Human Mobility and Migration Patterns

Sharmishtha Dutta

CSCI 6250: Frontiers of Network Science
Fall 2021

November 22, 2021

Rensselaer

Introduction

- Research[1] shows that human mobility is barely random
- Useful for

  epidemiology
  spreading processes
  urban geography
  flow of resources in economics
  crisis response
  predicting population movement and so on

[1] Zhao, C., Zeng, A. and Yeung, C.H., 2021. Characteristics of human mobility patterns revealed by high-frequency cell-phone position data. EPJ Data Science, 10(1), p.5.

Selected Paper

- A universal model for mobility and migration patterns
- Filippo Simini, Marta C. Gonzalez, Amos Maritan, Albert-Laszlo Barabasi
- Nature, 2012
- Cited 1200+ times
- Major contribution: Proposed model overcomes six limitations of the previous prevailing framework, validated by experimental evaluations

Background

The gravity law (1946): inspired by Newton's gravity law

$$T_{ij} = \frac{m_i^{\alpha} n_j^{\beta}}{f(r_{ij})} \tag{1}$$

Where

- T_{ij} = Number of commuters between locations i and j per unit time
- r_{ij} = The distance between i and j
- α and β are adjustable exponents
- f is chosen to fit the empirical data

Limitations of the gravity law

1. No rigorous derivation of equation 1
   - Entropy maximization leads to $\alpha = \beta = 1$
   - No functional form for $f(r)$
2. It lacks theoretical guidance
   - Practitioners choose a range of functions for $f(r)$
   - Up to nine parameters needed to fit the empirical data
3. Requires previous traffic data
   - The value of α, β is chosen after fitting the model to data
   - Areas of major interest often lack data

Limitations of the gravity law (cont.)

4. Has systematic predictive discrepancies [figure 1]
   Predicts similar values for two locations that has similar population and distance
5. Predicts that T_{ij} increases without limit
   T_{ij} is proportinal to location populations
   increasing the destination population (e.g. T_j) results in a $T_{ij} > T_i$
6. Deterministic - cannot account for fluctuations in the number of travellers

Example of Systematic predictive discrepancy

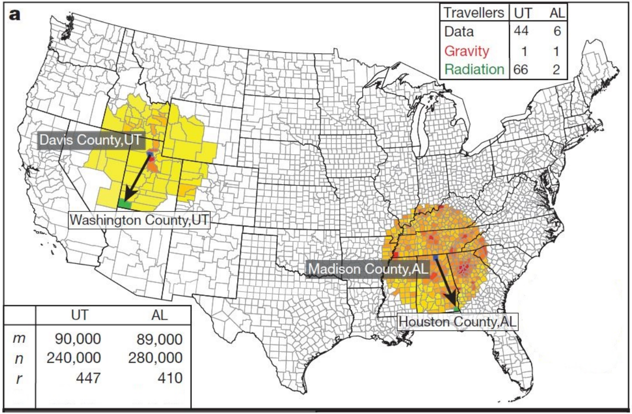

Figure 1: Systematic predictive discrepancy

Proposed model

Fundamental Observations

- Commuting is a daily process
- Source and destination determined by job selection
- Job opportunity depends on the population around a specific location [figure 2]
- Use counties as location unit
- Assume, individuals choose lesser jobs closer to their home county

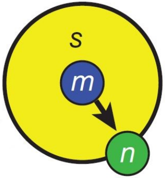

Figure 2: yellow zone denotes s_{ij}

Proposed model

Radiation model

The average flux T_{ij} from location i to location j, distanced at r_{ij} is:

$$T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \tag{2}$$

- T_i = the total number of commuters that start their journey from location i
- s_{ij} = the total population in the circle of radius r_{ij} centred at i (excluding m_i, n_j)[figure 2]

$T_i = m_i(N_c/N)$ where, N_c = the total number of commuters, N = the country's population

Resolving the limitations of gravity law

1. The radiation model [equation 2] has a rigorous derivation[2]
2. The radiation model [equation 2] has no free parameters[3]
3. The variable has been changed from distance - r_{ij} to population density - s_{ij}[4]
    - The population density around Utah is significantly lower than the United States average
    - Work opportunities within the same radius are ten times smaller in Utah than in Alabama
    - Commuters in Utah have to travel farther to find comparable employment opportunities.

