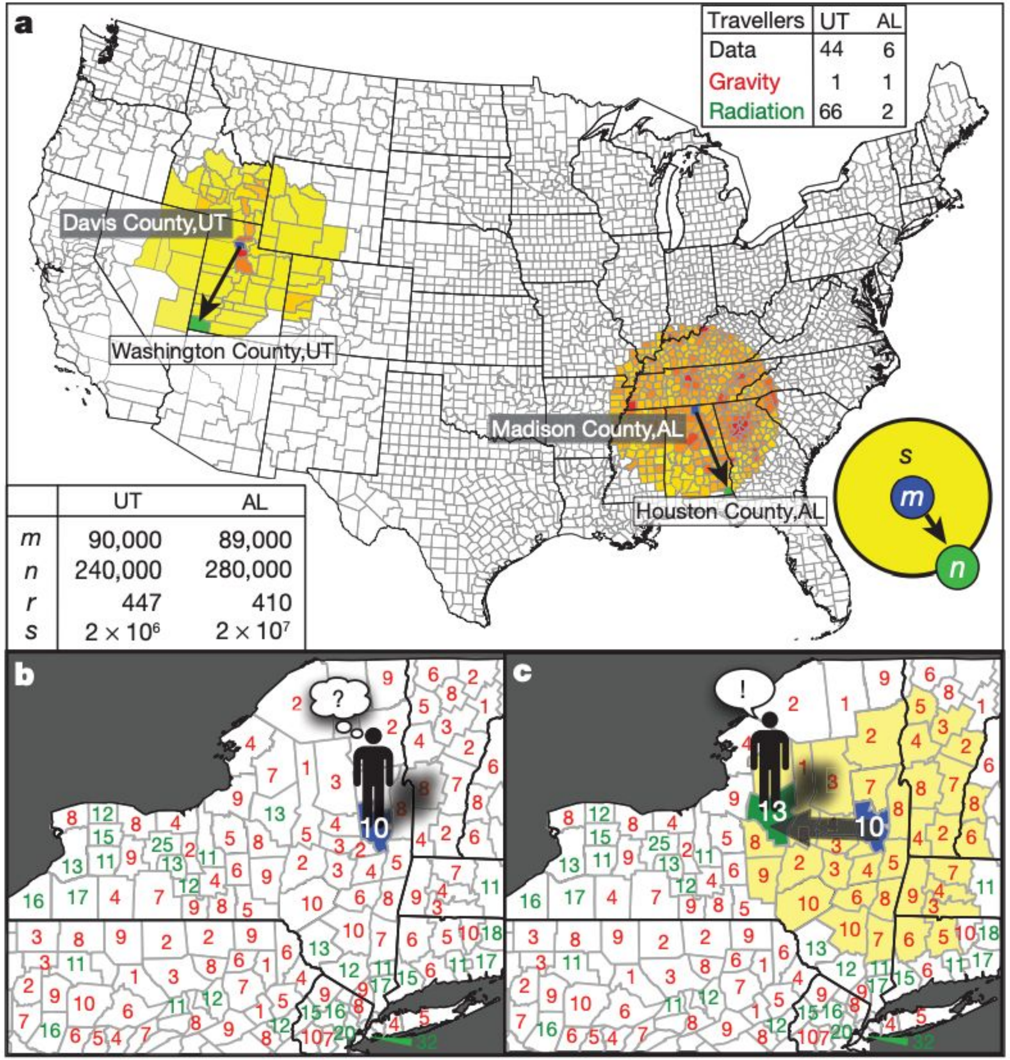
- Usually, the population or the total number of arrivals are assumed to be proportional to the number of “real opportunities” in a location
- L is the constant probability of accepting an opportunity destination
  - As in the case of the gravity model, the value of L is adjusted in order to obtain simulated flows as close as possible to observed data

Radiation model

The radiation model elaborates on the IO hypothesis and assumes that the choice of a traveler's destination consists of these steps:

1. each opportunity in every location is assigned a **fitness** z chosen from some distribution $p(z)$, i.e., the quality of the opportunity for the traveler
2. the traveler ranks all opportunities according to their distance from the origin location
3. the traveler chooses the closest opportunity with a fitness higher than the traveler's fitness threshold (a random number extracted from $p(z)$)

