National Research Council of Italy

Institute of Electronics Computer and Telecommunication Engineering

Mining cellular traffic collected by mobile operator network probes

M. Fiore, CNR-IEIIT, Italy

Contributors:
D. Naboulsi, A. Furno, R. Stanica, M. Gramaglia

Outline

1. **Introducing mobile traffic data**
    -   Context
    -   Mobile traffic data collection

2. **Overview of the research field**
    -   Literature classification
   -   Examples of results

3. **Reading list**

1
Introducing mobile traffic data
Context and data collection

Context (I)

- **Mobile communication usage is growing (very) fast**
  - Compound annual growth rate of **146%** in 2006-2012
  - Global mobile data traffic grew **81%** in 2013…
  - …and it will increase nearly **11-fold** by 2018
  - Mobile data traffic is growing **faster** than Internet traffic did back at the turn of the millennium!
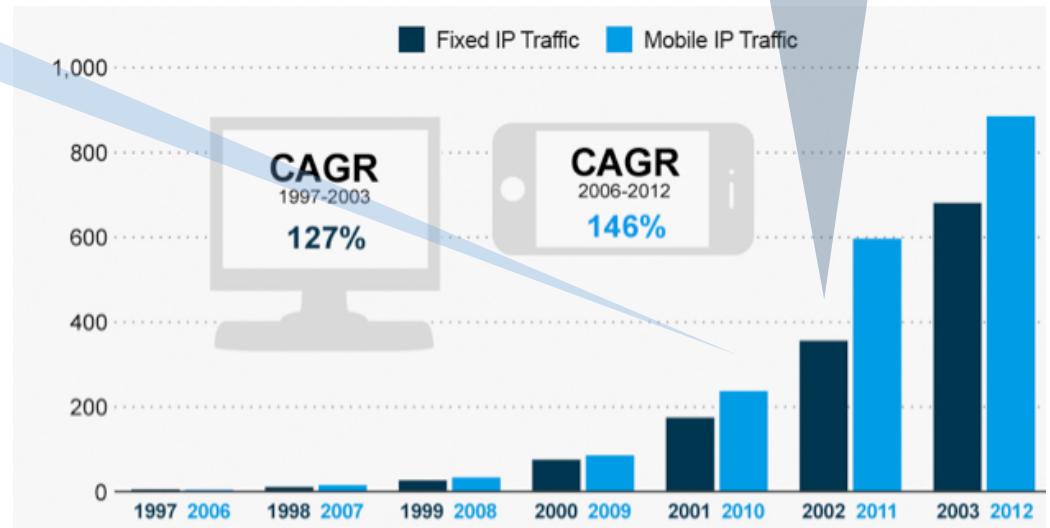
*[source: Cisco Visual Networking Index]*

Context (II)

- **The surge of mobile activity...**
  - **Mobile devices** popularity rise them to status symbols
  - Mobile **apps and services** are driving societal changes

- **...and some collateral effects of our interest**
  - **Digital activities are easily recorded, tracked, transferred**
    - Apple consolidated.db?
  - An unprecedented opportunity to **monitor human behavior** at very large spatiotemporal scales
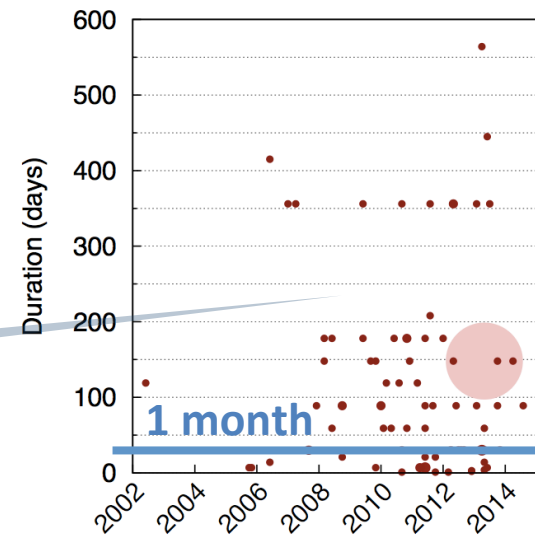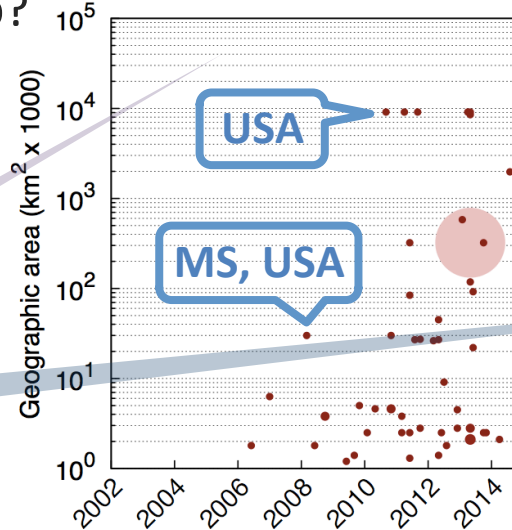
Context (III)

- **Relevance of mobile traffic data**

  - **Mobile service providers**
    - **tailoring of services** (and ads) to users
    - **reselling of** (aggregate?) data to third parties

