inventors for the digital world

Aline Carneiro Viana
aline.viana@inria.fr
Research Scientist
Inria Saclay - Ile de France

When completion challenges current human mobility knowledge

Guangshuo Chen (eBay), Marco Fiore (CNR), Carlos Sarraute (Grandata Labs)

HAL TR-0495, INRIA Saclay (under submission)

TMA, 18th June 2019

What make us human?

What make us human?

We think

Language skills, imagination, intelligence, mind-reading, creativity, big brain…

"Humans have a unique ability to understand the beliefs of another person (Credit: Thinkstock)"

We feel

Morality, foresight, forethoughts, compassion, blushing, kindness…

"Our open-ended ability to imagine and reflect on different situations, and our deep-seated drive to link our scenario-building minds together. (Credit: Suddendorf)"

We behave

Habits, culture, emotions, storytellers, ability to **control and alter our environment**…

"We take advantage of others' experiences, reflections and imaginings to prudently guide our own behaviour. (Credit: Suddendorf)"

ThoughtCo.: "What makes us human"
TEDx Talks: "Clues about the evolution of our extraordinary minds". Thomas Suddendorf
BBC: "The traits that make human beings unique"

Concert, 90s

Concert 2017

Pervasive connectivity makes our **real** life and **virtual** activities seamlessly merged together

New opportunity

Digital datasets are now mirroring human dynamics and interests

1. Enforces the understanding of human behavior
2. And consequently, the understanding of where, for what, and when services or resources are needed

UNITED NATIONS DEVELOPMENT PROGRAMME

Human Development Reports

"...mobility and migration have always been an instrinsic part of **human development**." by De Hass, Hein, 2009.

Reality Mining of Mobile Communications: Toward a New Deal on Data

ALEX PENTLAND, Massachusetts Institute of Technology (MIT)

An emerging—and truly global—nervous system

The ability to understand the patterns of human life by analyzing the digital traces that we leave behind will transform poor nations even more than rich nations.

My Daily Travel Survey

Between August 2018 and April 2019, the Chicago Metropolitan Agency for Planning (CMAP) conducted the My Daily Travel survey, asking households in northeastern Illinois to tell us about trips they made for work, school, shopping, errands, and socializing with family and friends. Collecting this information through a survey is the only way to accurately measure and understand people's changing daily travel patterns.

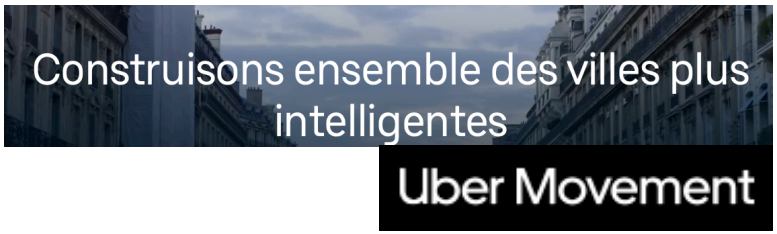
MIT News Browse or Search

Measuring the economy with location data

Startup's platform crunches anonymized smartphone GPS data to understand how people shop, work, and live.

Rob Matheson | MIT News Office
March 27, 2018

Press Inquiries RELATED



Deciphering mobility properties

02

Marta Gonzalez et al. 2008

Recurrence and time periodicity of visited locations

Frequent travels to a limited number of places

Population trip distance distribution follows a truncated power law

Confinement of movement

Movements within a geographical distance bounded by the radio of gyration
Long distances beyond this radio are rare

Understanding individual human mobility patterns

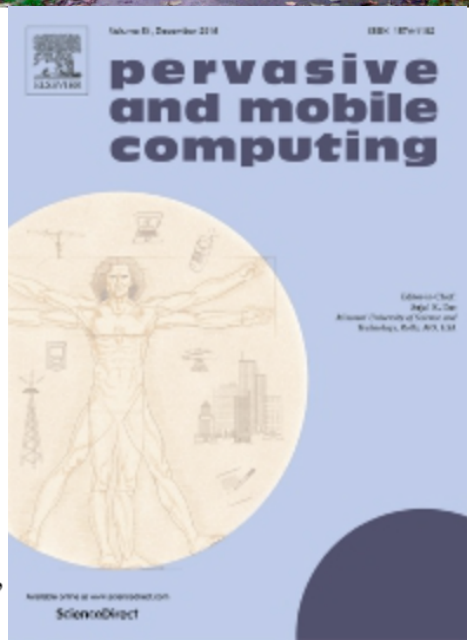
Marta C. González[1], César A. Hidalgo[1,2] & Albert-László Barabási[1,2,3]

Eduardo Mucceli et al. 2016

People from 8 cities in 3 continents, appearing in 3 data sources

Tendency to follow the "path of least resistance" or "desire lines"

Repetitiveness and Confinement of movement



On the Regularity of Human Mobility

Eduardo Mucelli Rezende Oliveira[a,b,*], Aline Carneiro Viana[b], Carlos Sarraute[c], Jorge Brea[c], Ignacio Alvarez-Hamelin[d]

Chaoming Song et al. 2010

Recurrence and time periodicity of visited locations

Frequent travels to a limited number of places

High theoretical predictability of human mobility (upper bound: 93%)

Thanks to frequent travels to a limited number of places

Science AAAS

Limits of Predictability in Human Mobility

Chaoming Song,[1,2] Zehui Qu,[1,2,3] Nicholas Blumm,[1,2] Albert-László Barabási[1,2]*

