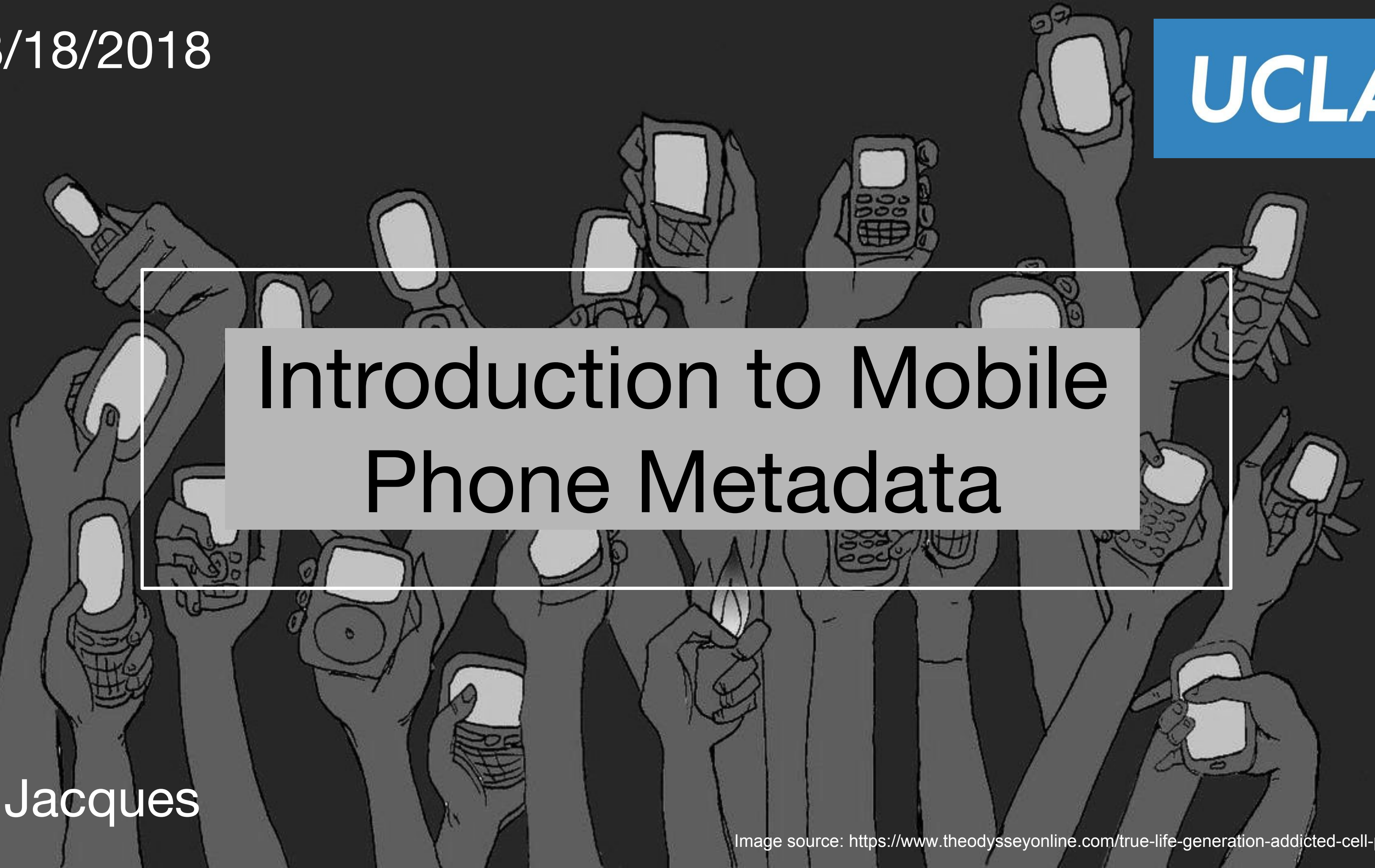
08/18/2018



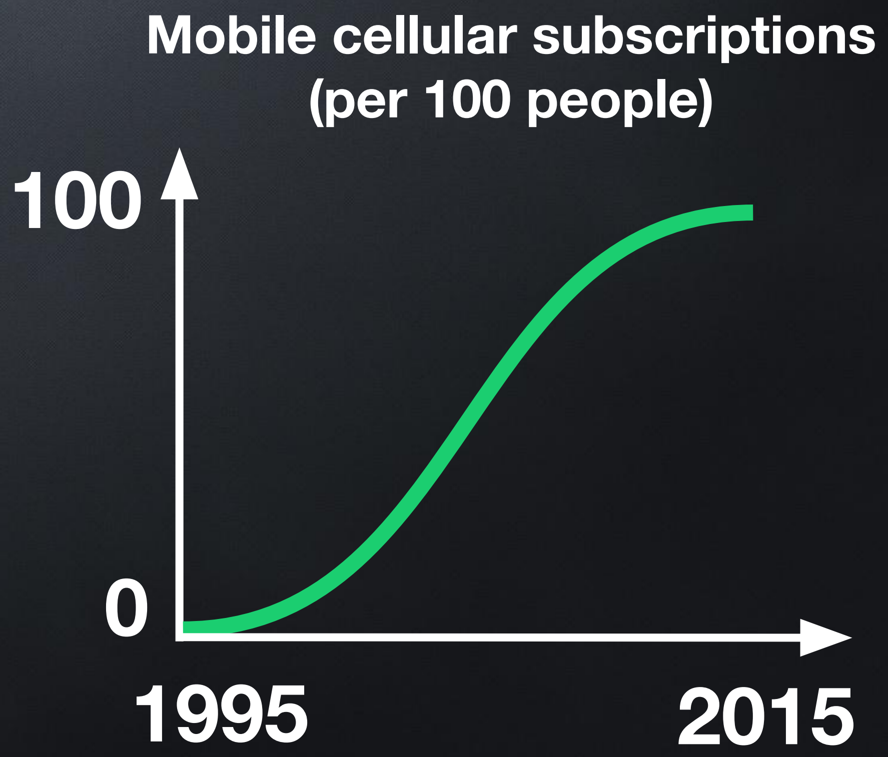
Introduction to Mobile Phone Metadata

D. Jacques

UCLA

Image source: https://www.theodysseyonline.com/true-life-generation-addicted-cell-phones

NO TECHNOLOGY HAS EVER SPREAD FASTER AROUND THE WORLD



International Telecommunication Union

SMARTPHONE (19% ownership in Senegal, 2015)

Computer, accelerometer, digital compass, gyroscope, GPS, microphone, camera…

-> **PERSONAL SENSING**

BASIC HANDSET

When, how, from where, with whom we communicate.

-> **DIGITAL TRACES** (CDRs, mobile money, top-up…)

Essentials of
MOBILE CELLULAR NETWORK

Base Transceiver Stations, grouped in Location Areas, are the basic units of a mobile network

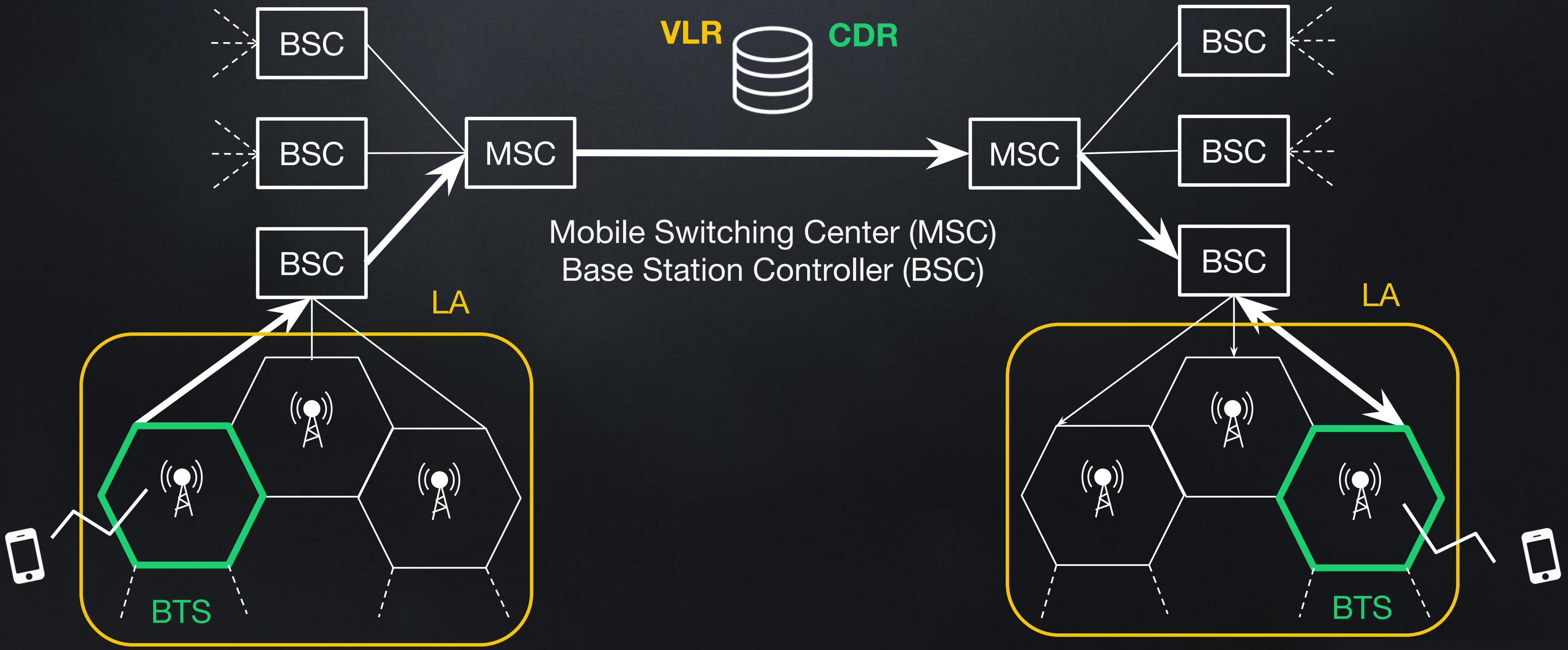
Base transceiver station (BTS)

Location Area (LA)