  - **Mobile operators**
    - face a huge challenge in accommodating the mobile demand
    - understand subscriber usage for more effective **network design and resource management**
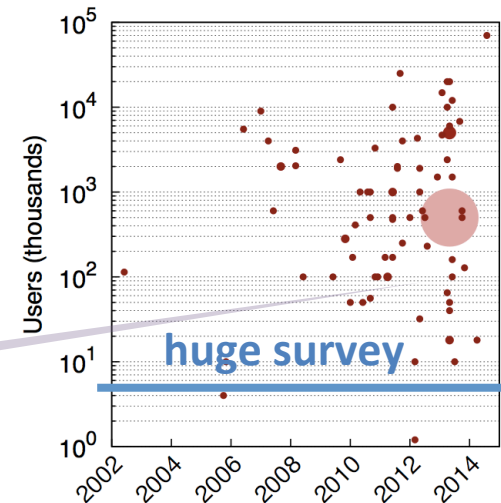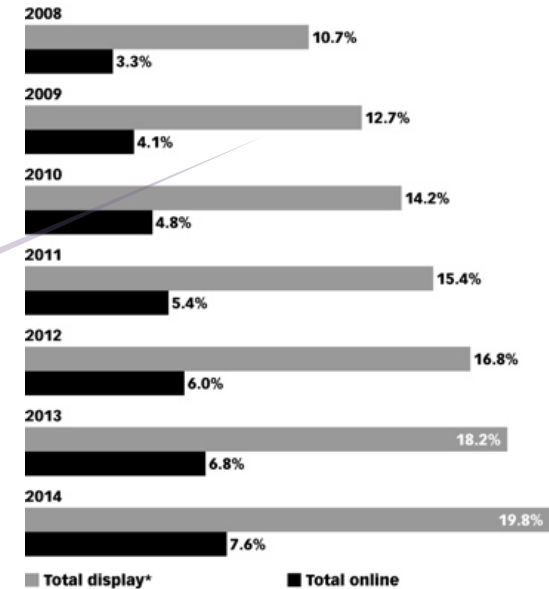
  - **Researchers**   C-RAN, SDN, NFV
    - **networking** (my background): see above
    - plenty of **other disciplines**: studies can be scaled to very large populations (impossible to achieve with traditional approaches)
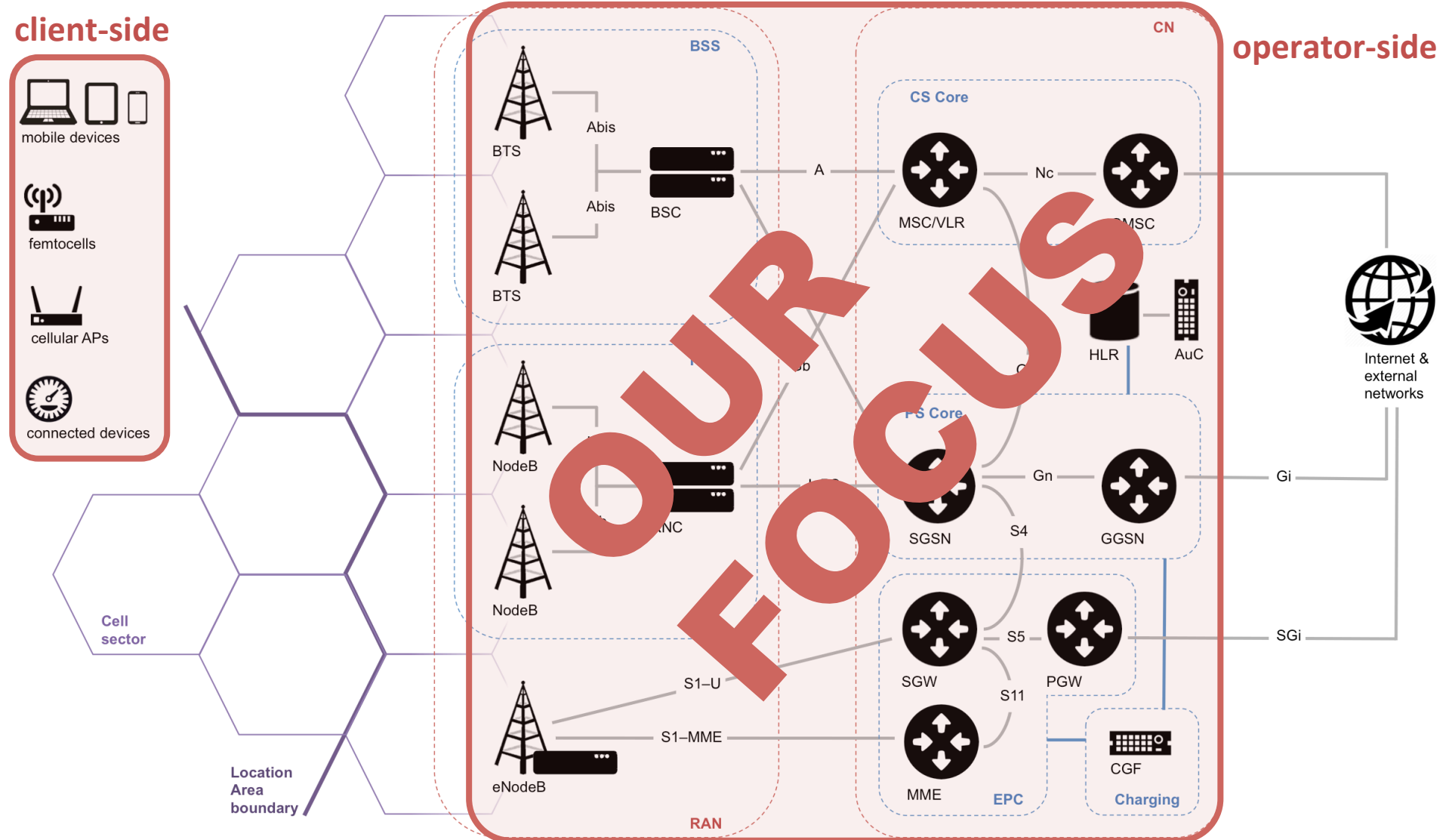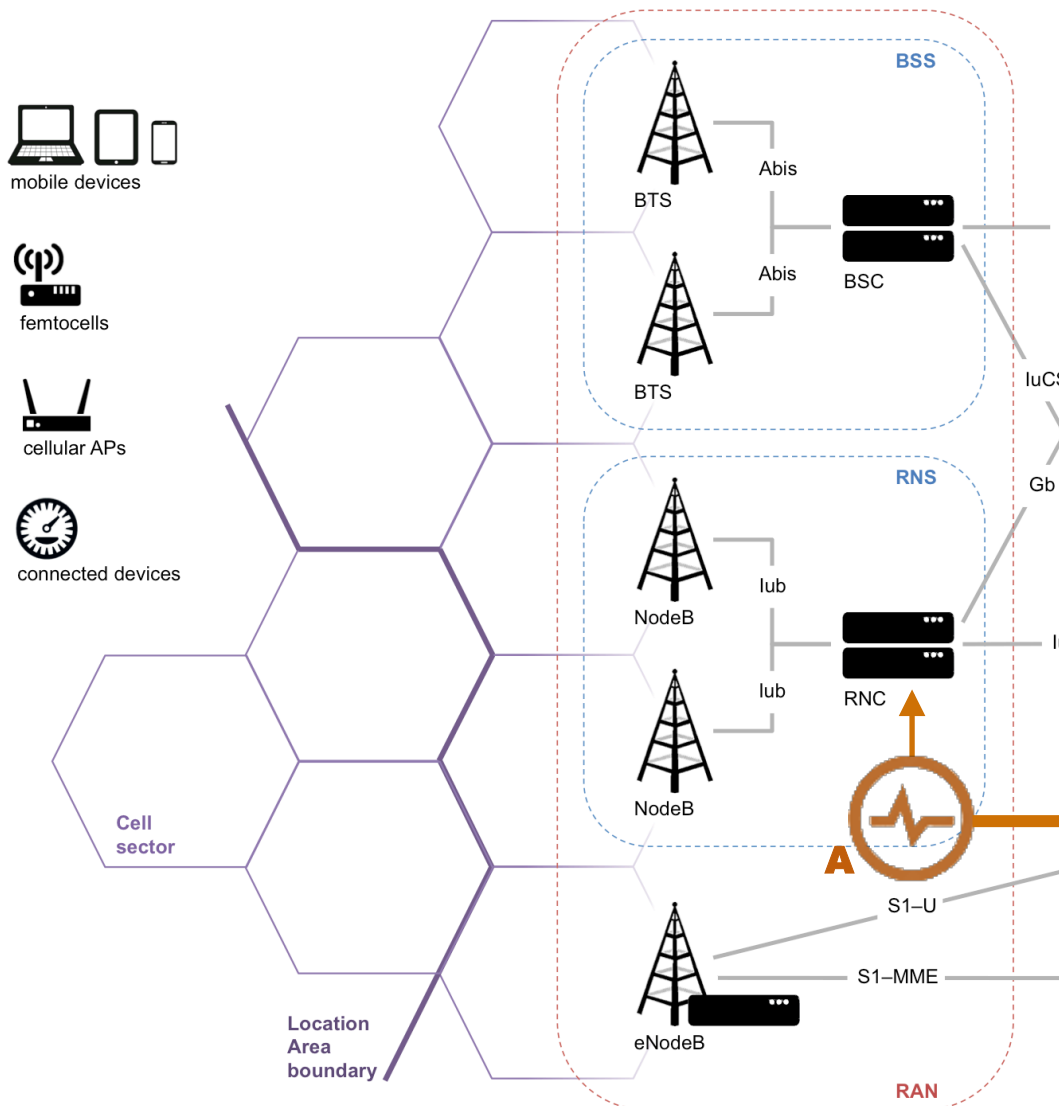

huge survey

Mobile traffic data collection (I)

Mobile traffic data collection (II)

RNC probes

Monitor signaling related to Radio Resource Control (RRC) operation. Fine-grained state changes at each mobile device, e.g.