Inria inventors for the digital world

Yves-Alexandre de Montjoye et al. 2013

Uniqueness of individual trajectories

4 random locations identify 1 user among 1.5 million, 95% of time

Unique in the Crowd: The privacy bounds of human mobility

Yves-Alexandre de Montjoye[1,2], César A. Hidalgo[1,3,4], Michel Verleysen[2] & Vincent D. Blondel[2,5]

SCIENTIFIC REPORTS

Call Detail Record (CDR) datasets: Beast or Beauty

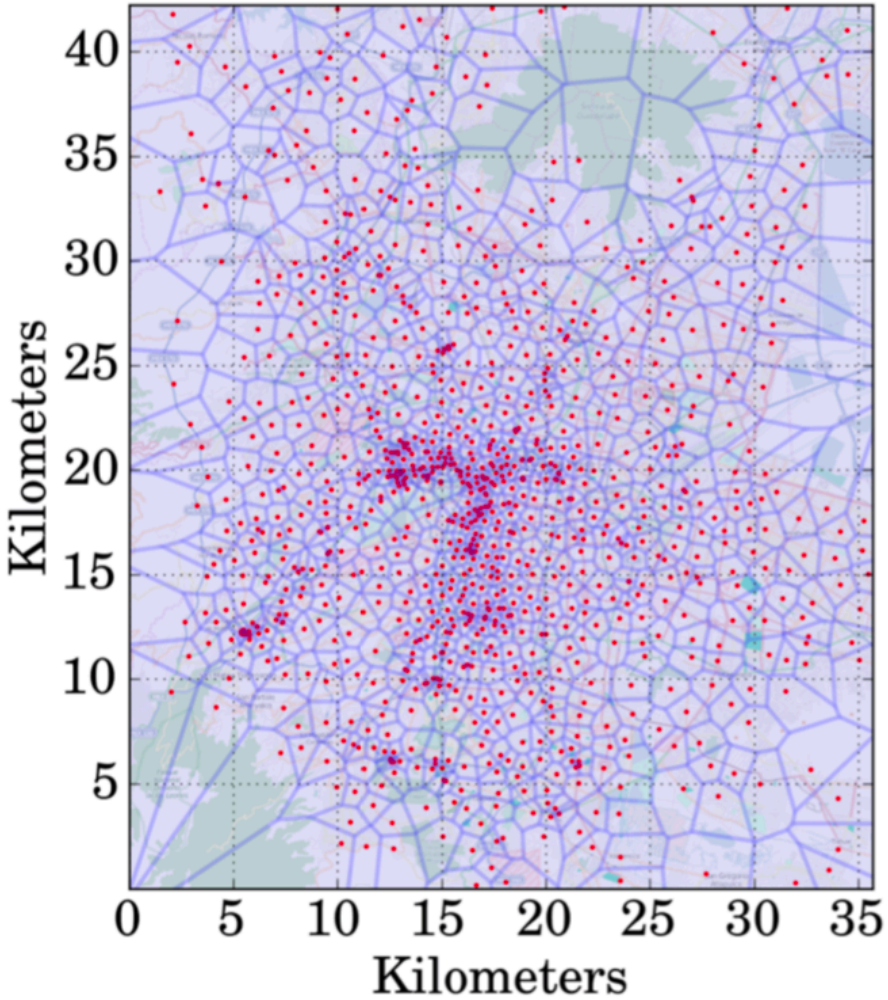
UserID	Event Time	Cell Tower	Caller	Callee	In/Out	Duration (s)
38DA6	2015-05-01 18:26:50	1921	38DA6	163B7	Out	52
78EC3	2015-05-01 14:16:09	2189	53808	78EC3	In	600
9FAFE	2015-05-01 23:20:09	2189	9FAFE	7BBF1	Out	41

Used in most of human mobility studies

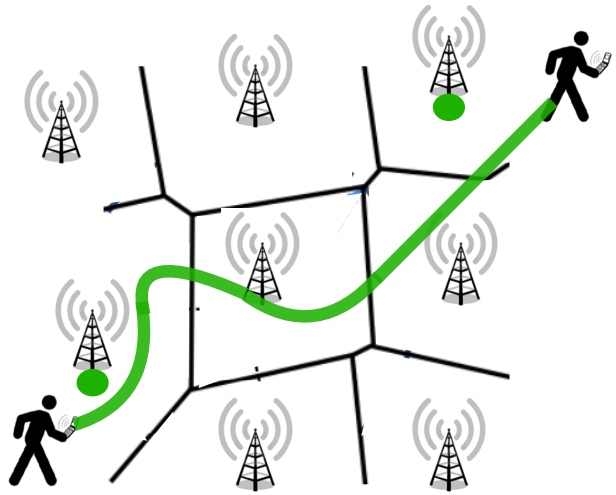
Large-scale and long-term mobility datasets