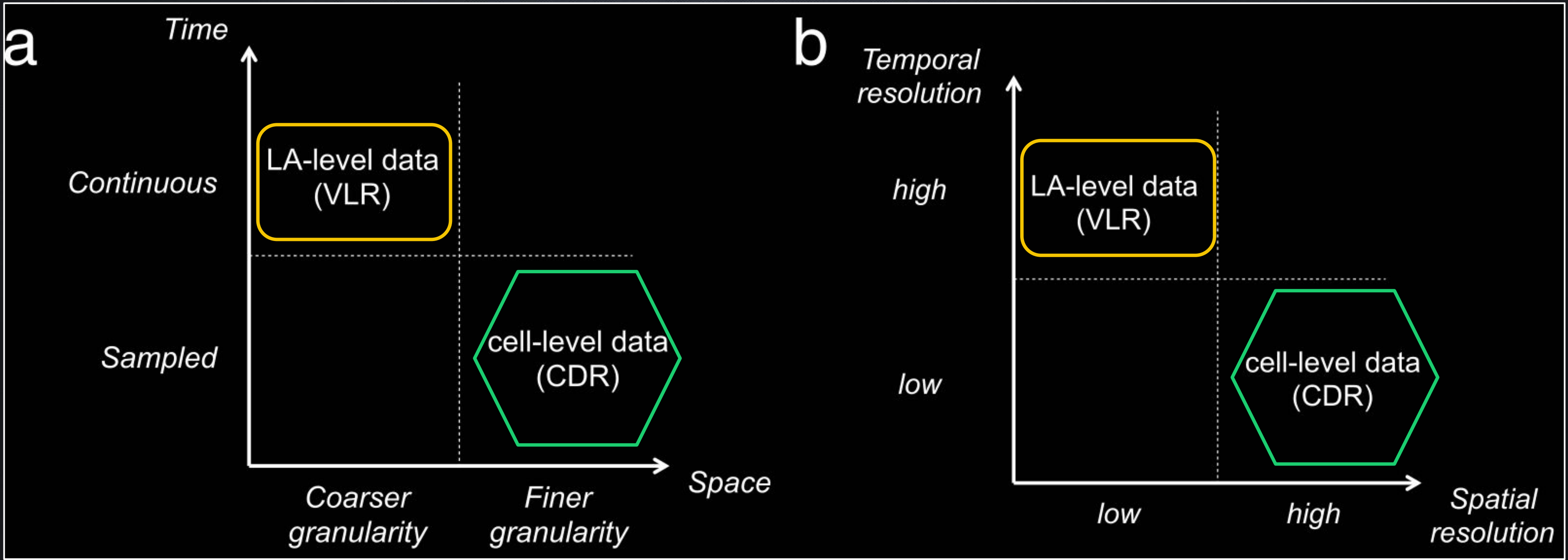
-> Visitor Location Register (VLR)

-> Call Detail Records (CDR)

The MNO queries the **VISITOR LOCATION REGISTER** (VLR) to locate the LA callee and stores the BTS location (+timestamp) of caller and callee in **CALL DETAIL RECORDS** (CDR) for billing purpose.



CDR has higher **spatial resolution** but lower **temporal resolution** than **VLR**



From: F. Ricciato, et al., Beyond the "single operator, CDR-only" paradigm: An interoperable framework for mobile network data analyses and population density, Pervasive and Mobile Computing (2016).

Most of the studies use of **CDR** because it is easier to access the data.

	CDR data	VLR data
Spatial resolution	**High (BTS level)**	Low (LA level)
Temporal resolution	Low	**Very High**
MS coverage/sampling	Possibly low	**Complete**
Risk of bias	High	**Low**
Data type	**Static**	Dynamic
Query Model	**Off-line or on-line**	On-line only

From: F. Ricciato, et al., Beyond the "single operator, CDR-only" paradigm: An interoperable framework for mobile network data analyses and population density, Pervasive and Mobile Computing (2016).

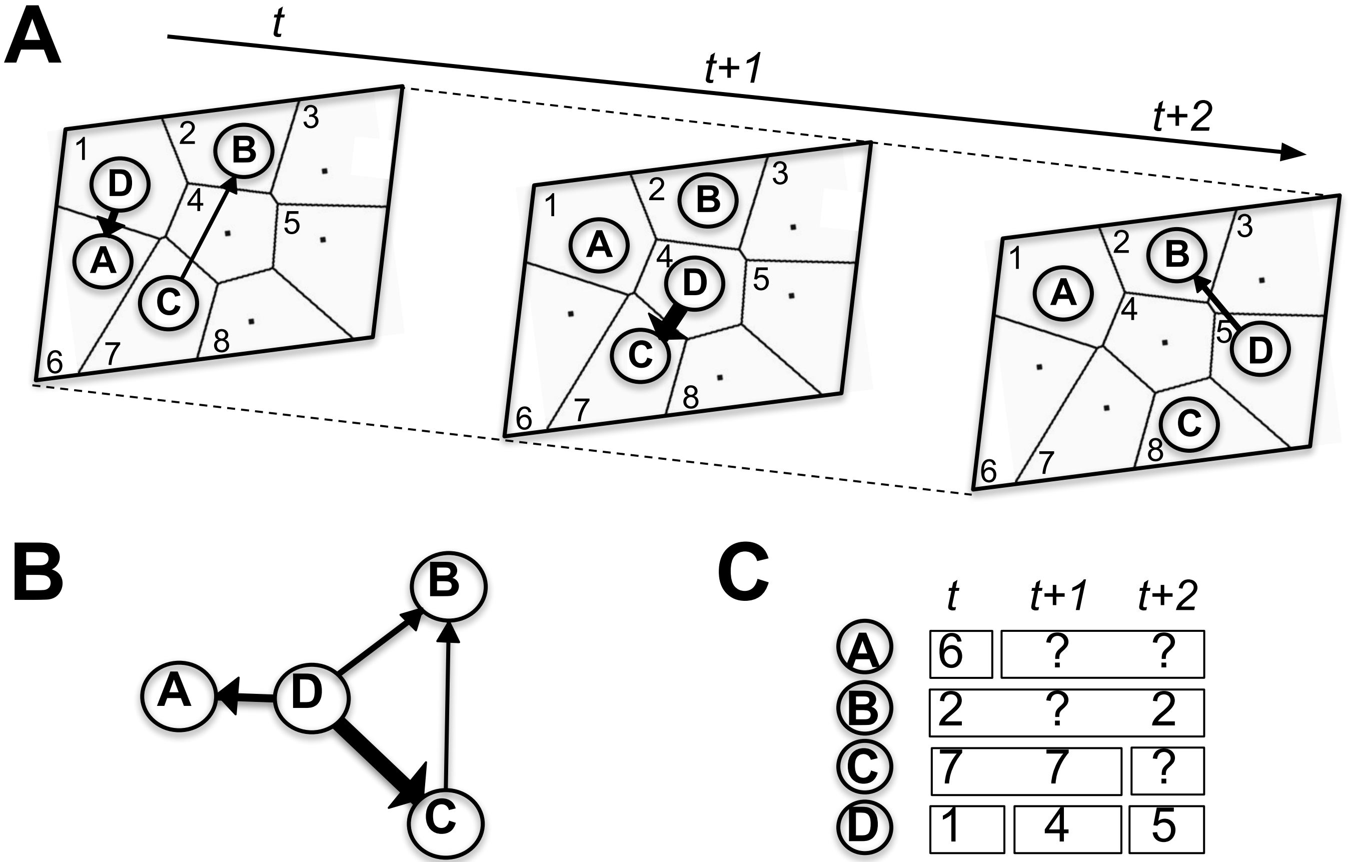
CALL DATA RECORDS looks like:

Who? Where? When? How long?

Caller SIM	Callee SIM	Outgoing BTS	Incoming BTS	Timestamp	Call duration (sec)
~~0458685984~~	0488595496	12	365	2018-01-18 15:22:12	456
0458685984	0458685984	12	25	2018-01-18 22:24:12	35
0469875254	0498563201	879	567	2018-01-19 08:47:10	125
(...)	(...)	(...)	(...)	(...)	(...)

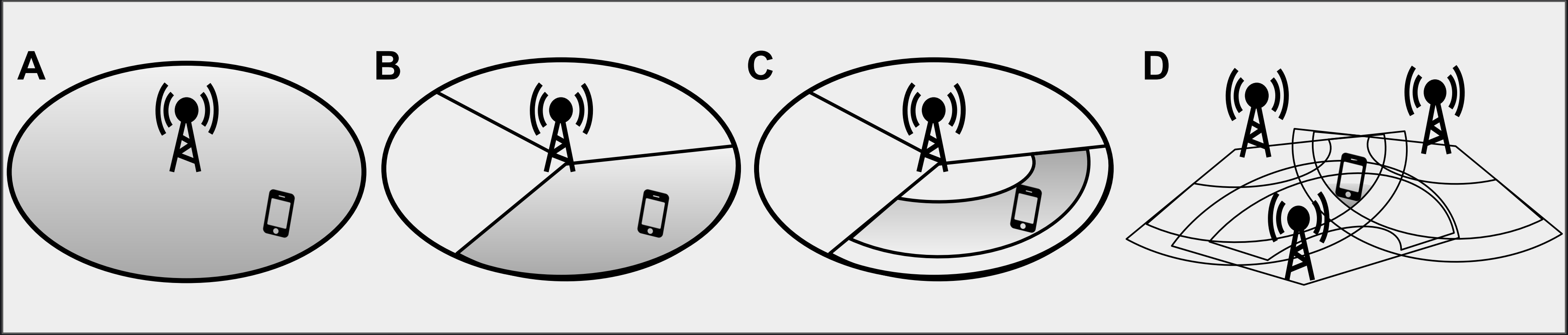
Pseudo-anonymization to preserve privacy

Geospatial, Dynamic, Directed Weighted Network



GEOSPATIAL DIMENSION

(A) Base station level
(B) Sector level (antenna)
(C) B + signal characteristics
(D) Triangulation