- attach and detach operations
- start and conclusion of sessions
- all HO and LU events

Plus, record key performance indicators on data transmission, such as the uplink and downlink throughput experienced by devices.

☺ maximum positioning precision

☹ not supported by all equipment large number of probes
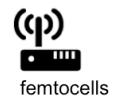
Mobile traffic data collection (III)

MSC probes
Similar to RNC probes, but only in the circuit-switched (CS) domain. Same type of information as RNC probes, but
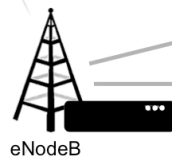- limited to voice/texting
- missing information on events managed locally by RNC, e.g., intra-RNC HO

☺ fair positioning precision
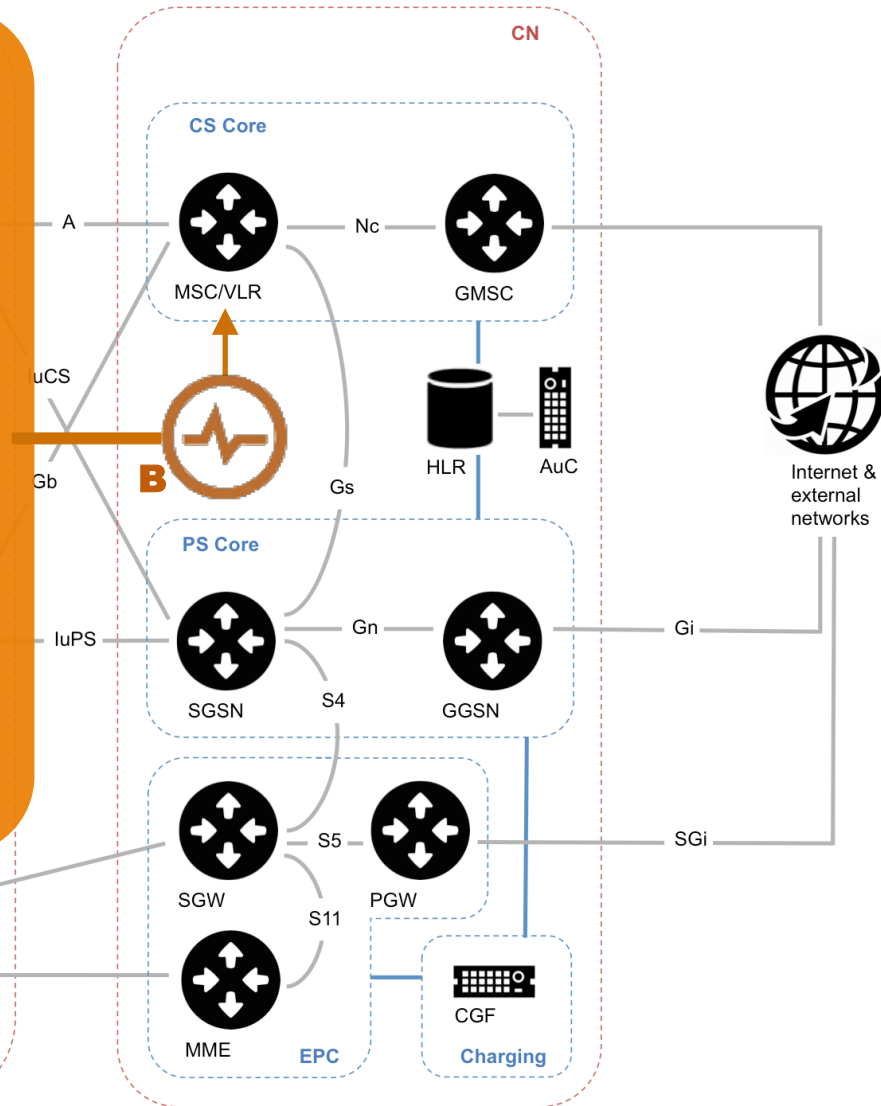 fair number of probes

☹ only voice/texting information

mobile devices

femtocells

cellular APs

connected devices

Cell sector

Location Area boundary

CN

CS Core

A — MSC/VLR — Nc — GMSC

IuCS

B

Gb

Gs

HLR   AuC

PS Core

IuPS — SGSN — Gn — GGSN — Gi

S4

S1–U — SGW — S5 — PGW — SGi

S11

S1–MME — MME — CGF

EPC   Charging

RAN

eNodeB

Internet & external networks
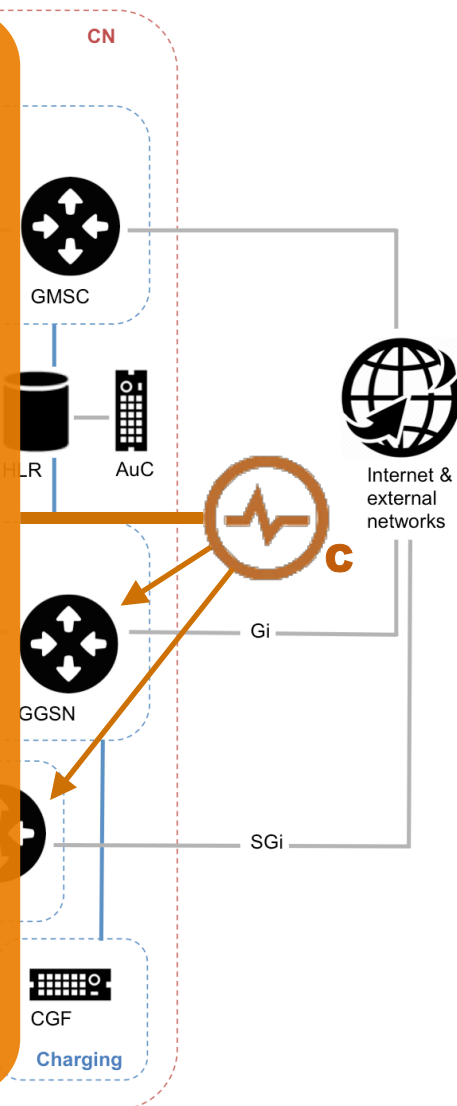
Mobile traffic data collection (IV)

GGSN/PGW probes

Monitor data transiting by packet-switched (PS) domain gateways in order to reach the Internet. For each data traffic session of every device, record Packet Data Protocol (PDP) Context information:

- device identifier
- session start and end time
- traffic volume
- type of service (at different levels of detail, from transport protocol, to service type, up to precise application)

These probes can also store device locations, corresponding to the cell where the device last updated its PDP Context from. They monitor Authentication, Authorization and Accounting (AAA) procedures triggered by PDP Context updates passing by GGSN/PGW.