[2] Solves limitation 1: The gravity law has no rigorous derivation of equation

[3] Solves limitation 2 and 3: No theoretical guidance found for gravity law and no functional form for $f(r)$

[4] Solves limitation 4: systematic predictive discrepancy

Resolving the limitations of gravity law

Equations

$$Gravity : T_{ij} = \frac{m_i^{\alpha} n_j^{\beta}}{f(r_{ij})}, Radiation : T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}$$

4. the number of travellers leaving from a location with population m to one with $n \to \infty$ saturates at $T_{n \to \infty} = \frac{m^2}{m+s} + \mathcal{O}(\frac{1}{n}) \leq m$ resolving the unphysical divergence highlighted in limitation five[5]
5. T_{ij} in the radiation model is a stochastic variable, predicting not only the average flux, but also its variance[6]
   - Can account for fluctuations in the number of travellers

[5] Solves limitation 5: gravity law predicts that T_{ij} increases without limit

[6] Solves limitation 6: gravity law cannot account for fluctuations in the number of travellers

National mobility fluxes in USA

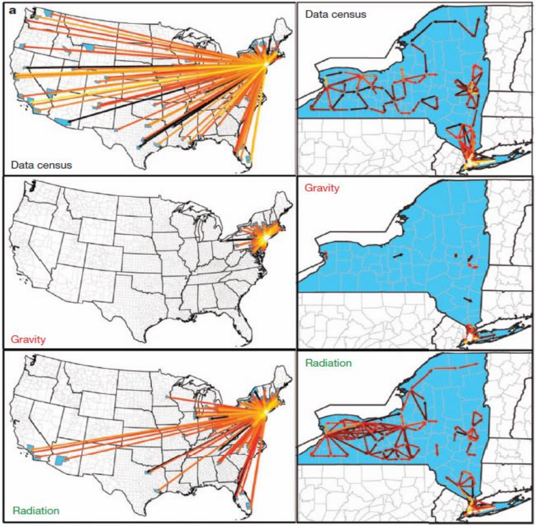

Figure 3: National mobility fluxes

Measured flux vs Predicted flux

- Grey points are scatter plot for each pair of counties
- A box is coloured green if the line $y = x$ lies between the 9th and the 91st percentiles in that bin and is red otherwise
- The black circles correspond to the mean number of predicted travellers in that bin.