Limitations of
CALL DATA RECORDS

THE **BEST-CASE SCENARIO** FOR POPULATION MAPPING USING MOBILE PHONE DATA

$$population_{it} = f(SIM_{it})$$

with

$$f_i() = f_j()$$

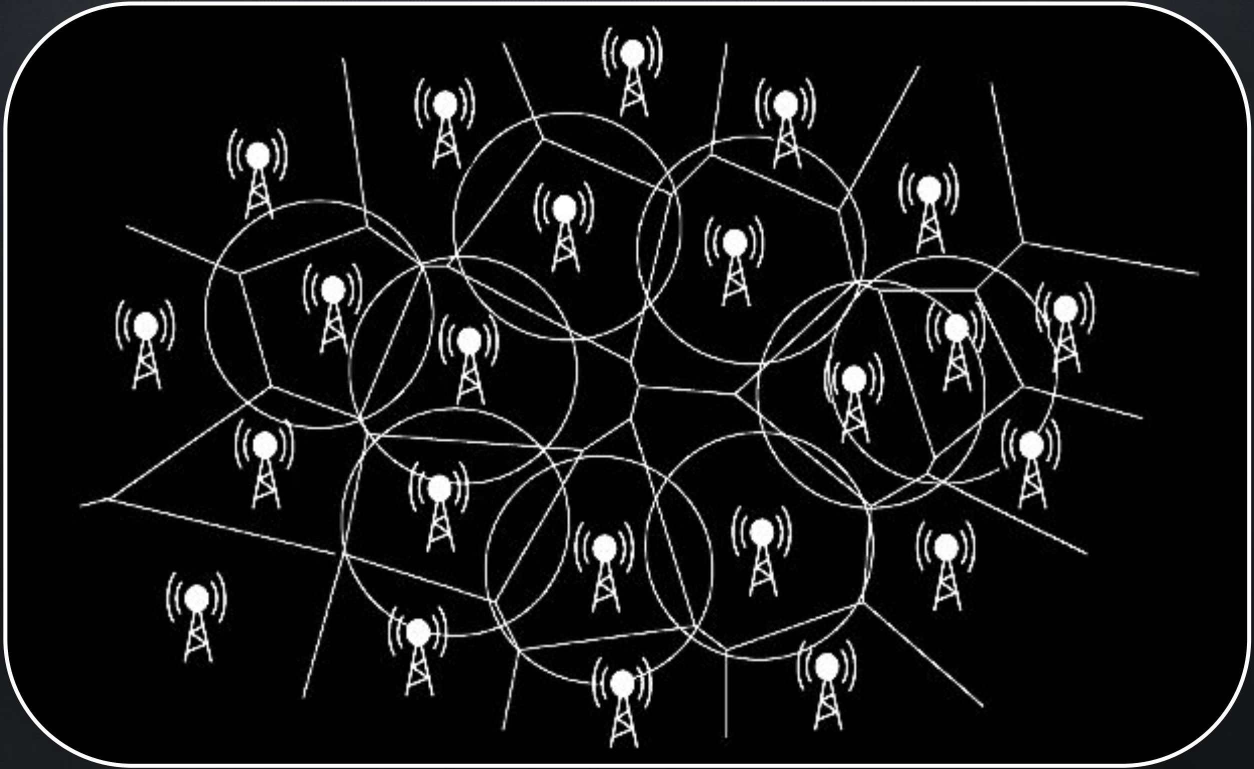
INDEPENDENT OF SPACE

$$f_t() = f_{t+1}()$$

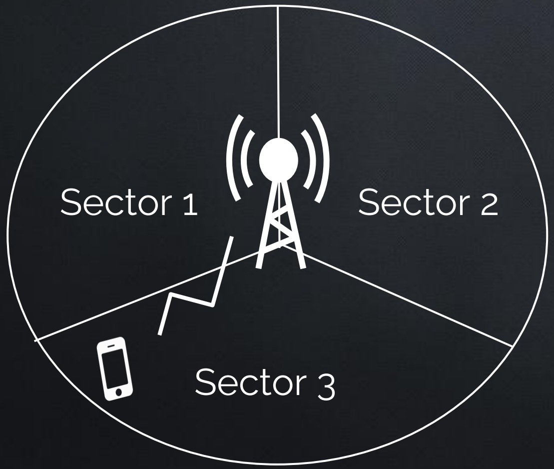
INDEPENDENT OF TIME

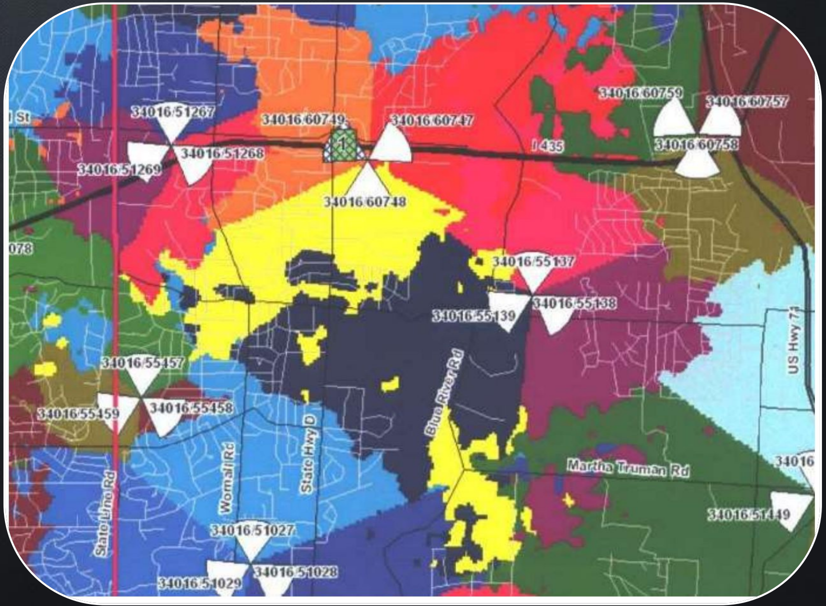
And *SIM* known at fine spatial and temporal granularity

Coverage of each antenna is usually approximated with **VORONOI/THIESSEN POLYGONS**



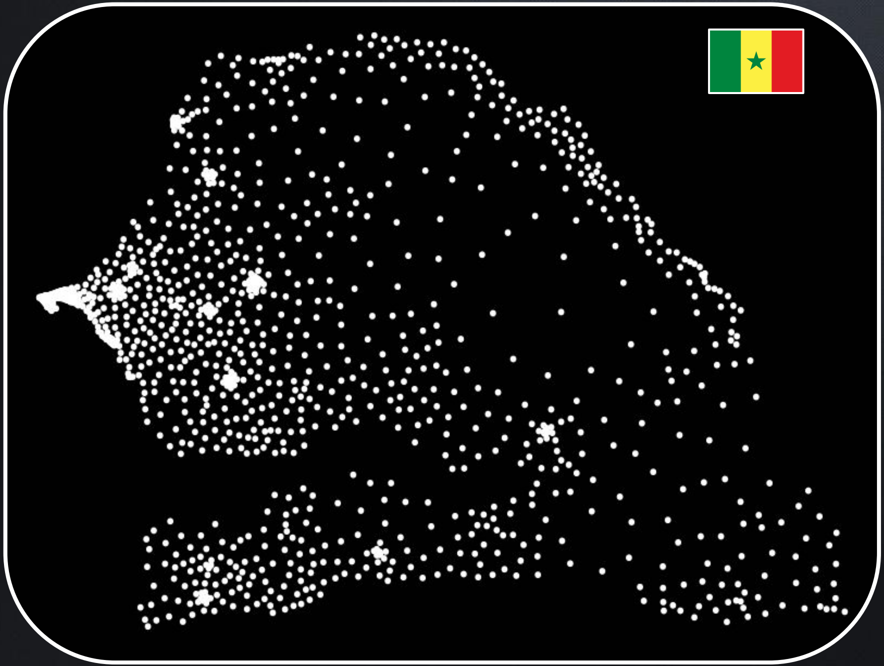
L A R G E difference between theoretical & actual antenna footprint

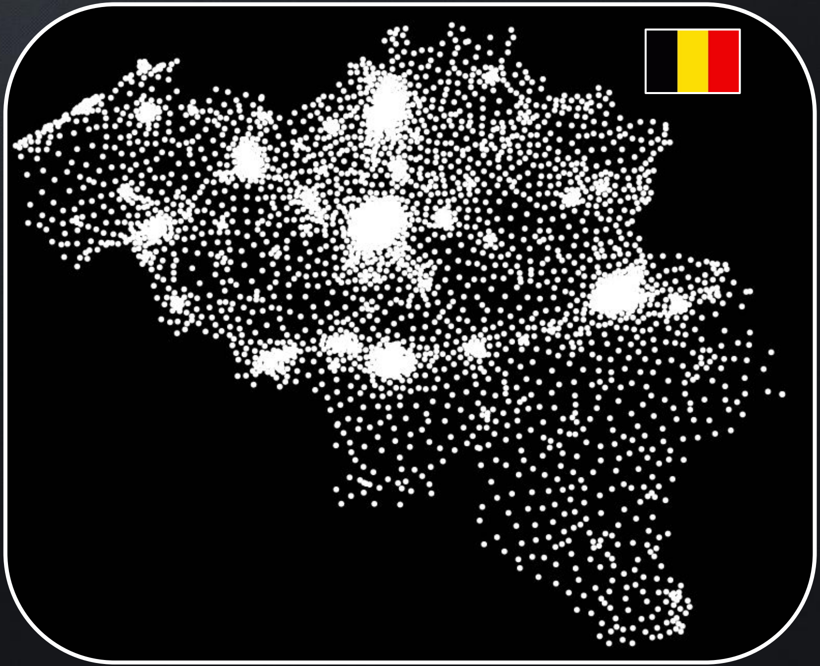




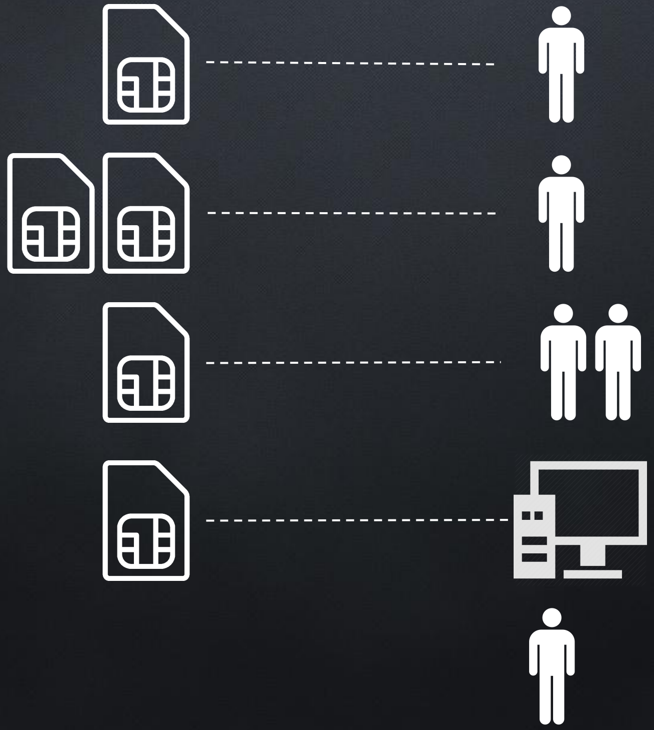
From:https://viewfromll2.com/2015/01/12/serial-the-failure-of-the-prosecutions-cellphone-theory-in-one-simple-chart/