☺ detailed information on IP traffic

   very large areas (e.g., countries) covered by a few probes

☹ no voice/texting information

   stale view of device location (infrequent PDP Context updates)

CN

GMSC

HLR   AuC

c

Internet & external networks

Gi

GGSN

SGi

CGF

Charging

Mobile traffic data collection (V)
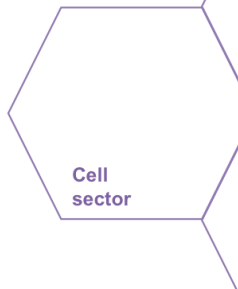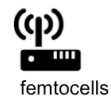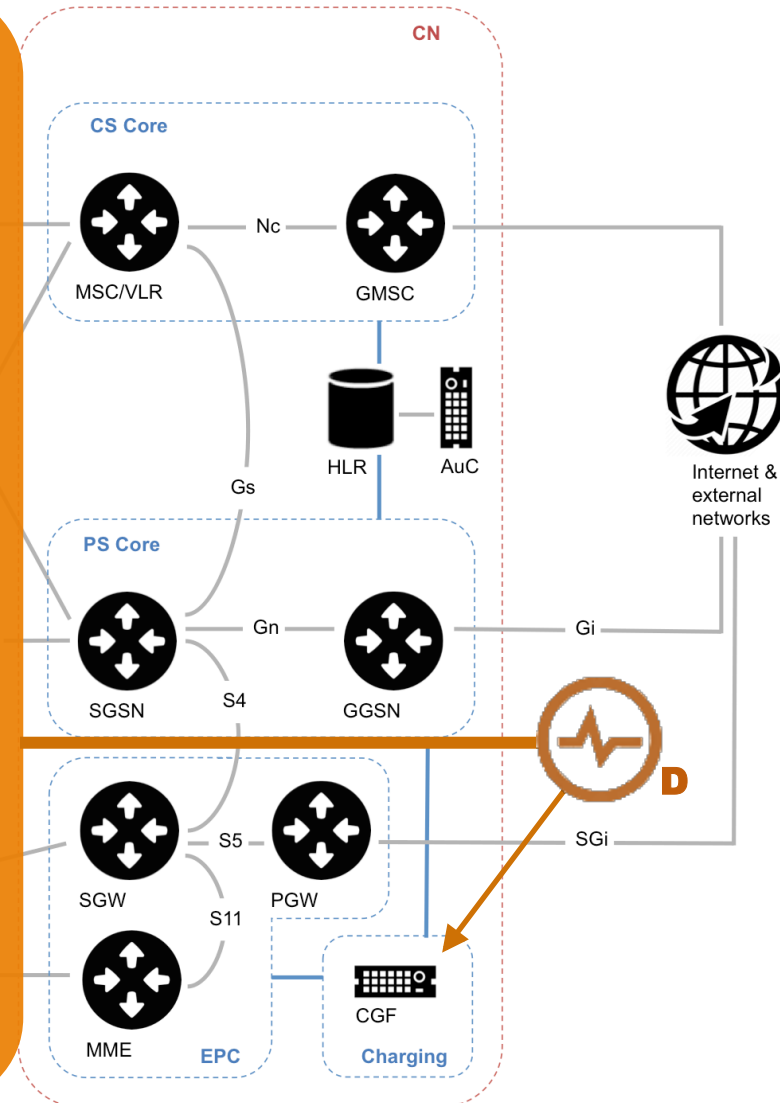

CGF probes

Monitor information stored for billing purposes, i.e., so-called Call Detail Records (CDR). These include, for each device activity:

- device identifier
- start (and possibly end) time
- cell where the activity starts (and possibly ends)
- type of activity: voice, texting, data (possibly including traffic volume)

In some rare cases, these probes can also record HO information.

☺ typically one single probe
    readily available, well formatted
    acceptable positioning precision

☹ no precise traffic information

mobile devices

femtocells

cellular APs

connected devices

Cell sector

CN

CS Core

MSC/VLR — Nc — GMSC

HLR AuC

Gs

PS Core

SGSN — Gn — GGSN — Gi

S4

SGW — S5 — PGW

S11

MME   EPC

CGF

Charging

Internet & external networks

SGi

D

Mobile traffic data collection (VI)

- **Takeaway lessons**
  - **Not all mobile traffic datasets are created equal!**
  - The **quality** of mobile traffic datasets is increasing
    - new probes, more extensive deployments in both RAN and CN
  - The **availability** of mobile traffic datasets is growing
    - operators become more open-minded
    - **challenges** are great occasions to access data!

- **Caveat**
  - Here, we are only considering **"half of the world"**
    - **client-side** mobile traffic data is an equally rich source of knowledge

Nokia Mobile Data Challenge
MIT Reality Mining
Microsoft GeoLife
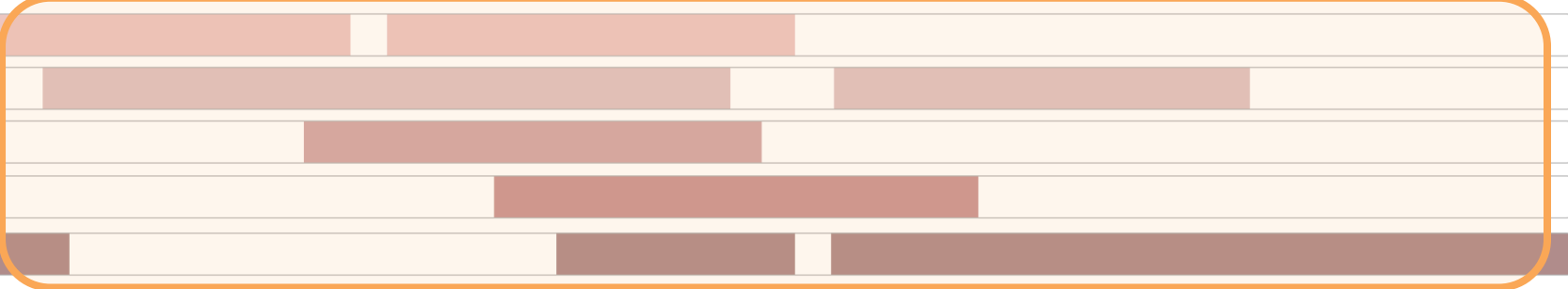
2
Overview of the research field
Using operator-side data