Simini et al., Nature 2012 A universal model for mobility and migration patterns



Travellers	UT	AL
Data	44	6
Gravity	1	1
Radiation	66	2

	UT	AL
m	90,000	89,000
n	240,000	280,000
r	447	410
s	2×10^6	2×10^7

- Each opportunity has a “value”, extracted from some distribution.
- Each individual has expectations, extracted from the same distribution.
- Principle of least effort: each individual chooses the closest opportunity that meets their expectations

Radiation model

Parameter-free: the model depends only on the populations

opportunities at the origin

opportunities at the destination

$$T_{ij} = O_i \boxed{\frac{1}{1 - \frac{m_i}{M}}} \frac{\boxed{m_i}\boxed{m_j}}{(m_i + \boxed{s_{ij}})(m_i + m_j + s_{ij})}$$

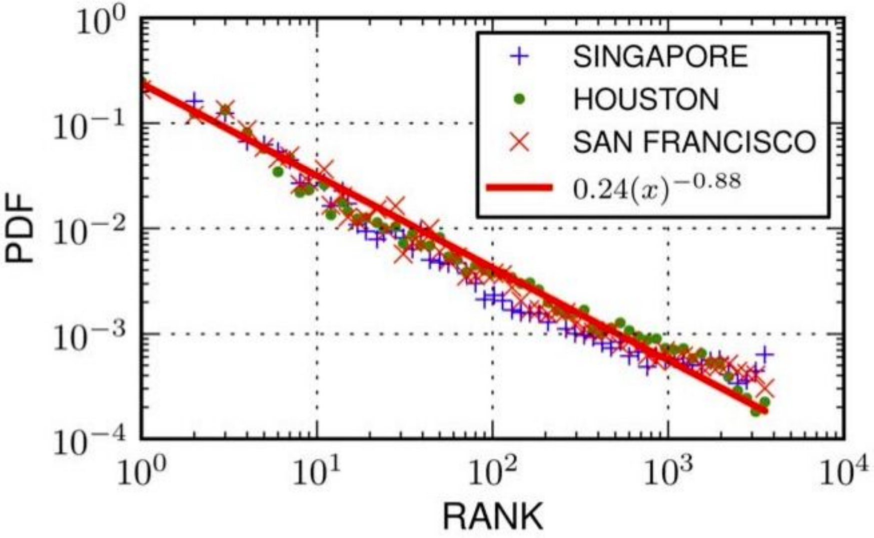
normalization factor

(so that the probability that a trip originating in the region ends in this location is 1)

opportunities in a circle of radius r_{ij} centered in the origin location i (excluding origin and destination)

Rank-distance model

- Data: Foursquare check-ins
  - 925,030 users over 6 months
  - 5 million places, 34 cities, 4 continents, 11 countries