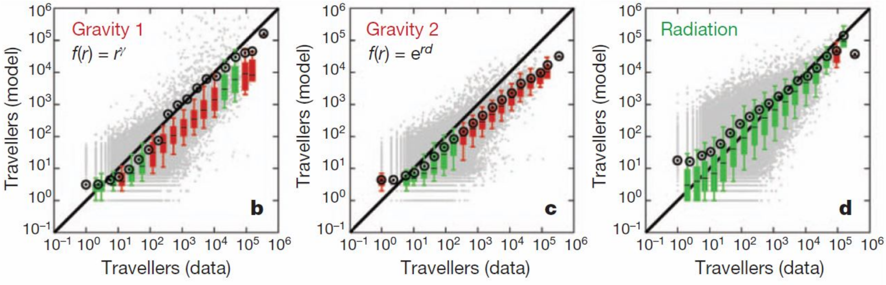

Figure 4: Measured flux vs Predicted flux

Probability measures

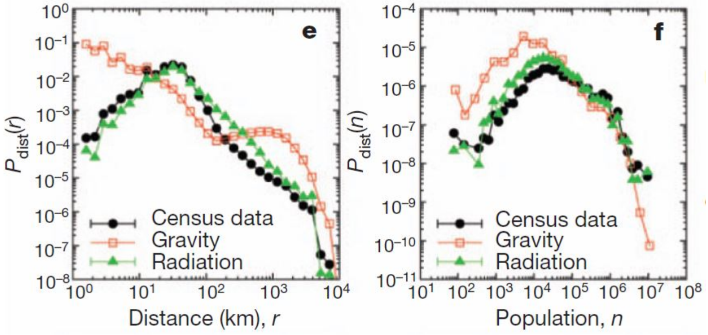

Figure 5: $P_{dist}(r)$: Probability of a trip between two counties that are at distance r and $P_{dist}(n)$: Probability of a trip towards a county with population n

Probability measures

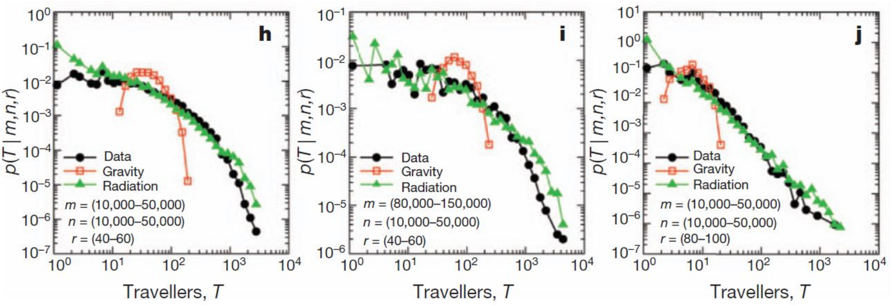

Figure 6: Conditional probability $P(T|m,n,r)$ to observe a flow of T individuals from a location with population m to a location with population n at a distance r for three triplets (m,n,r)

Generality of the model

- To verify the generality of the model, it is tested on four socio-economic phenomena: hourly travel patterns, migrations, communication patterns and commodity flows.
- An accurate quantitative description of mobility and transport is observed (spanning a wide range of time scales e.g. hourly mobility, daily commuting, yearly migrations)
- It captures diverse processes (commuting, intra-day mobility, call patterns, trade)
- The data is collected via a wide range of tools (census, mobile phones, tax documents) on different continents (America, Europe)

Beyond commuting

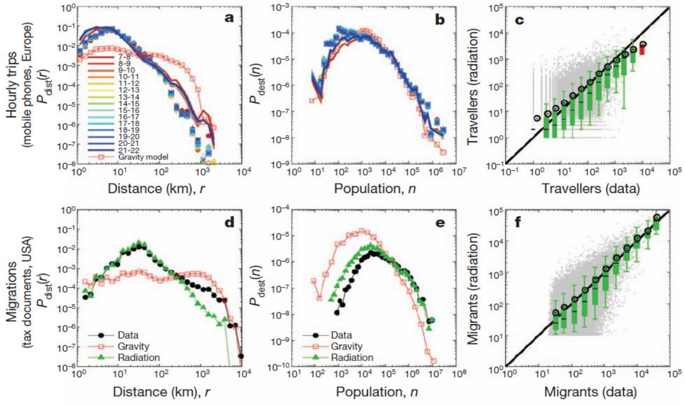

Figure 7: Testing the radiation model on **hourly** trips extracted from a mobile phone database of a western European country and long-term migration patterns among US counties **over a year**

Beyond commuting

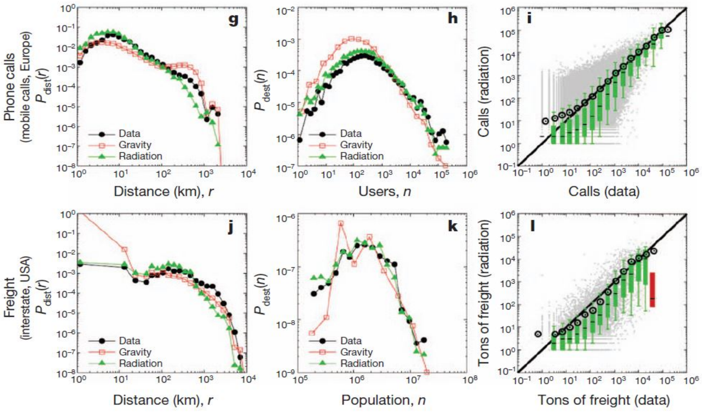

Figure 8: Phone call volume between municipalities **over 4-week period** and Commodity flows in the US during the **year** 2007