Literature classification

- **A highly multidisciplinary research field**
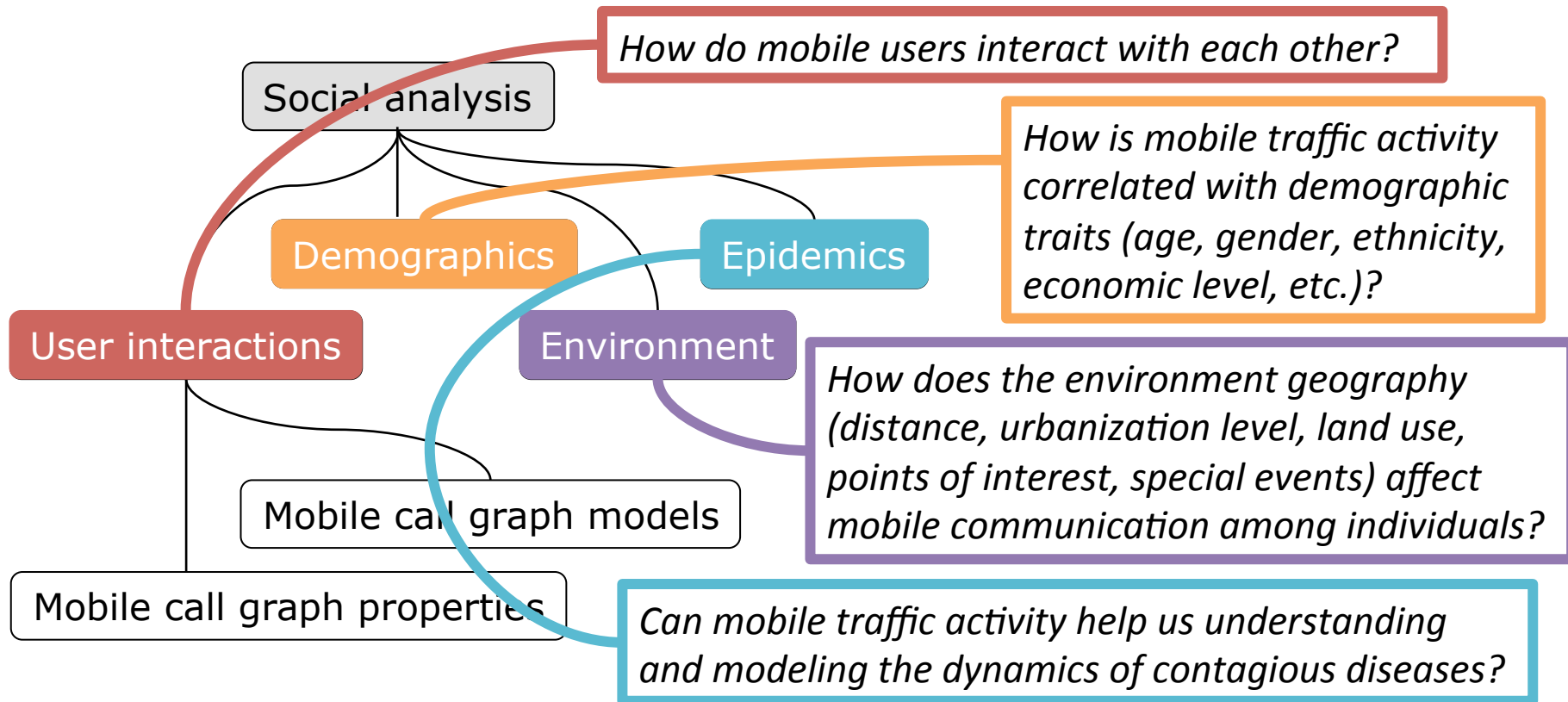 - Mobile traffic data is used in many domains (with overlaps)
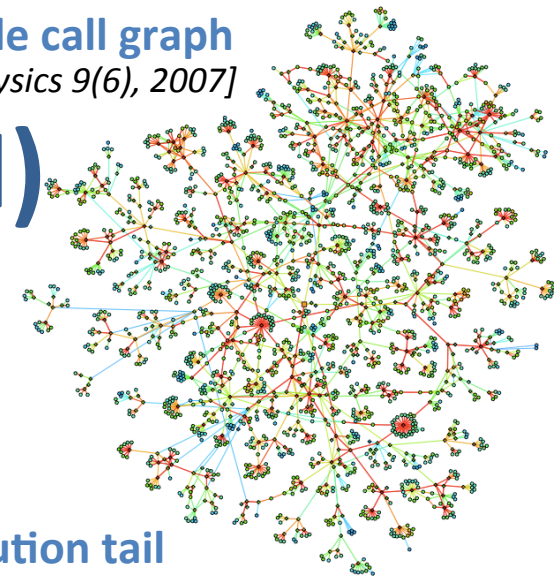
Social analysis (I)

- **Relating mobile traffic and social features**
  - Physics, sociology, epidemiology, networking

How do mobile users interact with each other?

How is mobile traffic activity correlated with demographic traits (age, gender, ethnicity, economic level, etc.)?

How does the environment geography (distance, urbanization level, land use, points of interest, special events) affect mobile communication among individuals?

Can mobile traffic activity help us understanding and modeling the dynamics of contagious diseases?

Social analysis

Demographics

Epidemics

User interactions

Environment

Mobile call graph models

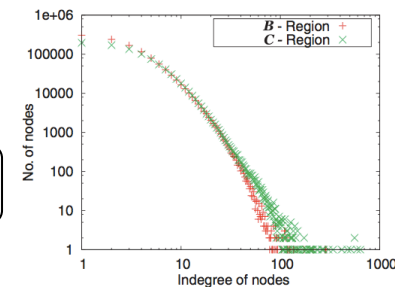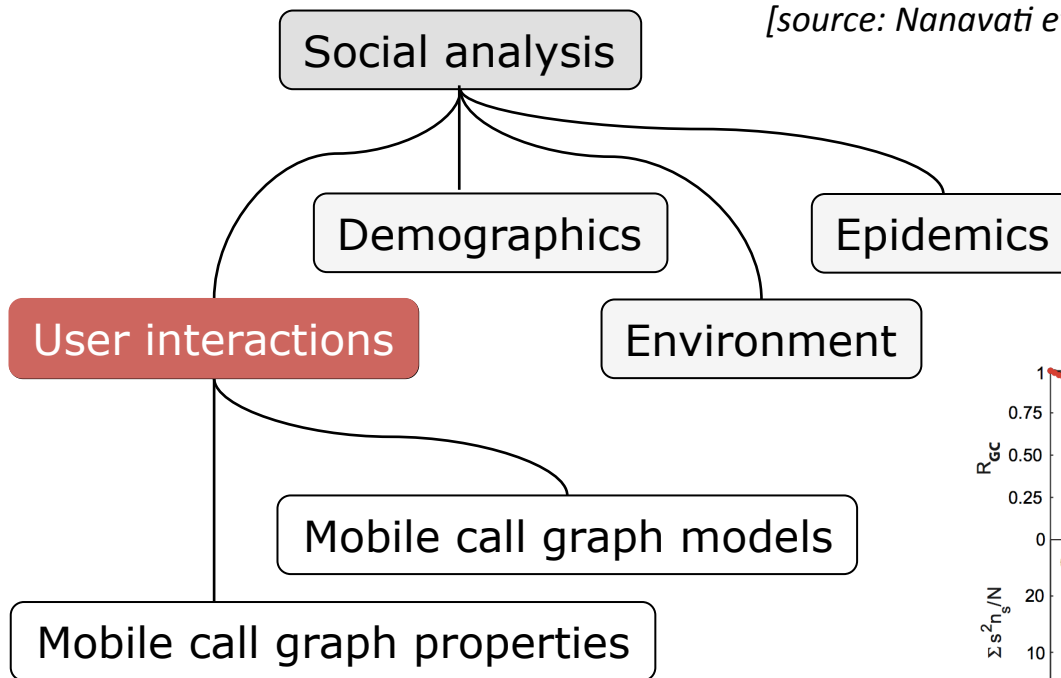Mobile call graph properties

Social analysis (II)

- ## **Examples: user interactions**
  - Mobile call graph analysis

power-law degree distribution tail
[source: Nanavati et al., ACM CIKM, 2006]