Large populations, large geographic regions, long periods of time

Call Detail Record (CDR) datasets: Beast or Beauty



BUT, sparse in space and irregularly distributed in time



Outcome: incomplete and imprecise trajectories

Inria
inventors for the digital world

Call Detail Record (CDR) datasets: Beast or Beauty

Studies focus on few highly active individuals only

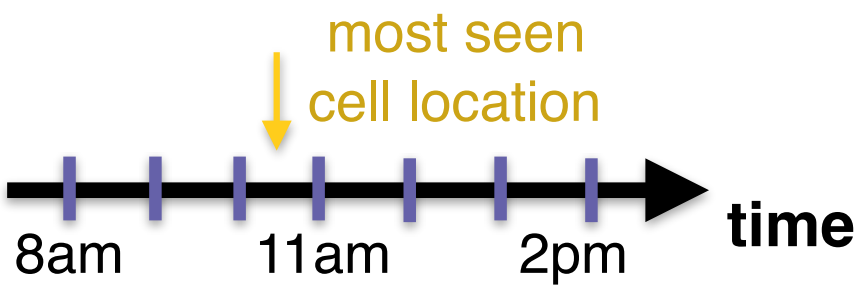
- User filtering is common : e.g., 0.45% in Song 2010 (out of 10 million)
- Waste a vast population with substantial mobility information
- Keep only very active-in-communication users

Partial reconstruction that leaves gaps in user mobility

- Potential biases due to missing locations are difficult to assess
- Spatiotemporal Interpolation, which are ineffective

Call Detail Record (CDR) datasets: Beast or Beauty

Time discretisation, i.e., increase time resolution



03

How we mitigate sparsity and irregularities?

- G. Chen, A. C. Viana, M. Fiore, C. Sarraute. **Individual Trajectory Reconstruction from Mobile Network Data.** HAL TR-0495, INRIA Saclay. June 2019. (under submission)

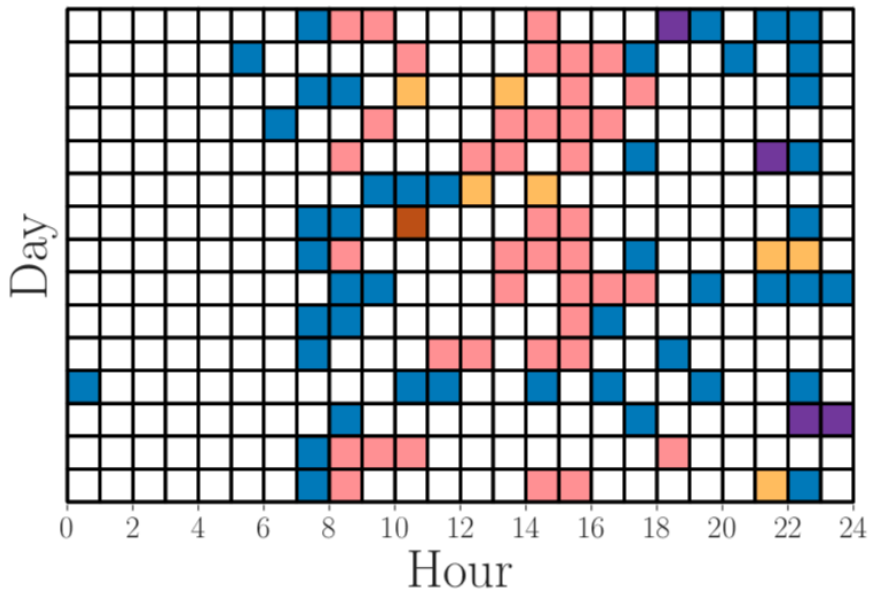
Trajectory reconstruction

Estimating a complete trajectory

reconstructed trajectory

$$\arg\min_{\hat{L}_{\mathcal{T}}} \frac{1}{|\mathcal{T}|} \sum_{i=1}^{N} \|\mathbf{l}_i - \hat{\mathbf{l}}_i\|,$$

original trajectory



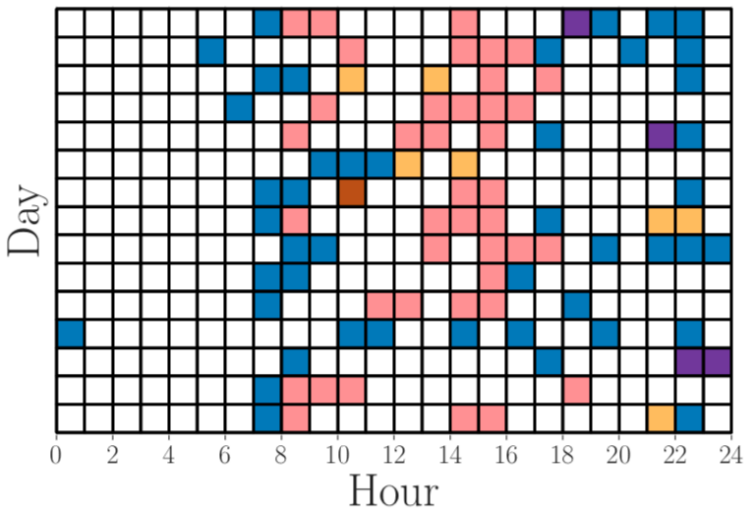
Context-enhanced Trajectory Reconstruction (CTR)