The DENSITY OF ANTENNA (~spatial resolution) is **higher** in urban areas





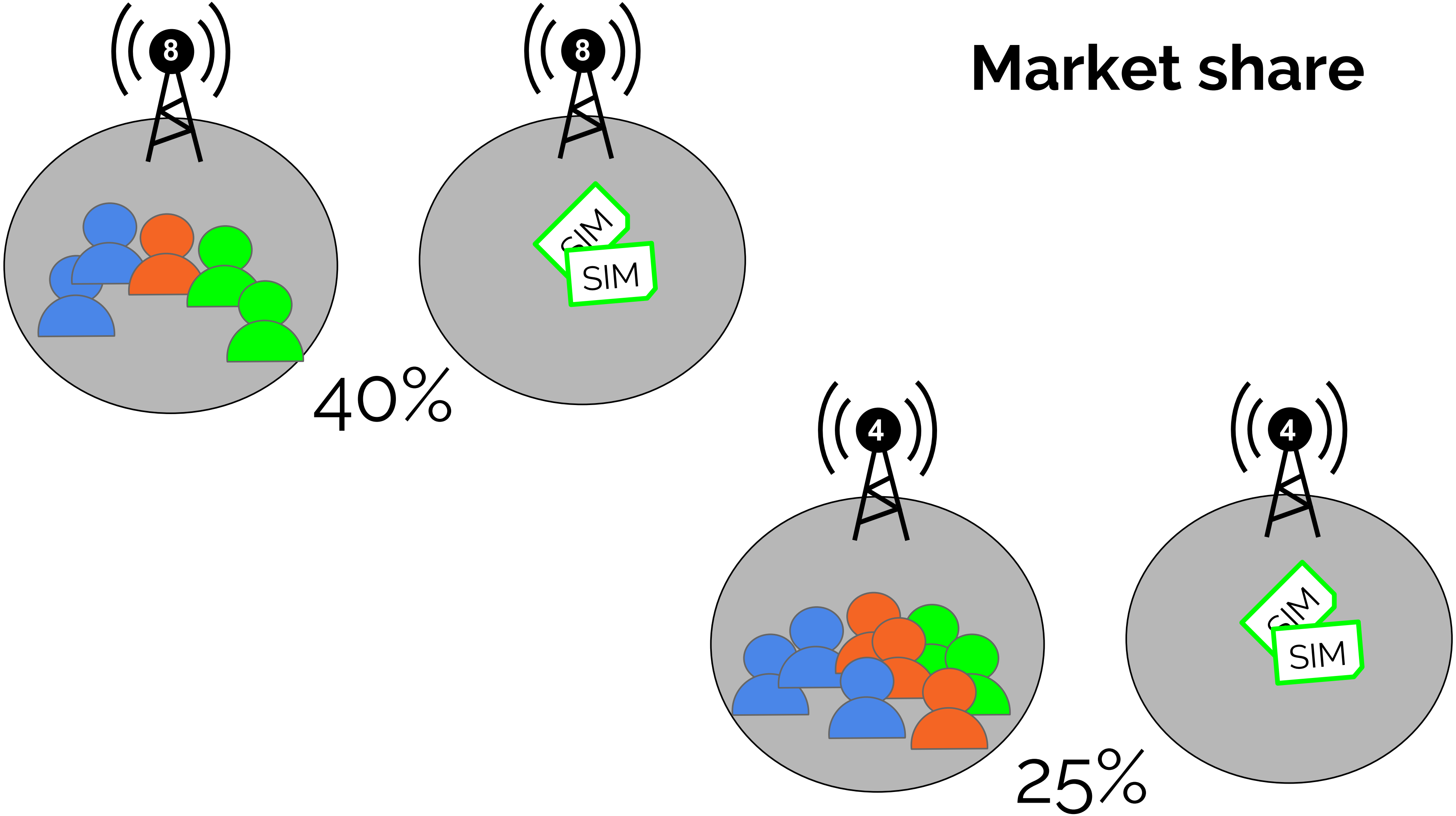
SIM CARD ≠ PERSON



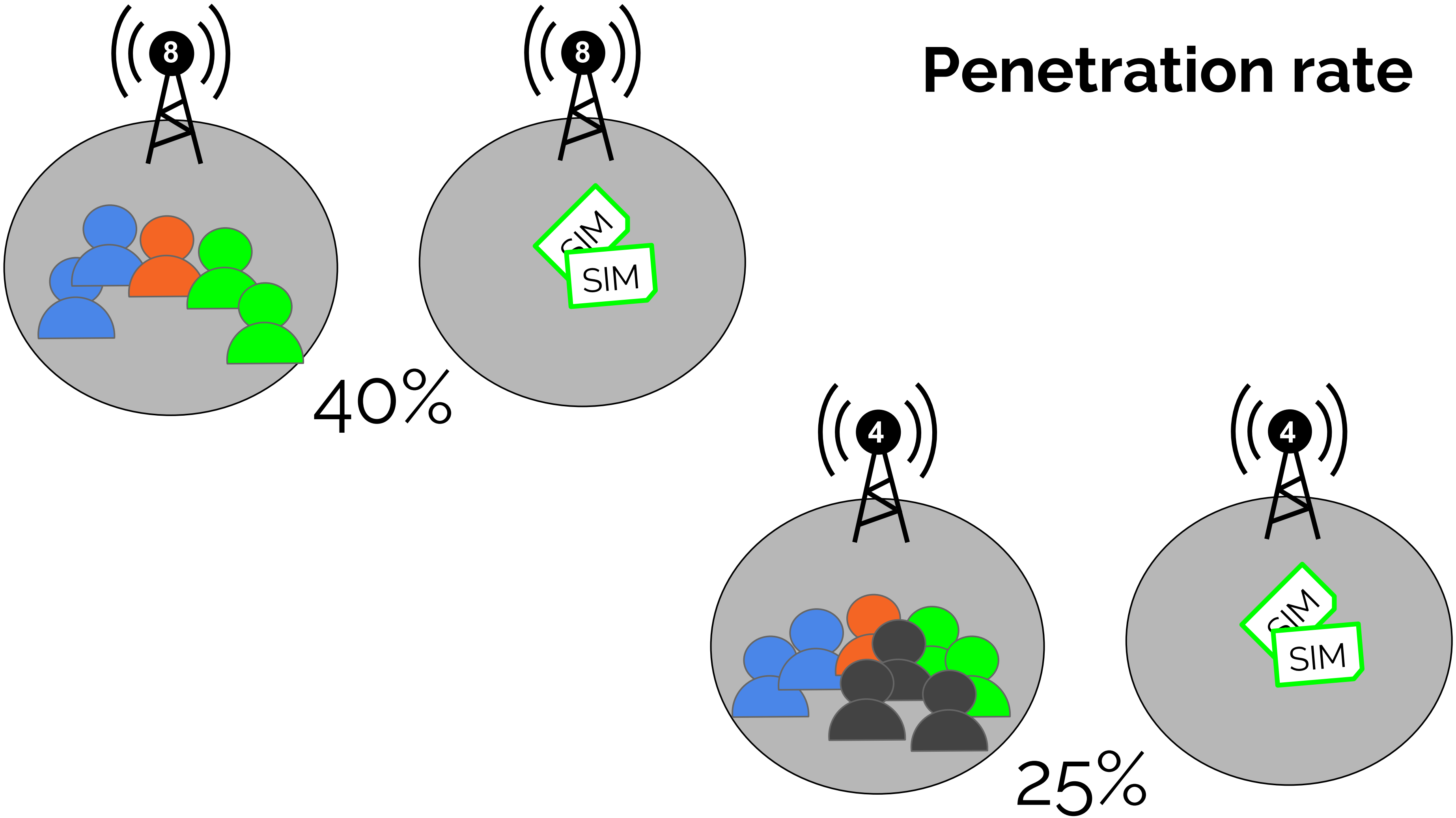
1:1	ideal case
Many : 1	work / private
1 : Many	family sharing
1 : 0	M2M
0 : 1	penetration rate/market share

Adapted from: Ricciato, Fabio, et al. Estimating population density distribution from network-based mobile phone data. Publications Office of the European Union, 2015.

Market share



Penetration rate



MP OWNERSHIP is biased towards wealthier, educated, urban and predominantly male individuals

Mobile Divides: Gender, Socioeconomic Status, and Mobile Phone Use in Rwanda 2010

Joshua Blumenstock
U.C. Berkeley School of Information
Berkeley, CA 94720
jblumenstock@berkeley.edu

Nathan Eagle
The Santa Fe Institute
Santa Fe, NM 87501
nathan@mit.edu

ABSTRACT

We combine data from a field survey with transaction log data from a mobile phone operator to provide new insight into daily patterns of mobile phone use in Rwanda. The analysis is divided into three parts. First, we present a statistical comparison of the general Rwandan population to the population of mobile phone owners in Rwanda. We find that phone owners are considerably wealthier, better educated, and more predominantly male than the general population. Second, we analyze patterns of phone use and access, based on self-reported survey data. We note statistically significant differences by gender; for instance, women are more likely to use shared phones than men. Third, we perform a quantitative analysis of calling patterns and social network structure using mobile operator billing logs. By these measures, the differences between men and women are more modest, but we observe vast differences in utilization between the relatively rich and the relatively poor. Taken together, the evidence in this paper suggests that phones are disproportionately owned and used by the privileged strata of Rwandan society.

MP OWNERSHIP is biased towards wealthier, educated, urban and predominantly male individuals

Heterogeneous Mobile Phone Ownership and Usage Patterns in Kenya

2012

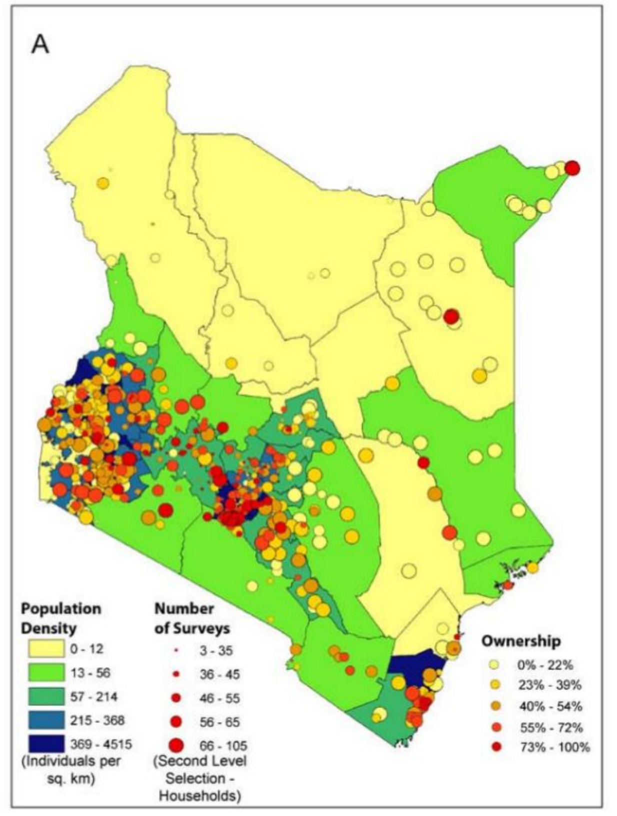
Amy Wesolowski[1], Nathan Eagle[2], Abdisalan M. Noor[3,4], Robert W. Snow[4], Caroline O. Buckee[2,5]*

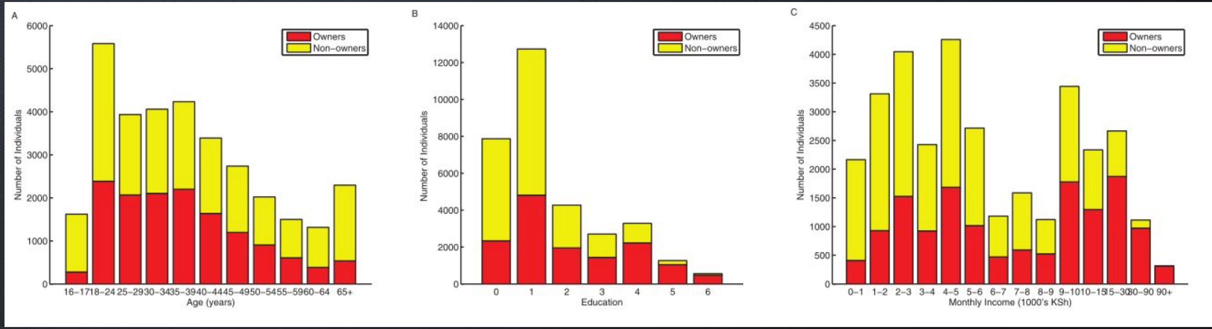
Abstract

The rapid adoption of mobile phone technologies in Africa is offering exciting opportunities for engaging with high-risk populations through mHealth programs, and the vast volumes of behavioral data being generated as people use their phones provide valuable data about human behavioral dynamics in these regions. Taking advantage of these opportunities requires an understanding of the penetration of mobile phones and phone usage patterns across the continent, but very little is known about the social and geographical heterogeneities in mobile phone ownership among African populations. Here, we analyze a survey of mobile phone ownership and usage across Kenya in 2009 and show that distinct regional, gender-related, and socioeconomic variations exist, with particularly low ownership among rural communities and poor people. We also examine patterns of phone sharing and highlight the contrasting relationships between ownership and sharing in different parts of the country. This heterogeneous penetration of mobile phones has important implications for the use of mobile technologies as a source of population data and as a public health tool in sub-Saharan Africa.