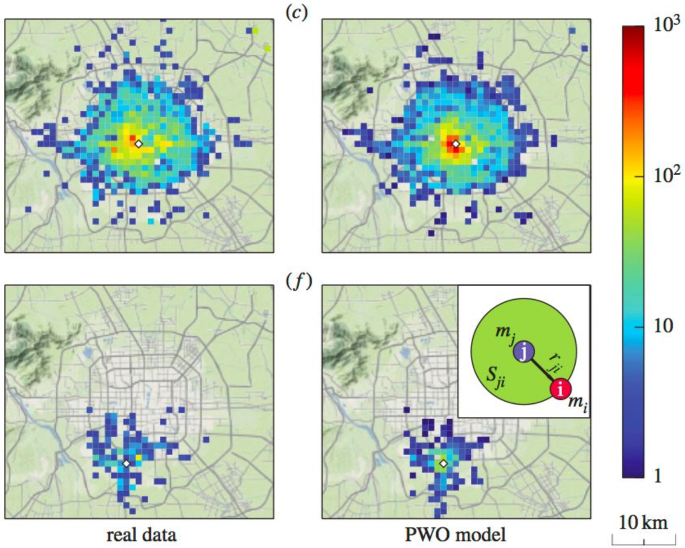


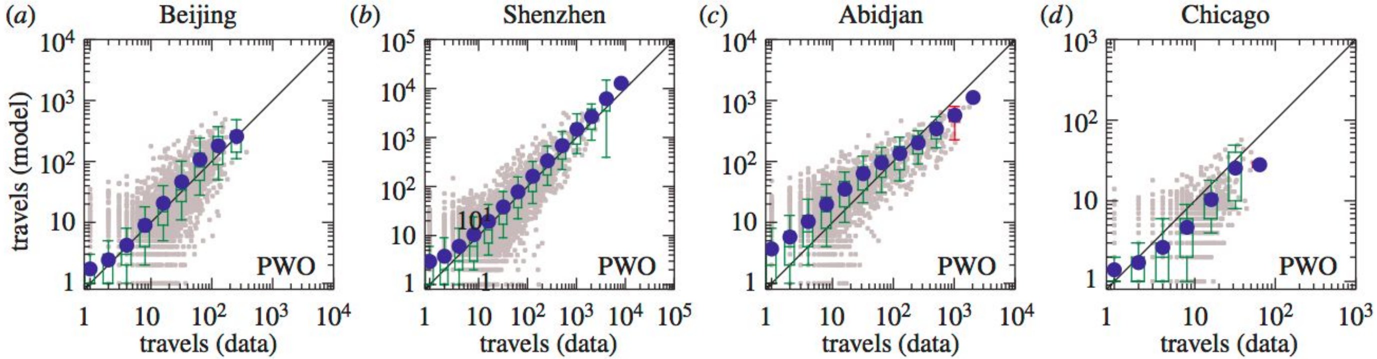
$$p_{ij} \propto \frac{1}{(m_i + s_{ij})^{\alpha}}$$

Rank-distance model

The PWO model considers the intervening opportunities centered at the destination:

$$p_{ij} \propto m_j \left(\frac{1}{m_i + m_j + {\color{red} s_{ji}}} - \frac{1}{M} \right)$$





Intervening Opportunities

PROs ✓

- parameter-free (Radiation and PWO)
- performance comparable to Gravity models

CONs ✗

- underfitting
- overdispersion

Validation of collective models

Common metrics to compare OD matrices

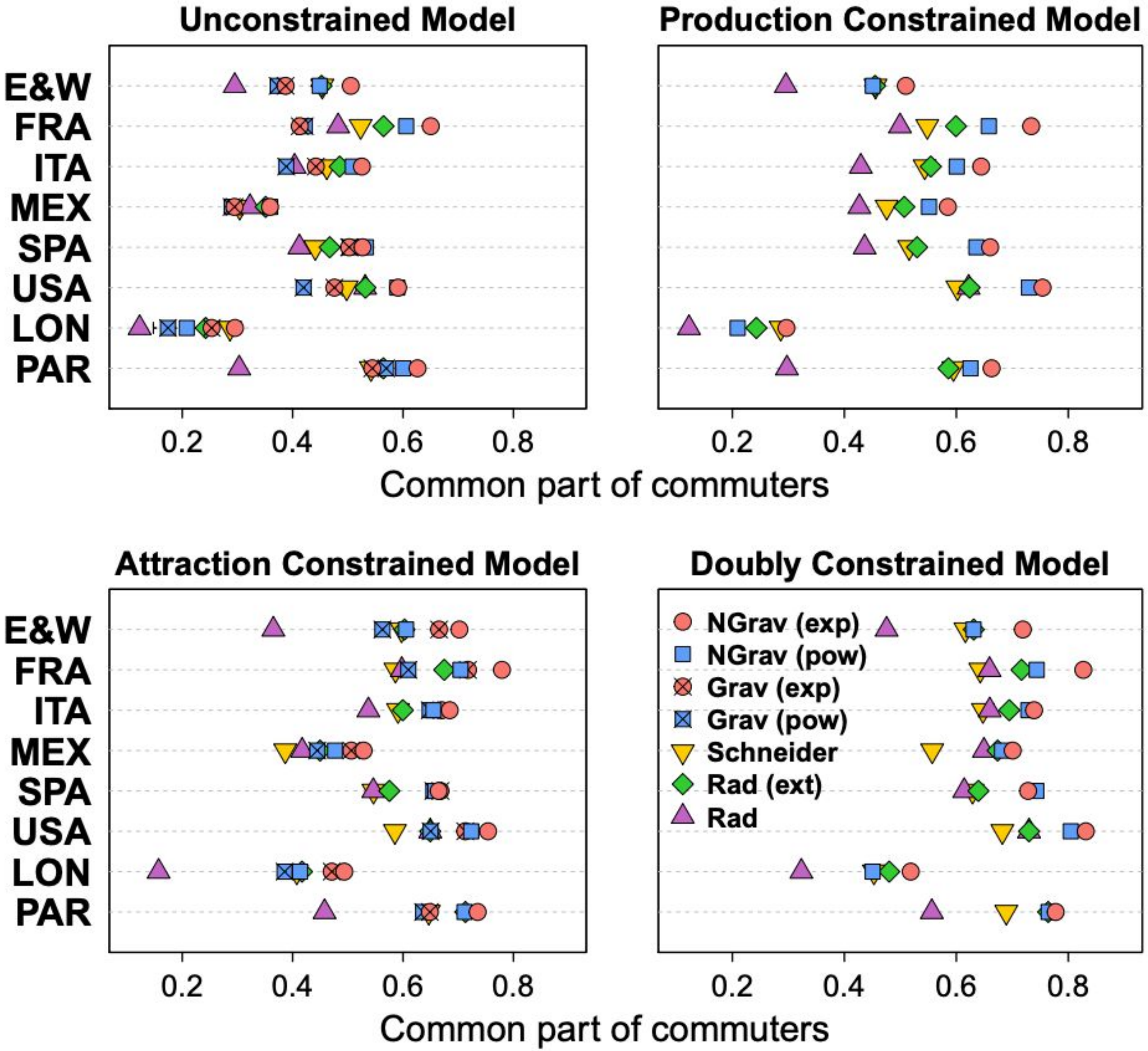
- Sorensen-Dice similarity
  (Common part of commuters)