Allows completing all missing trajectory data with good accuracy

Leverages human mobility features

Context-enhanced Trajectory Reconstruction - CTR

completeness=0.3



1. Overnight invariance

2. Regularity of movement: in space and time

3. Pattern of long static phases with fast movements in between

Tendency to spend a substantial amount of time at a few fixed locations

Transitions among these locations are instead fairly rapid

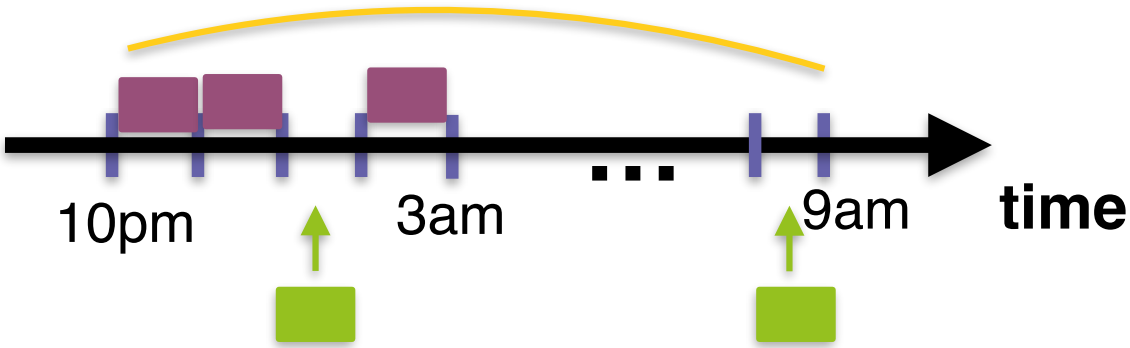
CTR - Full trajectory reconstruction

Nighttime reconstruction

- Most frequent location [10pm,9am]
- Home if in 80% of positions (over many days)

if location is unknown in [10pm,9am]: complete with home location

most seen (80% of time) cell location = **home** tag



CTR - Full trajectory reconstruction

Tensor Factorization (TF)

Exploit redundancies to recover missing data

BUT does not account for contextual specificities

It treats each dimension and each single data equally

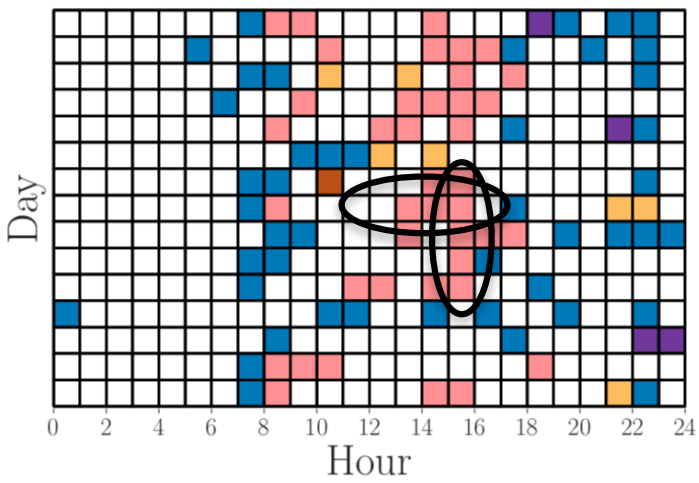
BUT human mobility patterns exhibit non-uniform importance of dimensions and values

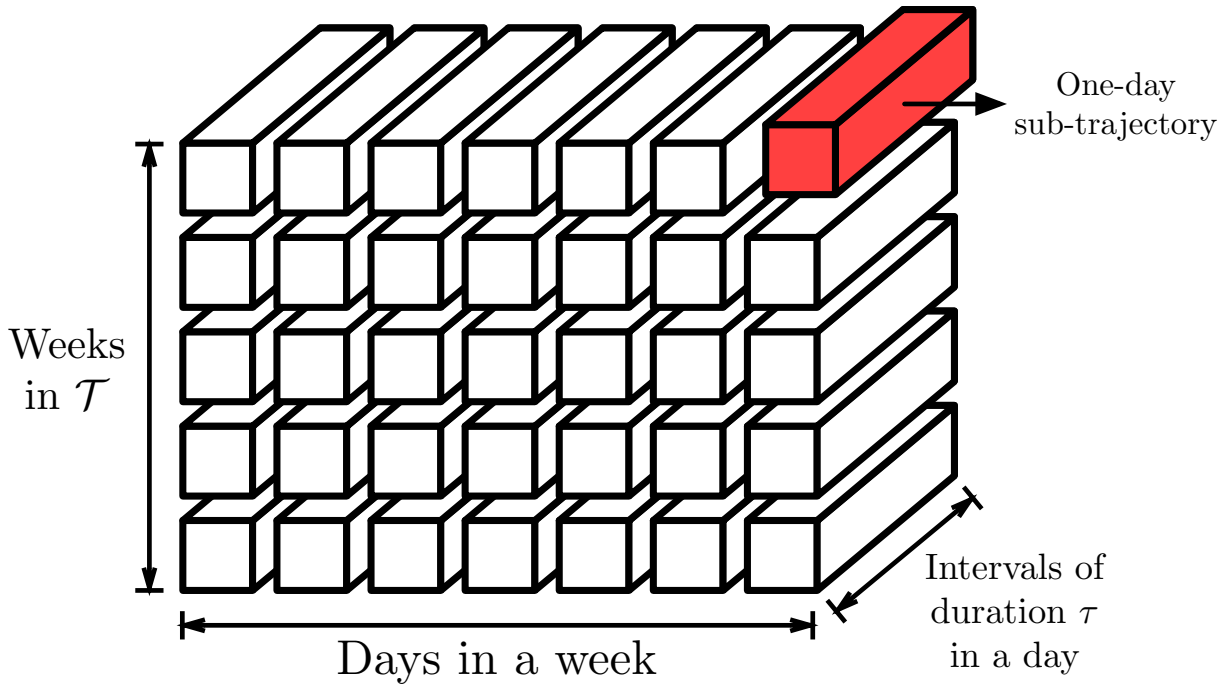
CTR - Full trajectory reconstruction

Customized Tensor Factorization

1) Reducing the diversity of locations at a same hour across days - repetitive patterns