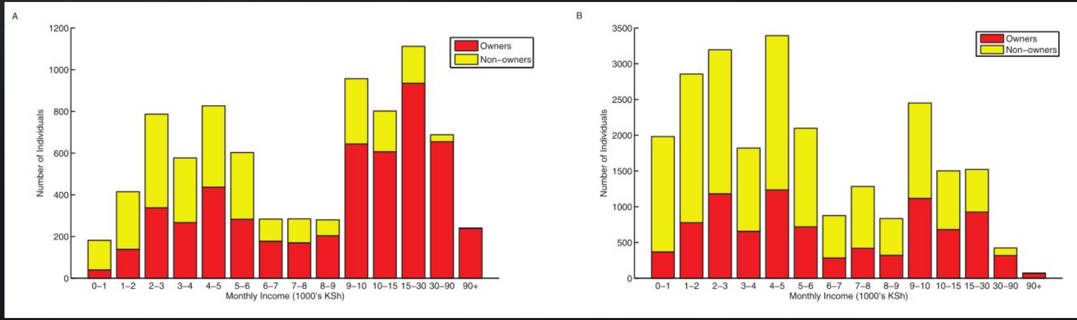
MP OWNERSHIP is biased towards wealthier, educated, urban and predominantly male individuals

OWNER / NON-OWNER





AGE EDUCATION INCOME



INCOME (URBAN) INCOME (RURAL)

But MOBILITY estimates is robust to biases in phone ownership

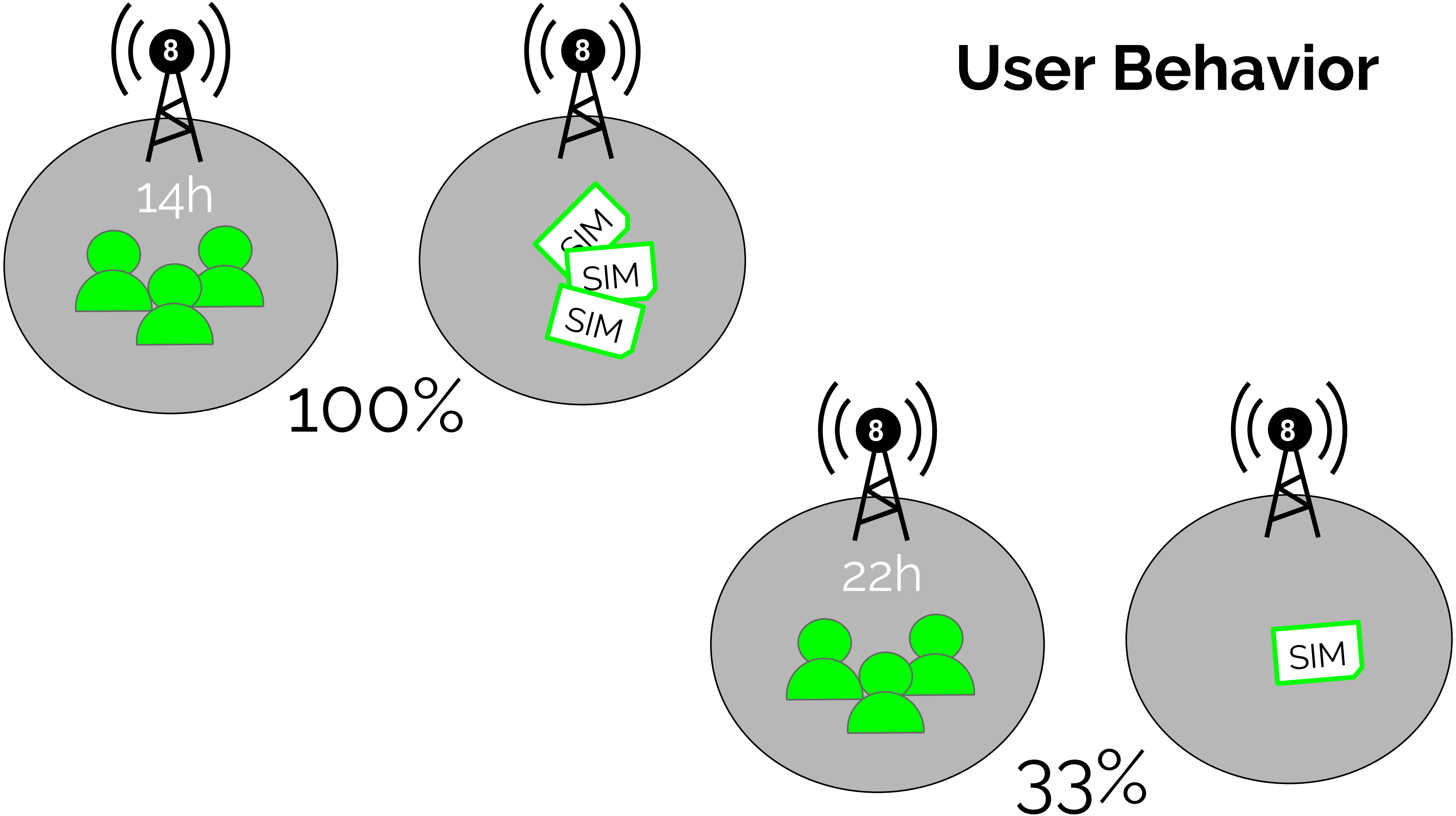
The impact of biases in mobile phone ownership on estimates of human mobility

2013

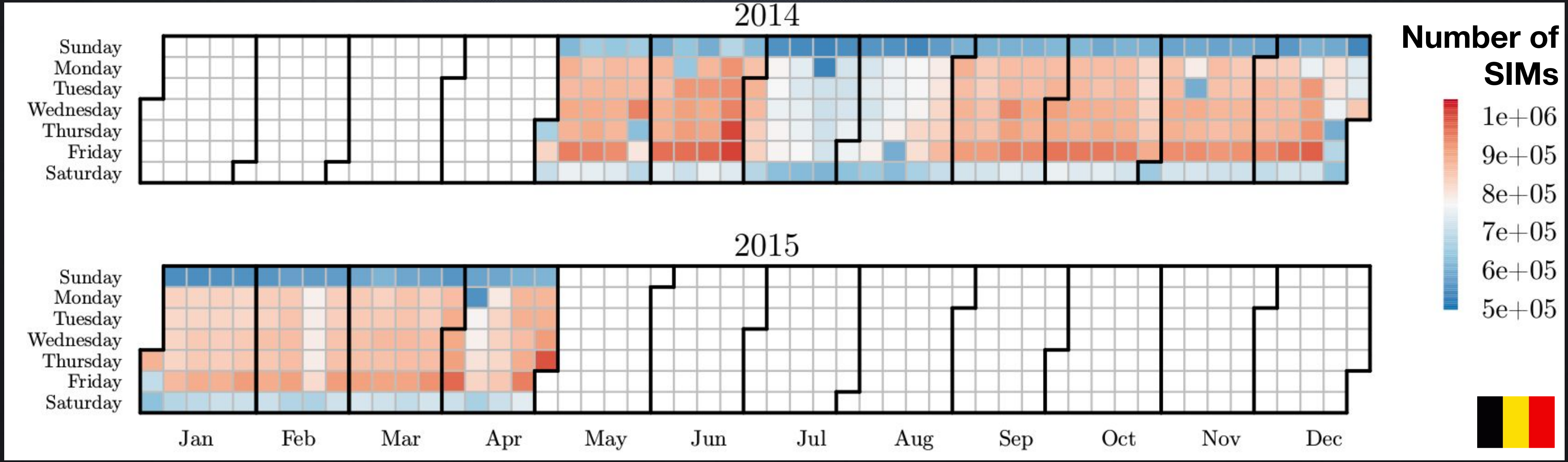
Amy Wesolowski[1], Nathan Eagle[2,3], Abdisalan M. Noor[4,5], Robert W. Snow[4,5] and Caroline O. Buckee[3,6]

Mobile phone data are increasingly being used to quantify the movements of human populations for a wide range of social, scientific and public health research. However, making population-level inferences using these data is complicated by differential ownership of phones among different demographic groups that may exhibit variable mobility. Here, we quantify the effects of ownership bias on mobility estimates by coupling two data sources from the same country during the same time frame. We analyse mobility patterns from one of the largest mobile phone datasets studied, representing the daily movements of nearly 15 million individuals in Kenya over the course of a year. We couple this analysis with the results from a survey of socioeconomic status, mobile phone ownership and usage patterns across the country, providing regional estimates of population distributions of income, reported airtime expenditure and actual airtime expenditure across the country. We match the two data sources and show that mobility estimates are surprisingly robust to the substantial biases in phone ownership across different geographical and socioeconomic groups.