Generality of the model

Observation

The agreement with data of such diverse nature suggests that the hypotheses behind the model capture fundamental decision mechanisms that, directly or indirectly, are relevant to a wide span of mobility and transport-driven processes.

Follow up paper 1

Population flow drives spatio-temporal distribution of COVID-19 in China, Nature[7]

- Used a nationwide mobile phone data to track population outflow from Wuhan
- Linked this data to COVID-19 infection counts by location (prefecture level)
- Finding: A strong correlation between total population flow and the number of infections in each prefecture

[7] Jia, J.S., Lu, X., Yuan, Y., Xu, G., Jia, J. and Christakis, N.A., 2020. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature, 582(7812), pp.389-394.

Geographical distribution

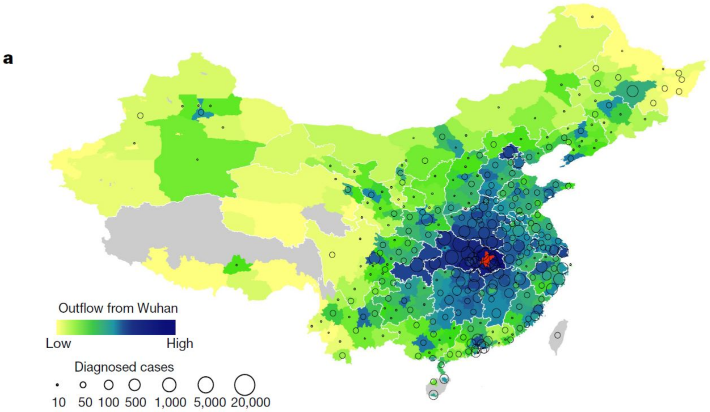

Figure 9: Geographical distribution of population outflow and confirmed COVID-19 cases as of 19 February 2020

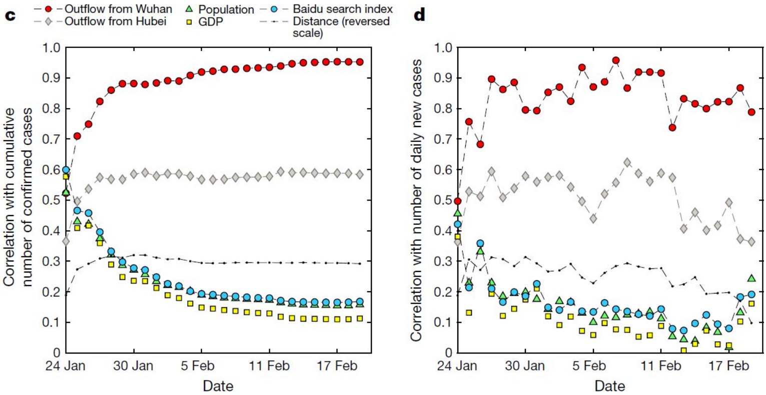

Figure 10: Relationship over time between the number of confirmed cases (cumulative until 19 February 2020) and the cumulative population inflow (up to 24 January 2020) from Wuhan, the cumulative inflow from Hubei province excluding Wuhan and others[8]

[8] the frequency of Baidu search terms related to the virus, the GDP, population and distance from Wuhan of the prefectures