Social analysis

Demographics

Epidemics

Environment

User interactions

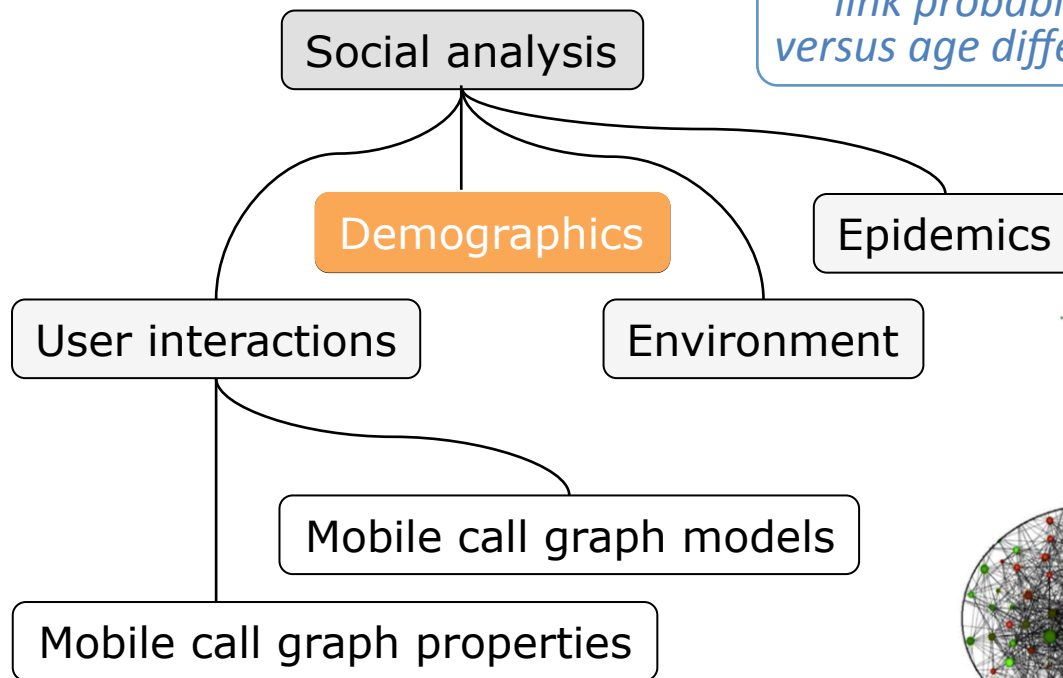Mobile call graph models

Mobile call graph properties
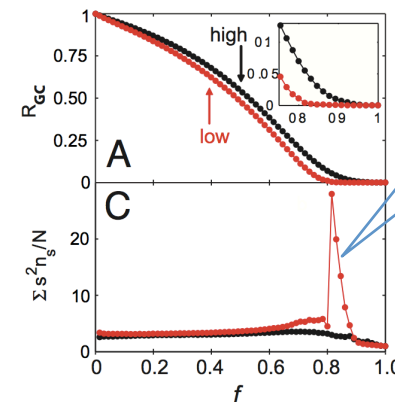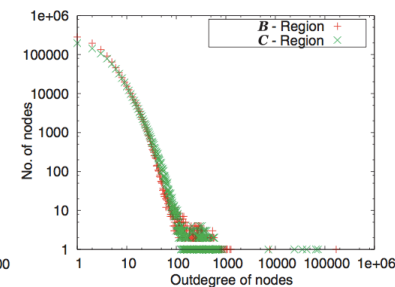
network collapse only if low-weight ties are removed first!

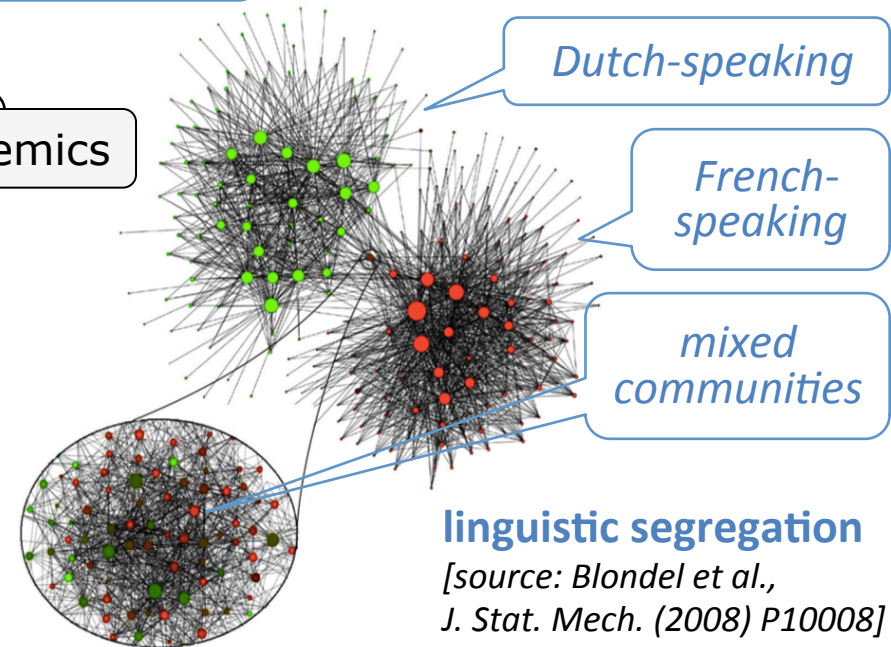**strength of weak ties**
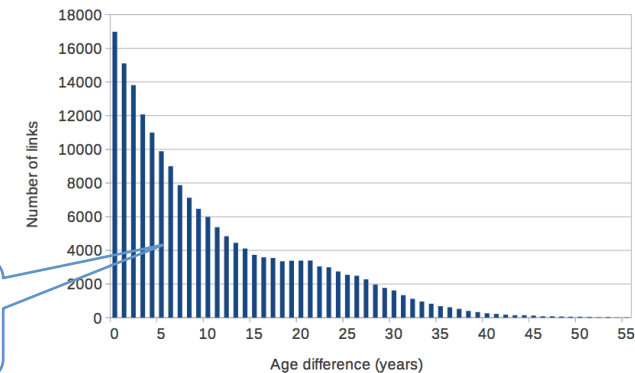[source: Onnela et al., PNAS, 104(18), 2007]

Social analysis (III)

communication matrix among ages

- **Examples: demographics**
  - Homophily and segregation

*link probability versus age difference*

Social analysis

User interactions

Demographics

Epidemics

Environment

Mobile call graph models

Mobile call graph properties

Dutch-speaking

French-speaking

mixed communities

linguistic segregation
[source: Blondel et al., J. Stat. Mech. (2008) P10008]

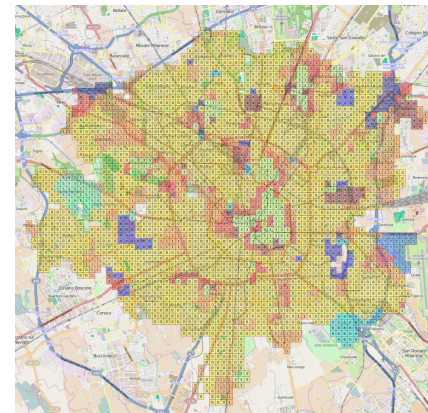# Social analysis (IV)

power-law call probability w.r.t. distance
[source: Onnela et al., PLoS ONE, 6(4):e16939, 2011]