2) Favouring similarity of locations in consecutive time intervals





CTR - Validation

Fraction of completeness

Ground-truth Internet session CDR (1,4k users, 2 weeks)

Sampling following CDR incomplete distributions (1,8 million users, 90 days)

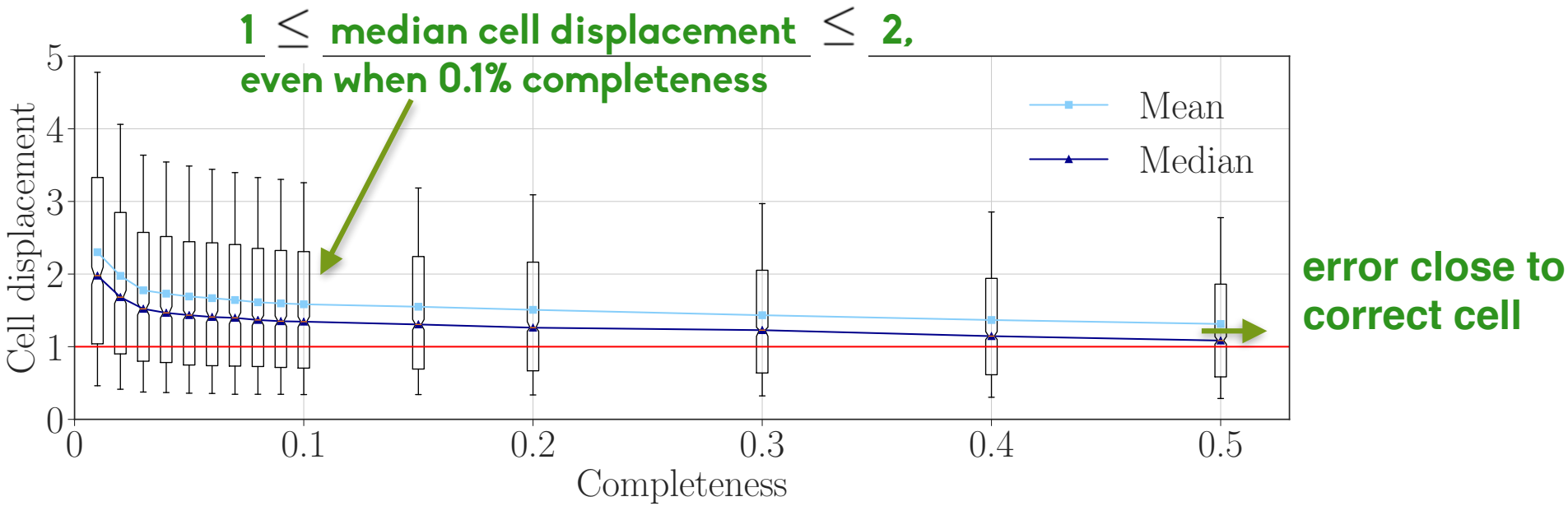


Figure 8 Cell displacement of locations inferred via CTR with respect to the actual ones, versus the completeness of the original trajectory. Candlesticks highlight the mean (light blue), median (dark blue), 25th and 75th percentiles (box), and 10th and 90th percentiles (errorbars). The horizontal line (red) highlights a one-cell displacement.