User Behavior

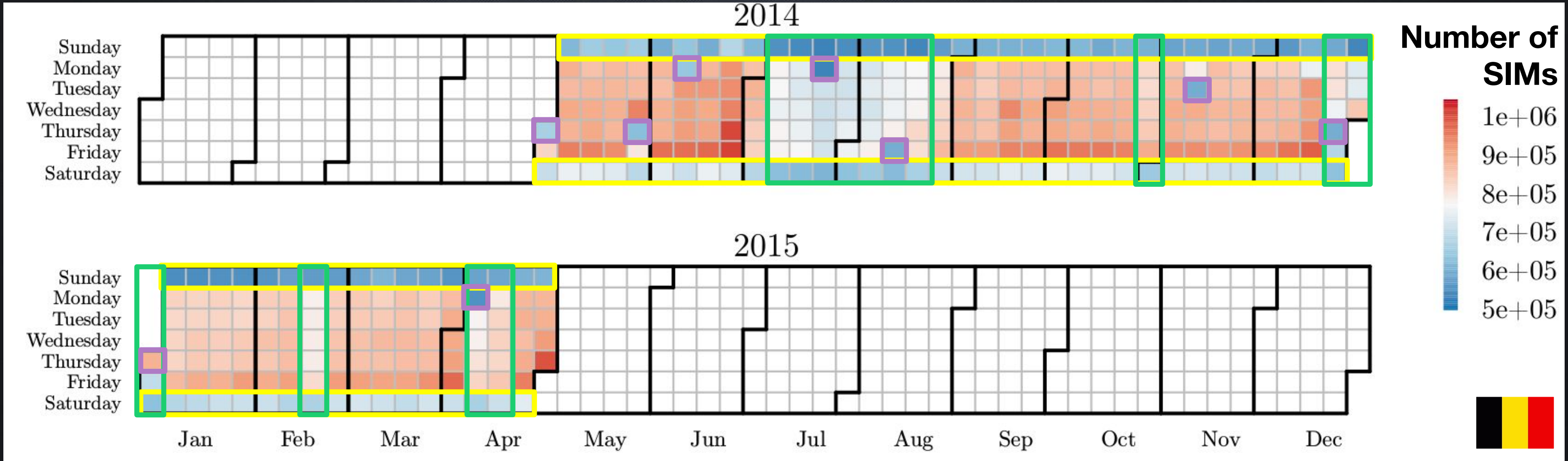


MOBILE PHONE USAGE is not independent of TIME

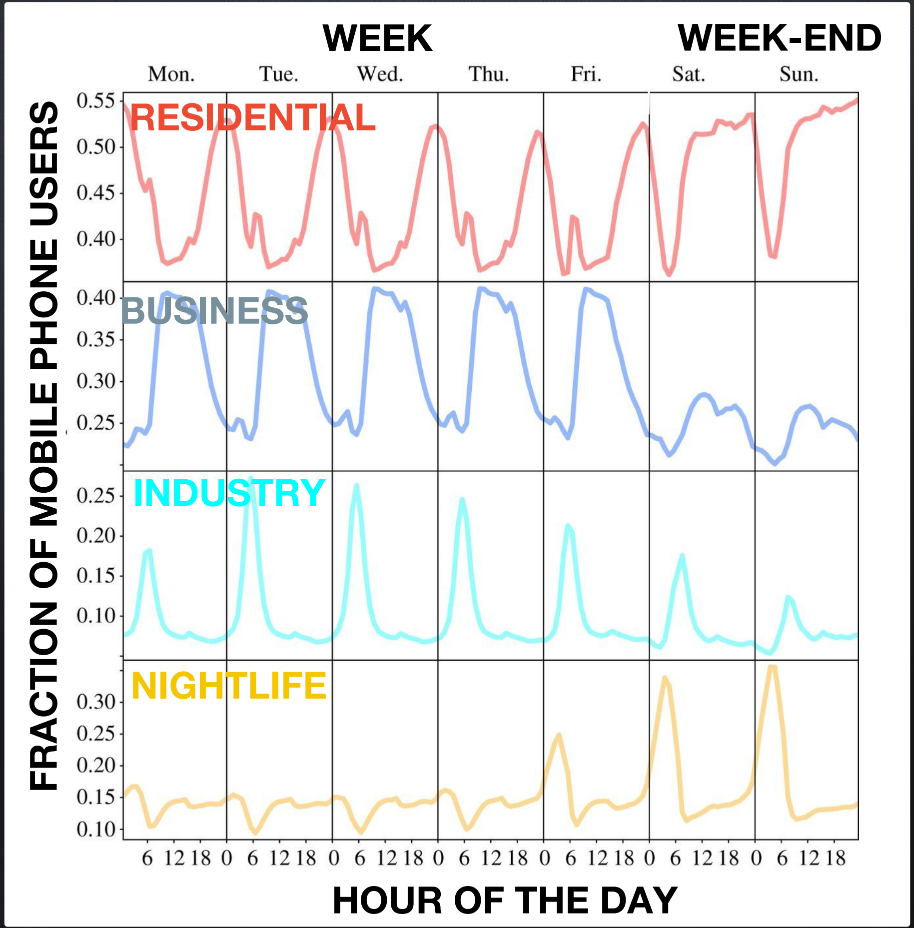


MOBILE PHONE USAGE is not independent of TIME

Week-end　Public holidays　School holidays



MOBILE PHONE USAGE is not independent of PLACE

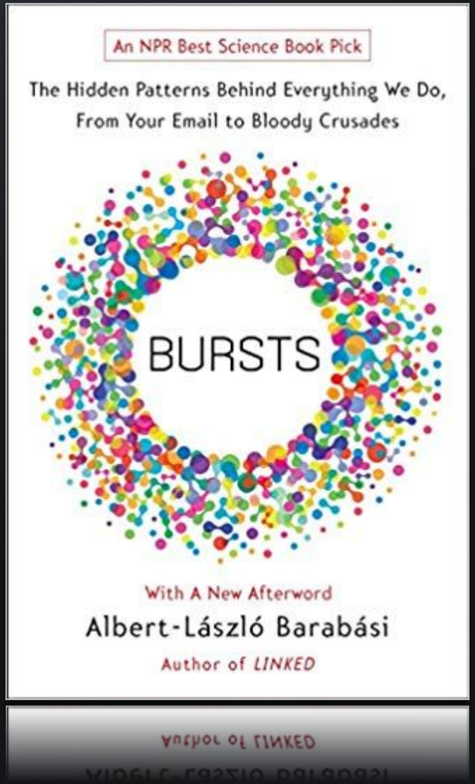


From: Lenormand, M., Picornell, M., Cantú-Ros, O. G., Louail, T., Herranz, R., Barthelemy, M., ... & Ramasco, J. J. (2015). Comparing and modelling land use organization in cities. Royal Society open science, 2(12), 150449.

Mobile Phone activity is BURSTY

Temporal resolution depends on user activity

(only first antenna is recorded during long call)



But HUMAN MOBILITY is highly predictable

Limits of Predictability in Human Mobility

Science AAAS

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

A range of applications, from predicting the spread of human and electronic viruses to city planning and resource management in mobile communications, depend on our ability to foresee the whereabouts and mobility of individuals, raising a fundamental question: To what degree is human behavior predictable? Here we explore the limits of predictability in human dynamics by studying the mobility patterns of anonymized mobile phone users. By measuring the entropy of each individual's trajectory, we find a 93% potential predictability in user mobility across the whole user base. Despite the significant differences in the travel patterns, we find a remarkable lack of variability in predictability, which is largely independent of the distance users cover on a regular basis.

But HUMAN MOBILITY is highly predictable

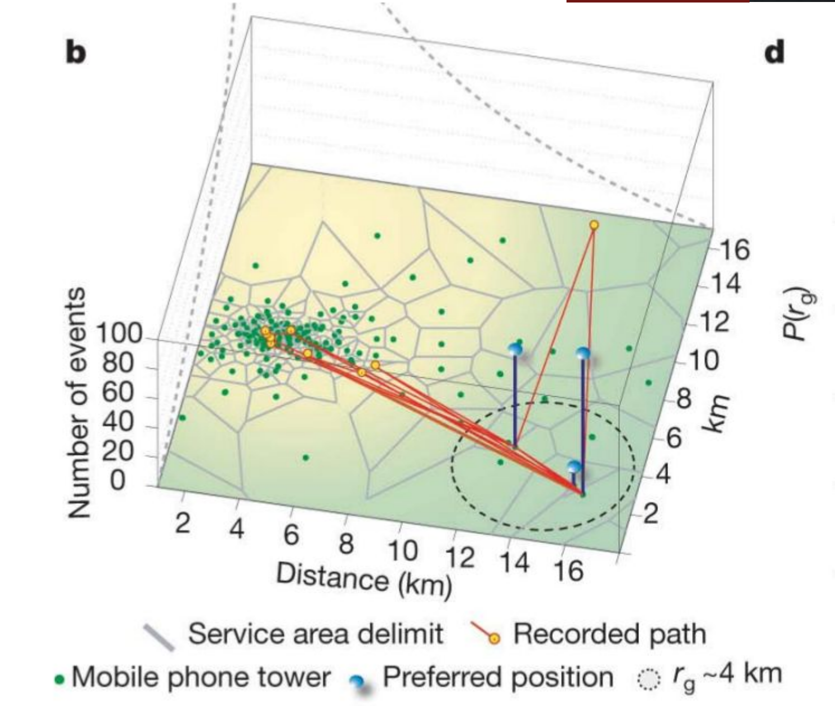
Understanding individual human mobility patterns

nature

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

2008

Despite their importance for urban planning[1], traffic forecasting[2] and the spread of biological[3–5] and mobile viruses[6], our understanding of the basic laws governing human motion remains limited owing to the lack of tools to monitor the time-resolved location of individuals. Here we study the trajectory of 100,000 anonymized mobile phone users whose position is tracked for a six-month period. We find that, in contrast with the random trajectories predicted by the prevailing Lévy flight and random walk models[7], human trajectories show a high degree of temporal and spatial regularity, each individual being characterized by a time-independent characteristic travel distance and a significant probability to return to a few highly frequented locations. After correcting for differences in travel distances and the inherent anisotropy of each trajectory, the individual travel patterns collapse into a single spatial probability distribution, indicating that, despite the diversity of their travel history, humans follow simple reproducible patterns. This inherent similarity in travel patterns could impact all phenomena driven by human mobility, from epidemic prevention to emergency response, urban planning and agent-based modelling.



ASSUMPTIONS HAVE TO BE MADE

1 SIM card equal 1 person.

The penetration rate is independent of time and space.

The market share between providers is independent of time and space.

Each phone/SIM location is accurately known.

Each phone/SIM location is known at any time.

Or Mobile phone usage is independent of time and space.

The data ‘belongs’ to **PRIVATE COMPANIES**.

Access is hard because of **BUSINESS-SENSITIVITY** and **PRIVACY** issues.

DATA HETEROGENEITY can be important among MNOs.

Access is generally limited to **AGGREGATED DATA** and subject to **NDA**.