- **Examples: environment**
  - Impact of distance and land use


```
Social analysis
├── Demographics
├── Epidemics
├── Environment
└── User interactions
    ├── Mobile call graph models
    └── Mobile call graph properties
```
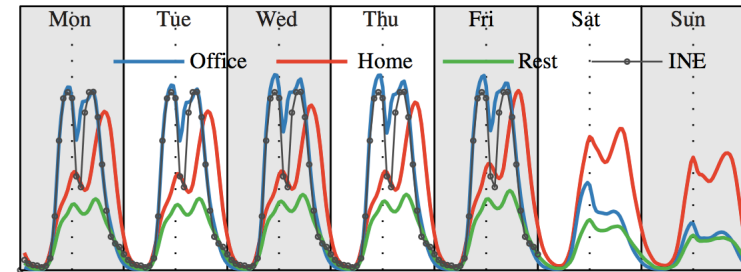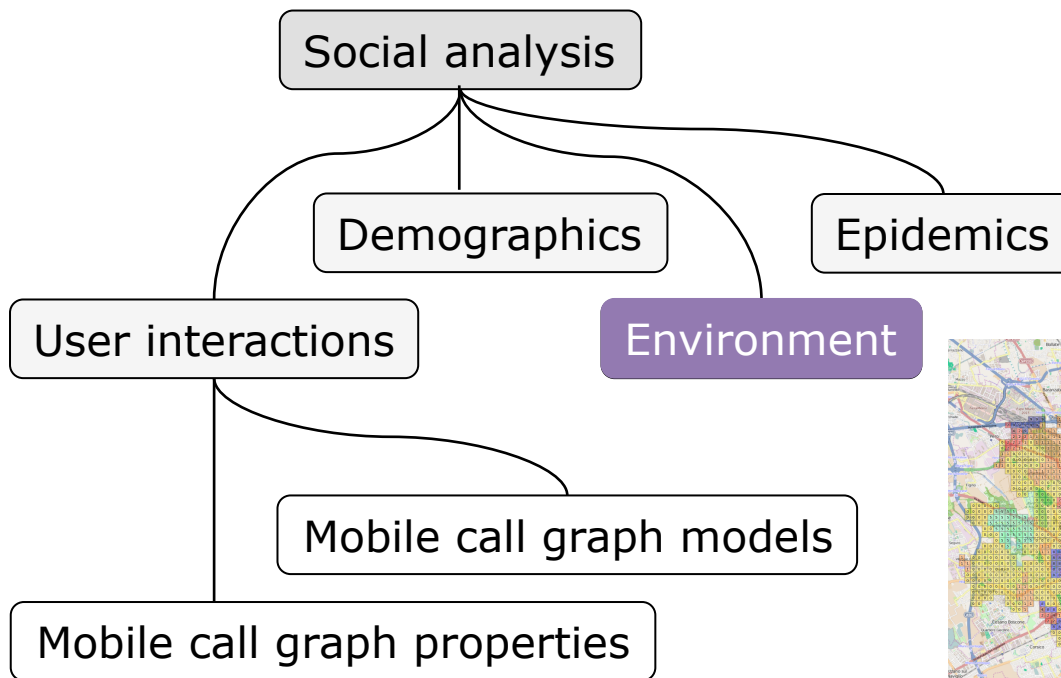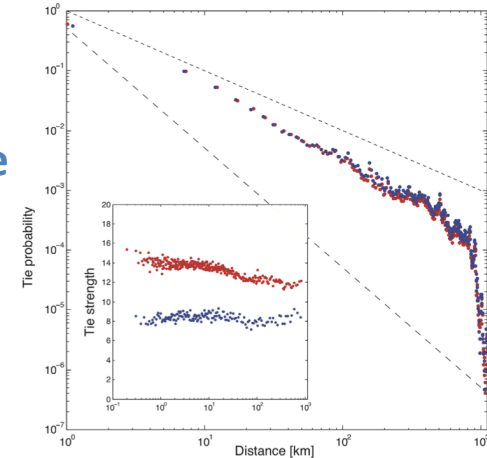

mobile traffic profiles depend on land use
[source: Csaji et al., Physica A, 392(6), 2013]

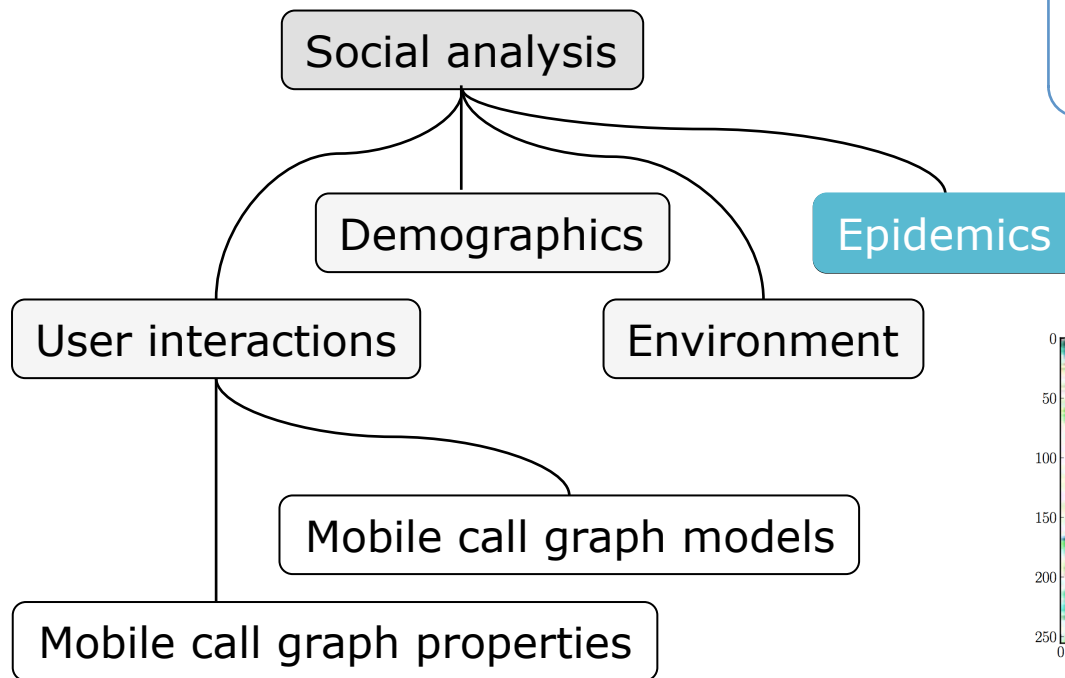**urban fabric detection from mobile traffic**
[source: Furno et al., in preparation]

Social analysis (V)

- **Examples: epidemics**
 - Contagion containment policies

viral information campaign is more effective than movement quarantine

[source: Lima et al., NetMob D4D, 2013]