inventors for the digital world

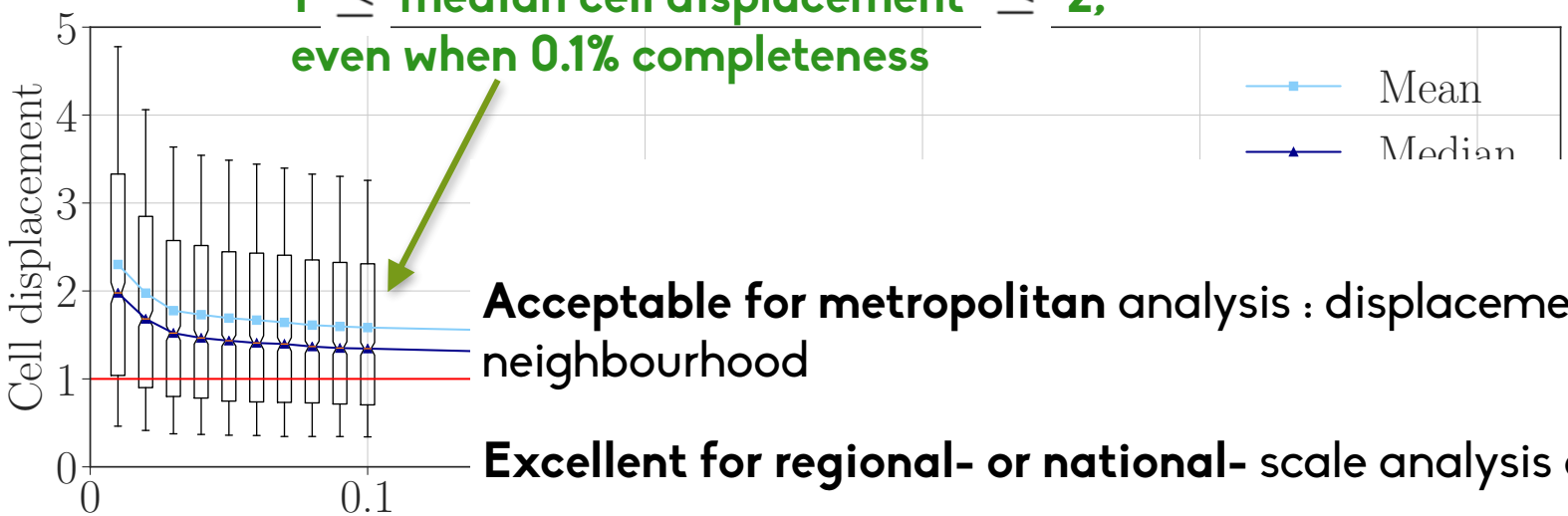
CTR - Validation

Ground-truth Internet session CDR (1,4k users, 2 weeks)

Fraction of completeness

Sampling following CDR incomplete distributions (1,8 million users, 90 days)

1 ≤ median cell displacement ≤ 2, even when 0.1% completeness



Acceptable for metropolitan analysis : displacement in nearby neighbourhood

Excellent for regional- or national- scale analysis of population

Figure 8 Cell displacement of locations inferred via CTR with respect to the actual ones, versus the completeness of the original trajectory. Candlesticks highlight the mean (light blue), median (dark blue), 25th and 75th percentiles (box), and 10th and 90th percentiles (errorbars). The horizontal line (red) highlights a one-cell displacement.

Revisiting Song et al. 2010's results

Conditions:

- about 8000 users, i.e., 0.45% out of 10 million users - **high filtering**
- **low** completeness : > 0.2, at least 0.5 calls/hour

Limits of Predictability in Human Mobility

Chaoming Song,[1,2] Zehui Qu,[1,2,3] Nicholas Blumm,[1,2] Albert-László Barabási[1,2]*

Science AAAS

Inria inventors for the digital world

Revisiting Song's predictability

Filtering favours individuals who are very active from a mobile communication and with higher mobility

i.e., users that are more difficult to anticipate

Larger population accounts for the vast majority of fairly static users, and reveals that people's movements are easier to anticipate

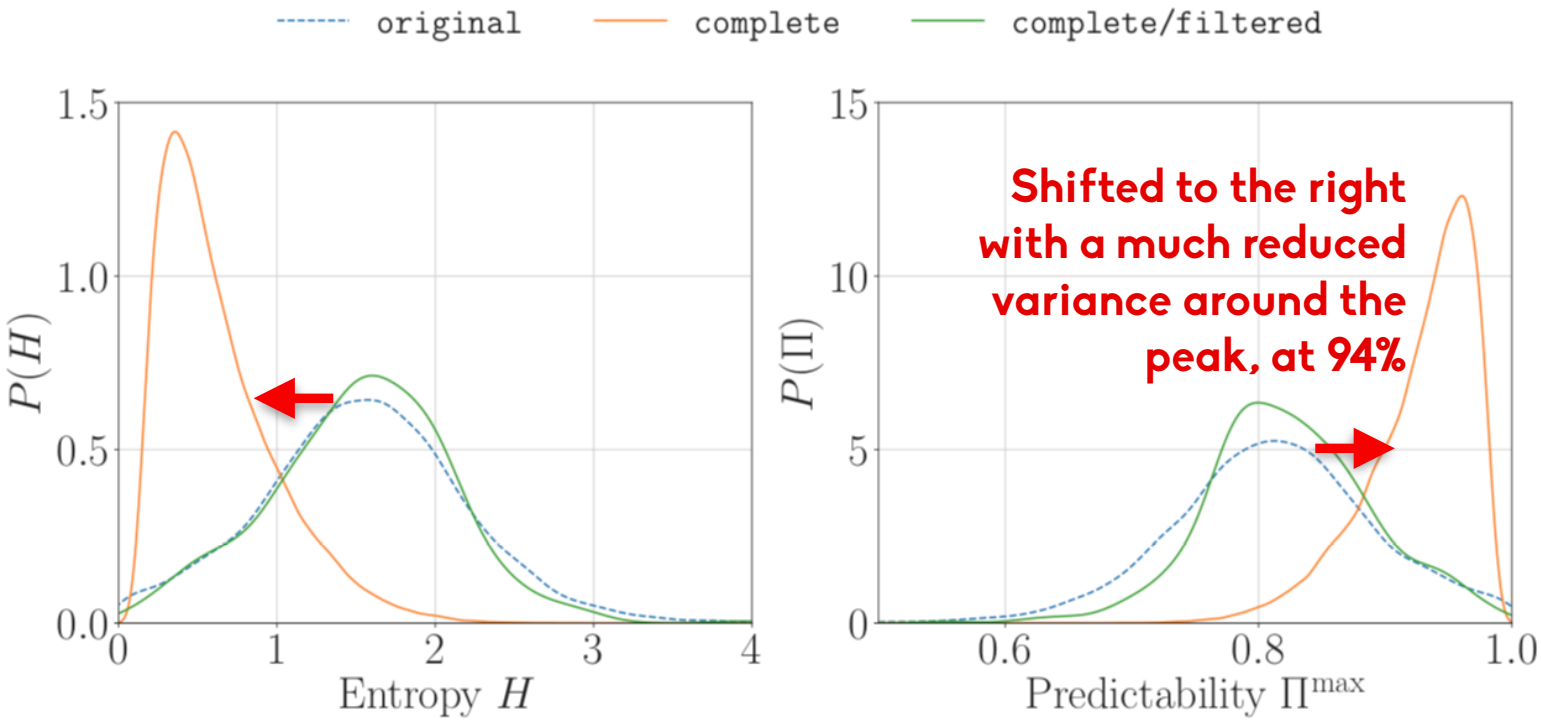


Figure 11 Revisited results from [4]. Distributions of the entropy and maximum theoretical predictability derived with (i) the incomplete original CDR-based trajectories of $8,000$ users filtered with the method in [4], (ii) the complete trajectories of all 1.7 million users, and (iii) the complete trajectories of the same $8,000$ users considered in the first case.