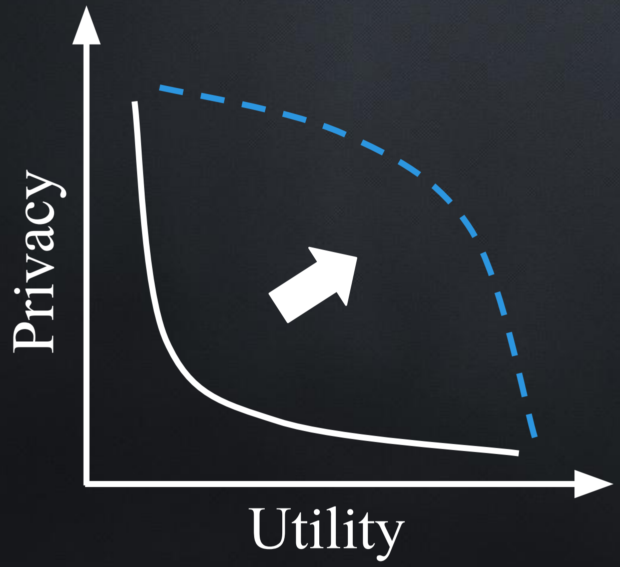
ANONYMIZATION is not sufficient to preserve privacy

Unique in the Crowd: The privacy bounds of human mobility **2013**

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

We study fifteen months of human mobility data for one and a half million individuals and find that human mobility traces are highly unique. In fact, in a dataset where the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals. We coarsen the data spatially and temporally to find a formula for the uniqueness of human mobility traces given their resolution and the available outside information. This formula shows that the uniqueness of mobility traces decays approximately as the 1/10 power of their resolution. Hence, even coarse datasets provide little anonymity. These findings represent fundamental constraints to an individual's privacy and have important implications for the design of frameworks and institutions dedicated to protect the privacy of individuals.

There is a trade-off between PRIVACY and UTILITY of personal data



OPAL

open algorithms for better decisions

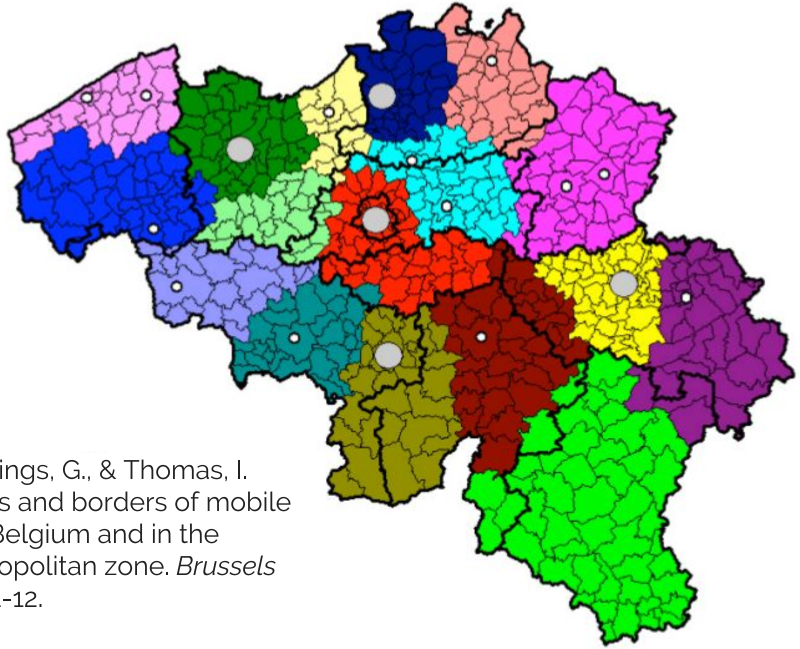
https://www.opalproject.org/general-overview/


D4D
challenge

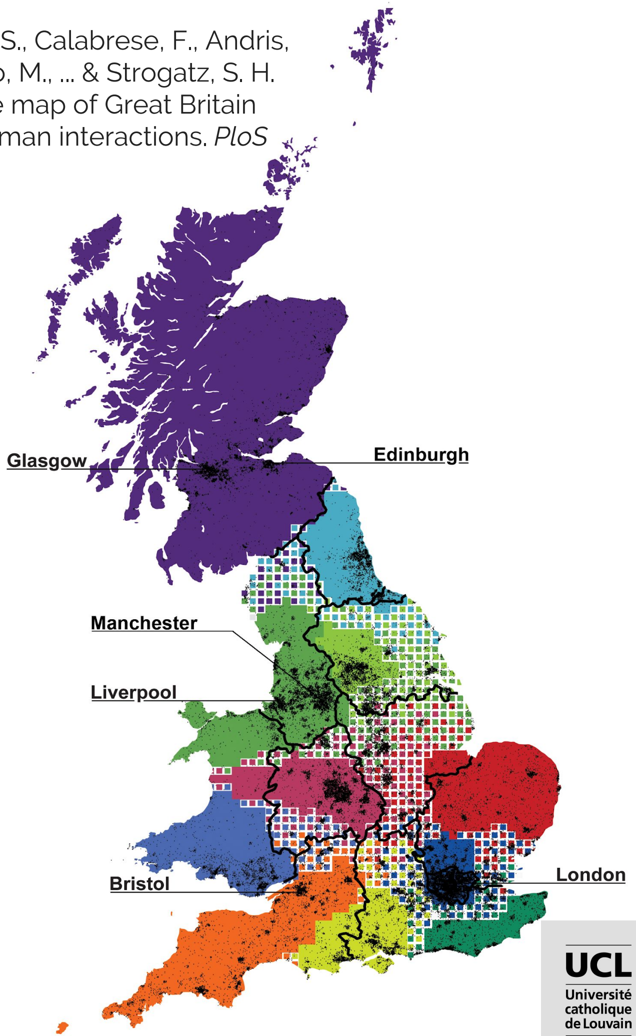
Examples of Applications

SOCIAL NETWORK
Community mapping

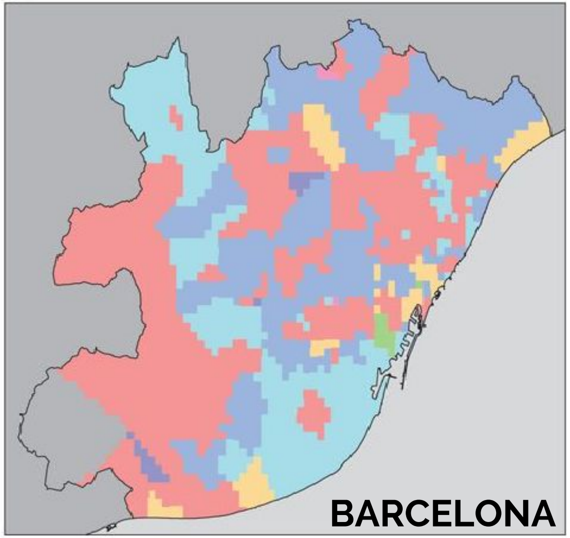
Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., ... & Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PloS one*, *5*(12), e14248.



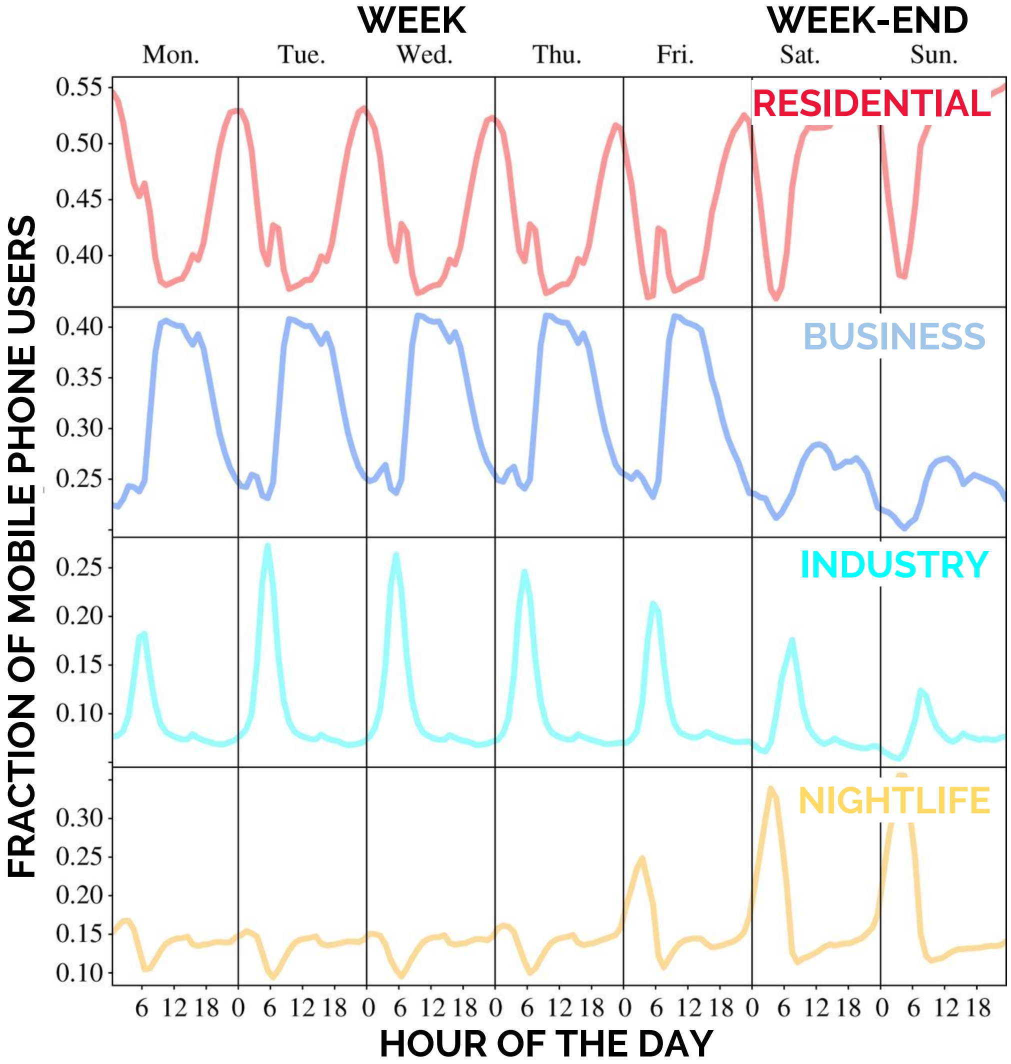
Blondel, V., Krings, G., & Thomas, I. (2010). Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone. *Brussels Studies*,*42*(4), 1-12.



TIME SERIES
Land use classification



Lenormand, M., Picornell, M., Cantú-Ros, O. G., Louail, T., Herranz, R., Barthelemy, M., ... & Ramasco, J. J. (2015). Comparing and modelling land use organization in cities. Royal Society open science, 2(12), 150449.

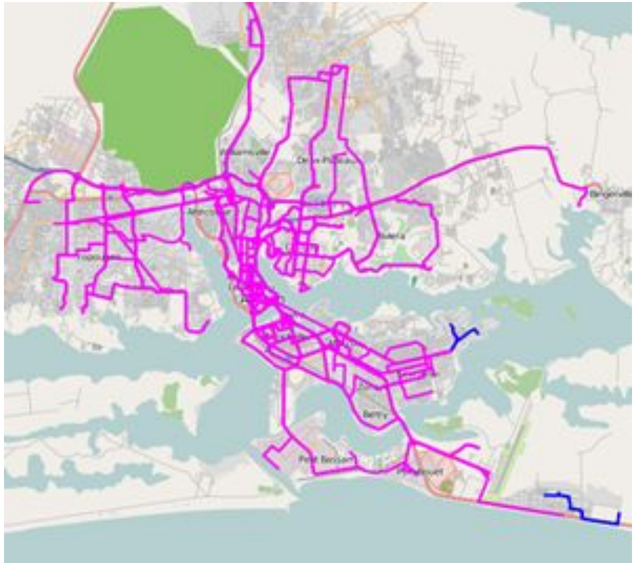


Mietzer, T. (2015). Urban Road Construction and Human mobility: Evidence from Dakar, Senegal. *Netmob 2015*.