Follow up paper 2

This paper[9] relates the number of infected individuals I in different provinces in China with the population migration, P_m, from Hubei:

$$I \propto P_m^{\phi} \tag{3}$$

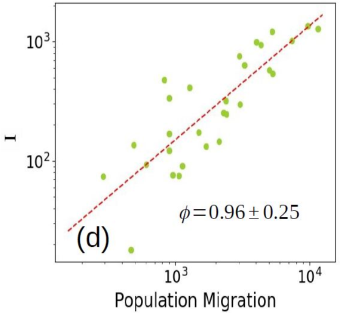

Figure 11: Migration vs infection rate: suggests an almost linear scaling relation

[9] Gross, B., Zheng, Z., Liu, S., Chen, X., Sela, A., Li, J., Li, D. and Havlin, S., 2020. Spatio-temporal propagation of COVID-19 pandemics. EPL (Europhysics Letters), 131(5), p.58003.

Follow up paper 3

This paper[10]:

- Describes the data and methodology that power Facebook Disaster Maps
- The method collects Facebook usage in areas impacted by natural hazards
- Goal: describe how the population is affected by and responding to the hazard

[10] Maas, P., Iyer, S., Gros, A., Park, W., McGorman, L., Nayak, C. and Dow, P.A., 2019, May. Facebook Disaster Maps: Aggregate Insights for Crisis Response & Recovery. In KDD (Vol. 19, p. 3173).

Background

- The paper describes 5 maps: Population, Movement, Power Availability, Network Coverage, Displacement
- Collects location data from Facebook users
- Privacy protection mechanisms to obscure the identity and actions of individuals and small groups while preserving the population-level insights
  - Adding random noise, spatial smoothing, dropping small counts
- Baselines are derived from 5-to-13 weeks of pre-crisis data

$$z\ score = \frac{c - \mu_{baseline}}{max[\sigma_{baseline}, \sigma_{min}]} \tag{4}$$

Where, c is the data observed in crisis time and $\sigma_{min} \approx 0.1$ handles the case where $\sigma_{baseline} = 0$

Result: Population map

The counts include people with location services enabled on their mobile device.[11]

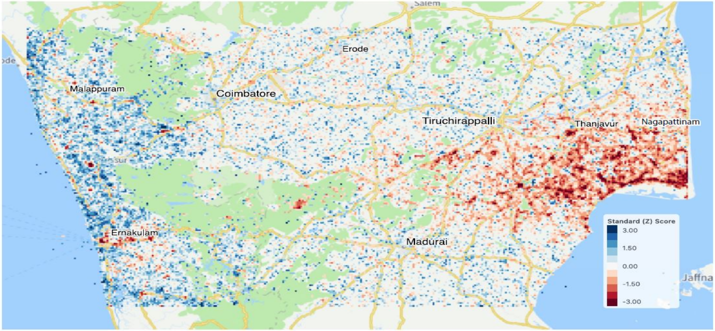

Figure 12: Population map for part of South India on November 17, 2018, in the aftermath of Cyclone Gaja

[11]If the same person appeared at multiple locations in a time, latest and most frequent location is counted

Result: Power availability map

In android devices, it is possible to observe when a user connects it to a power source. The power could come from places like cars, generators, or external batteries, which gives a proxy for the state of the power grid in an area.

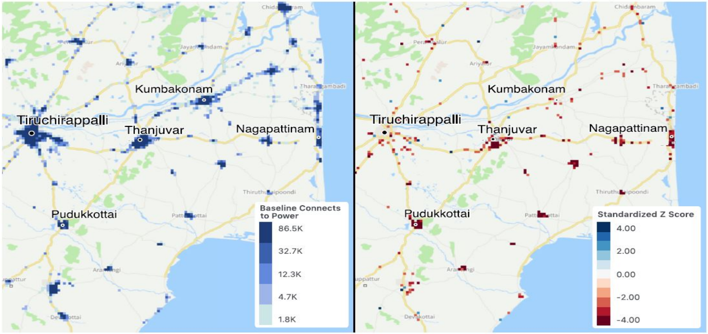

Figure 13: Power Availability maps for Cyclone Gaja on November 17, 2018. The map on the left shows the mean power connections from the baseline time period, while the map on the right shows the crisis-time z-scores.

Questions?