Revisiting De Montjoye et al. 2013's results

Conditions:

- 1.5 million users
- very **low** completeness : a user only has approximately **19 unique locations** observed in a month from 4 200 cell towers

Unique in the Crowd: The privacy bounds of human mobility

Yves-Alexandre de Montjoye[1,2], César A. Hidalgo[1,3,4], Michel Verleysen[2] & Vincent D. Blondel[2,5]

SCIENTIFIC REPORTS

Revisiting uniqueness in the crowd results

Sparsity brings fake uniqueness and hides diverse patterns of mobility

Completeness brings many similar users trajectories, what limit uniqueness

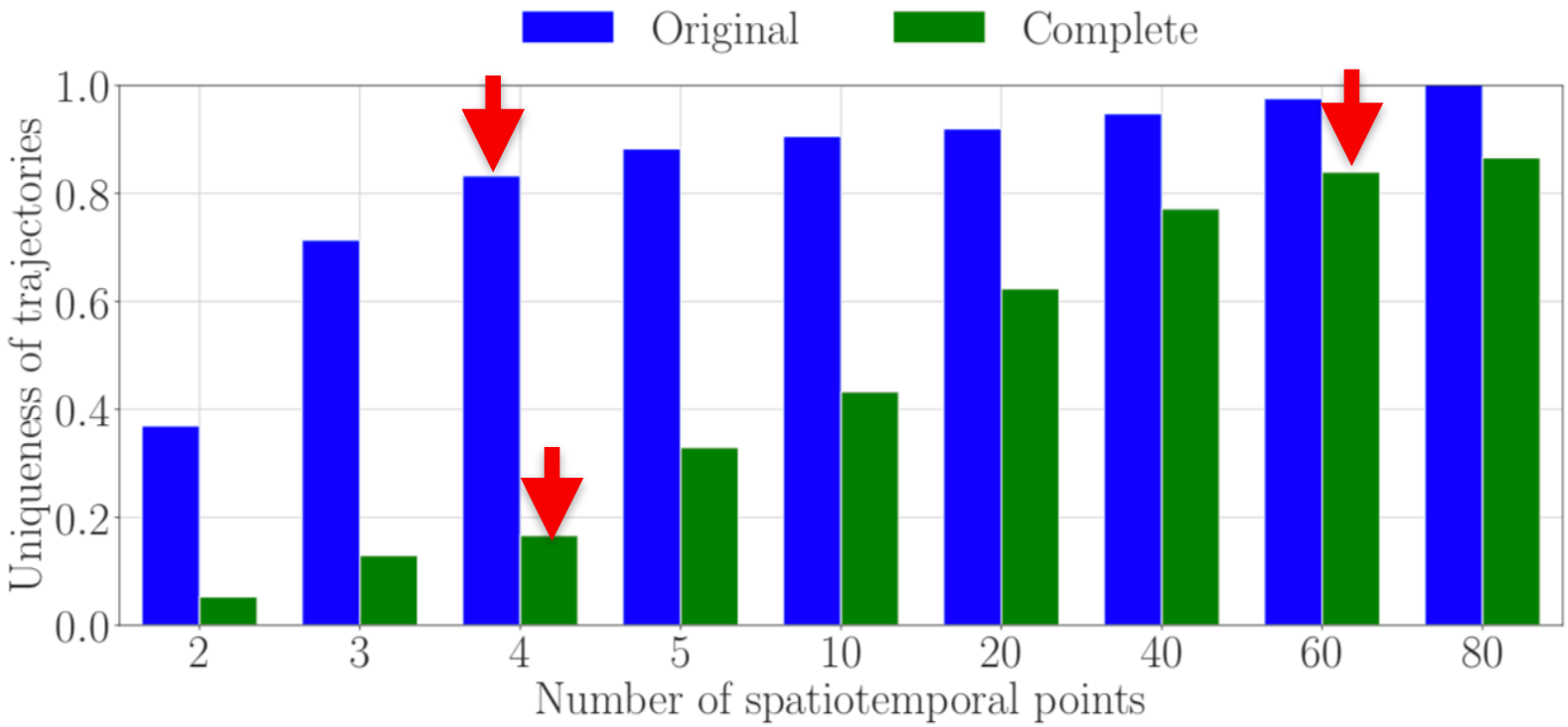


Figure 10 Revisited results from [5]. Ratio of trajectories with uniqueness $|U| = 1$ for a number of spatiotemporal points in the abscissa, for original (left) and complete (right) trajectories.