MOBILITY

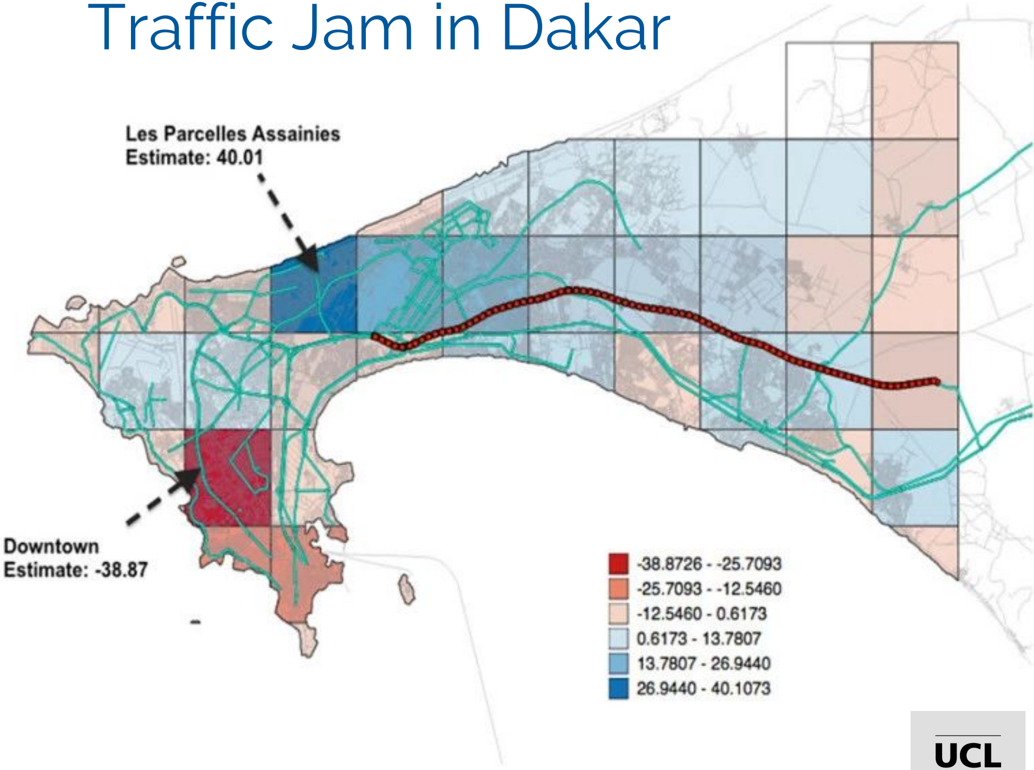
Transportation

Bus routes in Ivory Coast



Berlingerio, M., Calabrese, F., Di Lorenzo, G., Nair, R., Pinelli, F., & Sbodio, M. L. (2013, September). AllAboard: a system for exploring urban mobility and optimizing public transport using cellphone data. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 663-666). Springer Berlin Heidelberg.

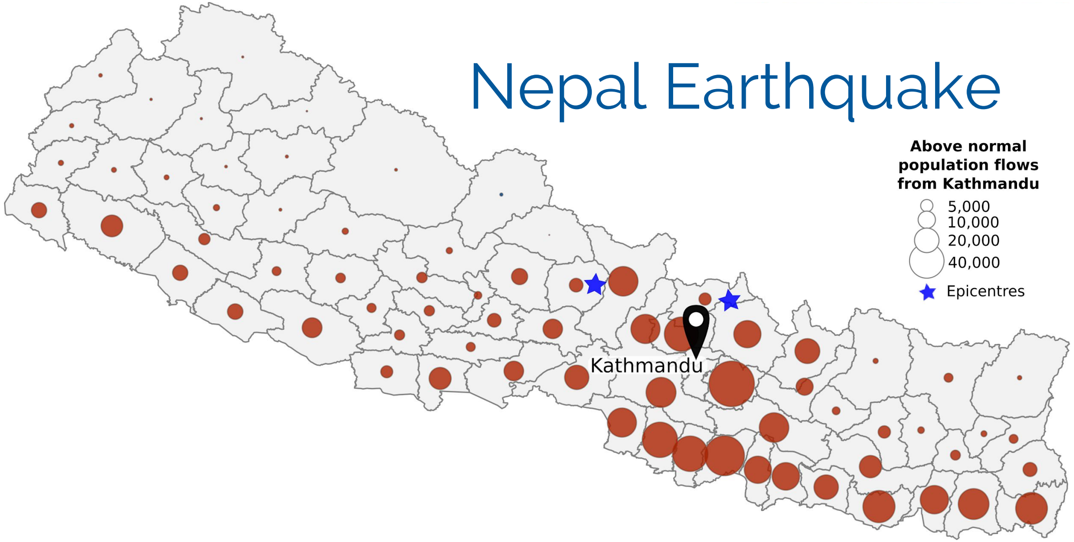
Traffic Jam in Dakar



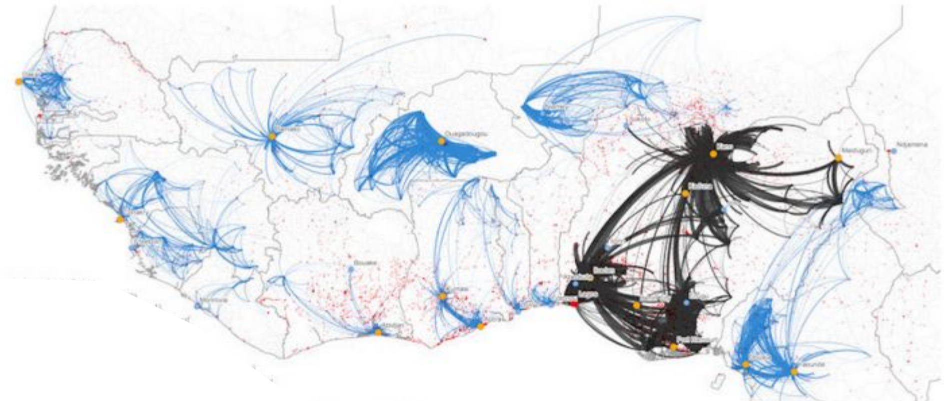
UCL
Université catholique de Louvain

MOBILITY
Epidemiology
Crisis

Nepal Earthquake

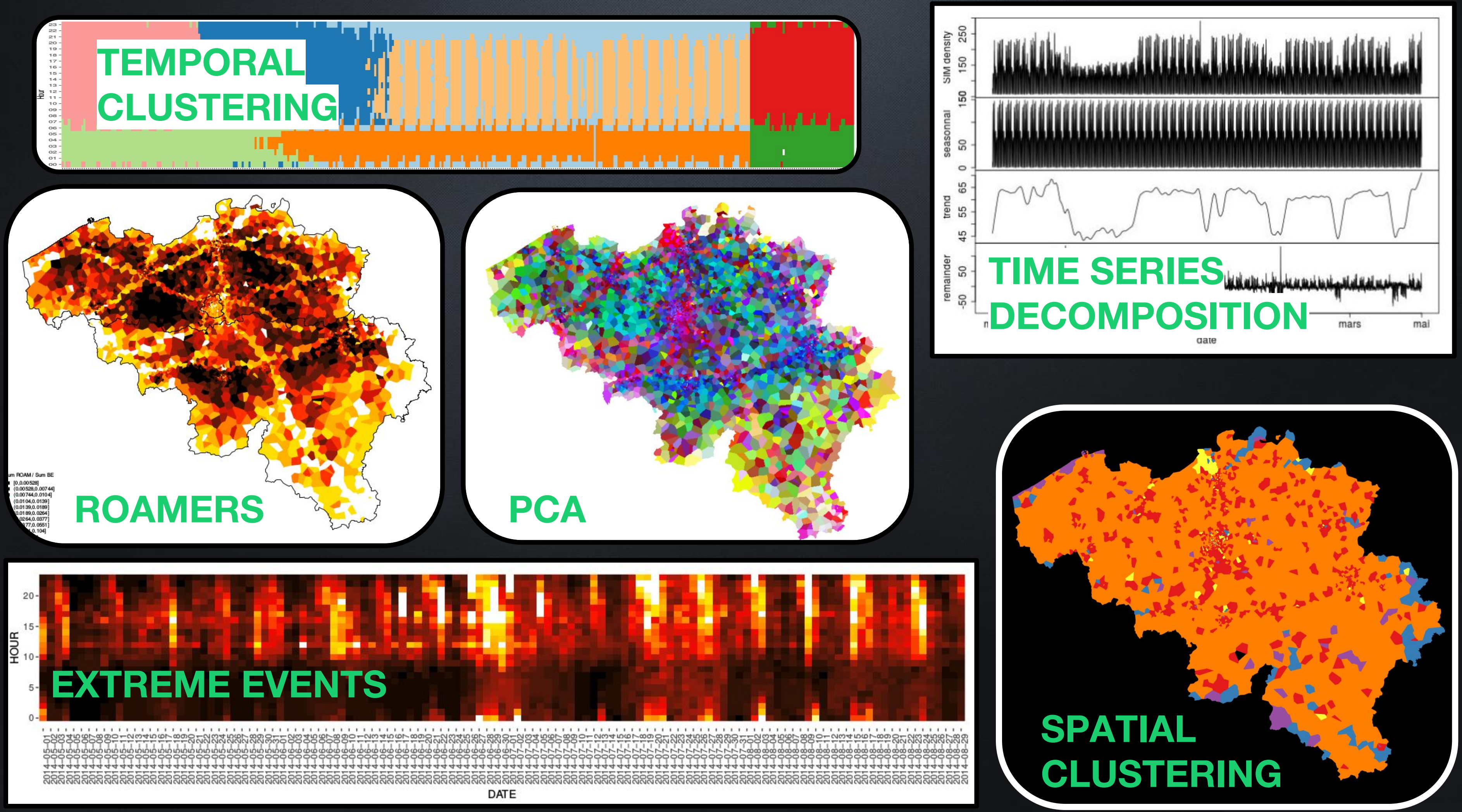


Ebola outbreak



FLOWMINDER.ORG

UCL
Université catholique de Louvain



TEMPORAL CLUSTERING
TIME SERIES DECOMPOSITION
SIM density
seasonnal
trend
remainder
mars
mai
date
ROAMERS
PCA
EXTREME EVENTS
HOUR
DATE
SPATIAL CLUSTERING