$$\frac{\sum_{ij} \min(T_{ij}^{e}, T_{ij}^{m})}{\sum_{ij} T_{ij}^{e}}$$

- Root Mean Squared Error

$$\sqrt{\frac{\sum_{ij} (T_{ij}^{e} - T_{ij}^{m})^2}{n^2}}$$

- More (cosine similarity, correlation, …)


Unconstrained Model
Production Constrained Model
Attraction Constrained Model
Doubly Constrained Model
E&W
FRA
ITA
MEX
SPA
USA
LON
PAR
0.2
0.4
0.6
0.8
Common part of commuters
NGrav (exp)
NGrav (pow)
Grav (exp)
Grav (pow)
Schneider
Rad (ext)
Rad

References

- [paper] Human Mobility: Models and Applications, Barbosa et al., Physics Report, 2018, Section 4.2
- [paper] The P1 P2/D Hypothesis: On the Intercity Movement of Persons, Zipf, American Sociological Review, 1946
- [paper] Intervening Opportunities: A Theory Relating Mobility and Distance, Stouffer, American Sociological Review, 1940
- [paper] Gravity models and trip distribution theory, Schneider, Papers of the regional science association, 1959

References

- [paper] A universal model for mobility and migration patterns, Simini et al., Nature, 2012
- [paper] A tale of many cities: universal patterns in human urban mobility, Noulas et al., PloS One, 2012
- [paper] Systematic comparison of trip distribution laws and models, Lenormand et al., Journal of Transport Geography, 2016

Homeworks

to be delivered by Thursday, November 3rd, 2022

Homework 7.1

In how many different contexts has Newton's gravity laws been applied? Find at least five cases not mentioned at lesson and write a blog post about it.

In particular, for each case describe in detail:

- The scientific paper where this approach is described
- The data used for this, and whether they are publicly available
- The results found

Discuss a possible application of the gravity laws in a context that it has been not applied yet.

Homework 7.2

Download the flows for at least two different US States from this repository, create and plot a FlowDataFrame. Then:

- split the FlowDataFrame into a training set and a test set;
- train the Gravity and Radiation models on the training set
- test the models' goodness on the test set (qualitative and quantitative evaluation). Use population as location relevance.
- Compare the two models with appropriate plots and/or tables.
- Repeat using the number of Education facilities in each location instead of the popultion (i.e., total count of POIs and buildings related to all education facilities, e.g., school, college, kindergarten, etc.).

Submit a well-commented notebook.