04

Habits or regularity of behaviors

Tell a lot about you and are the basis for prediction accuracy

Translates in regularity, repetitiveness and patterns (sequences of events)

Time matters, sometimes more than location

Contextual specificities bring new directions to behavior understanding

… and more

If mobility is concerned, carefully select the data source

E.g. short-term mobility and granular datasets (GPS data) **vs** long-term mobility and sparse datasets (CDR data)

Don't neglect data preliminaries process

Carefully reduce temporal/spatial resolutions

Completion is often required

Literature limitations bring back very interesting research questions

The basic laws governing daily human mobility **remains still limited!!!**

Merci beaucoup!

You are **Human Beings!**

So think as one while leveraging behaviors!

Revisiting Gonzalez Nature 2008's results

Conditions:

- **1.67%** out of 6 million users **- high filtering**
- very **low** completeness : < 1 location per user per day

Understanding individual human mobility patterns

Marta C. González[1], César A. Hidalgo[1,2] & Albert-László Barabási[1,2,3]

inventors for the digital world

Aggregated movement

Sparsity risks to substantially overestimate the region within which the Levy flight behavior of human mobility occurs, especially for users who travel long distances

Individual movements

The truncated power-law behavior found by Gonzalez et al. is confirmed

Original: $\beta_{\Delta_r} = 1.77 \quad \beta_{r_g} = 1.65$

Complete: $\beta_{\Delta_r} = 1.78$ and $\beta_{r_g} = 1.68$

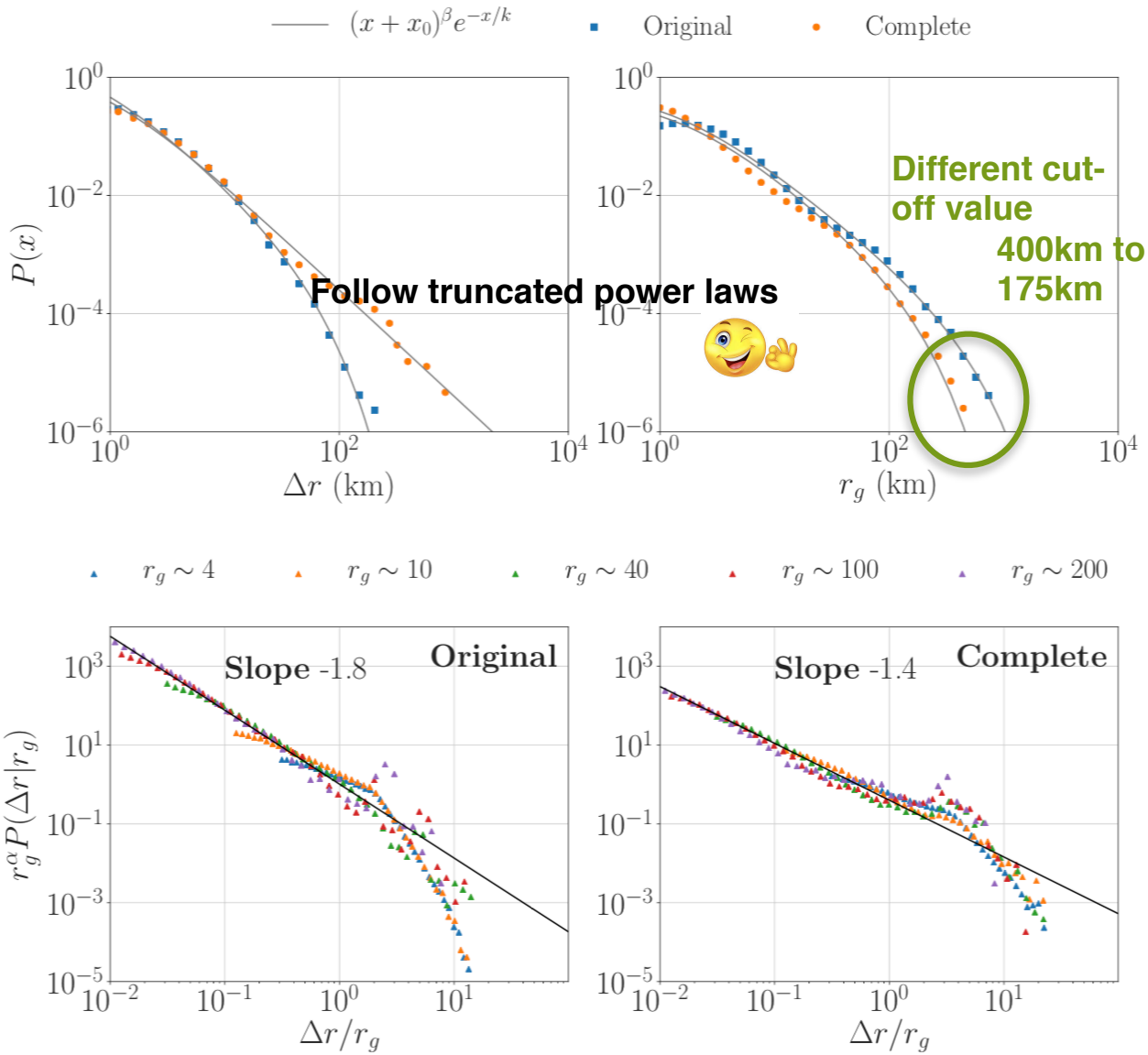


Figure 9 Revisited results from [3]. Distributions of travel distance Δr, radius of gyration r_g, and conditional travel distance Δr (with respect to $\Delta r/r_g$) for original and complete trajectories.