Social analysis

Demographics

Epidemics

User interactions

Environment

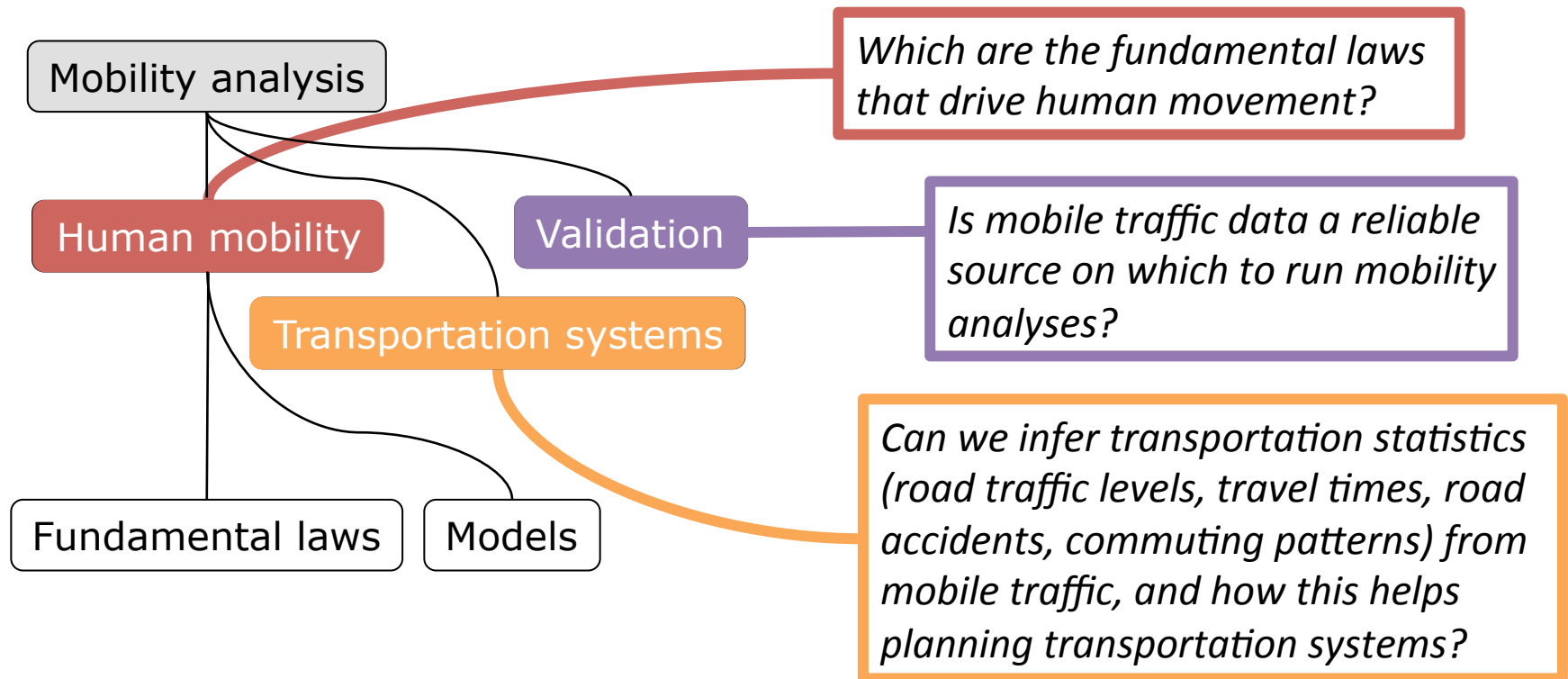Mobile call graph models

Mobile call graph properties

movement matrix among Ivory Coast subprefectures

call matrix among Ivory Coast subprefectures
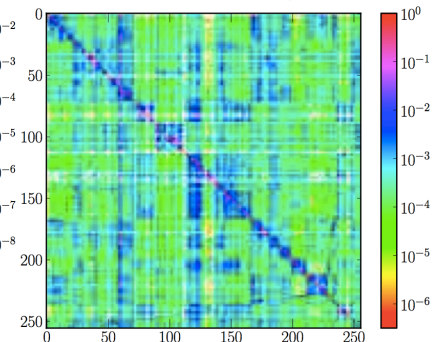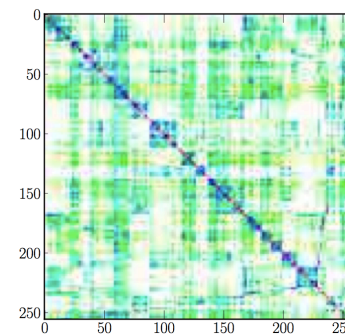
Mobility analysis (I)

- **Extracting mobility information from mobile traffic**
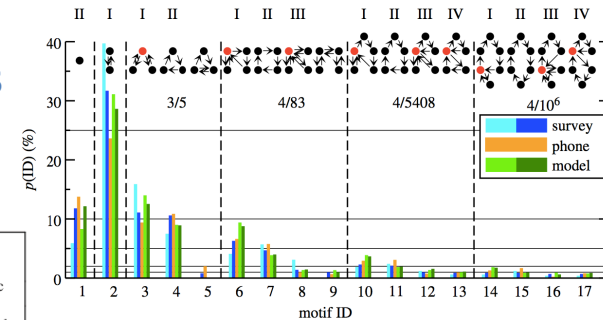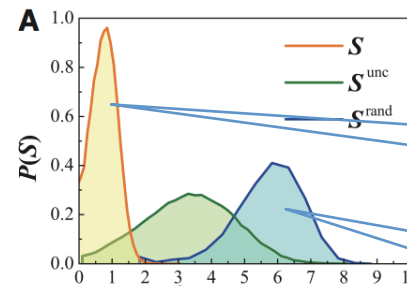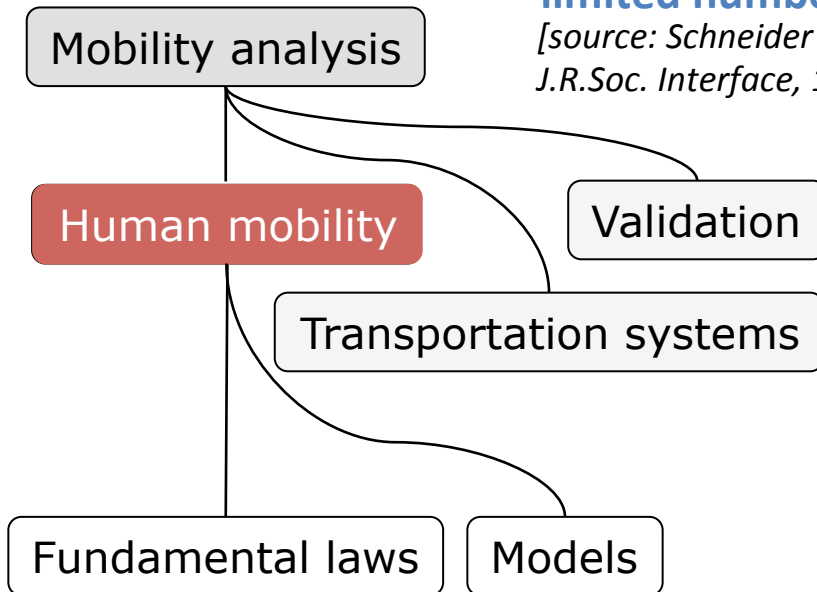 - Physics, sociology, transportation, networking, epidemiology

Mobility analysis

Human mobility

Validation

Transportation systems

Fundamental laws

Models

Which are the fundamental laws that drive human movement?

Is mobile traffic data a reliable source on which to run mobility analyses?

Can we infer transportation statistics (road traffic levels, travel times, road accidents, commuting patterns) from mobile traffic, and how this helps planning transportation systems?

Mobility analysis (II)

- ## **Examples: human mobility**
  - Constrained, routinary behaviors

Mobility analysis

Human mobility

Validation

Transportation systems

Fundamental laws   **Models**
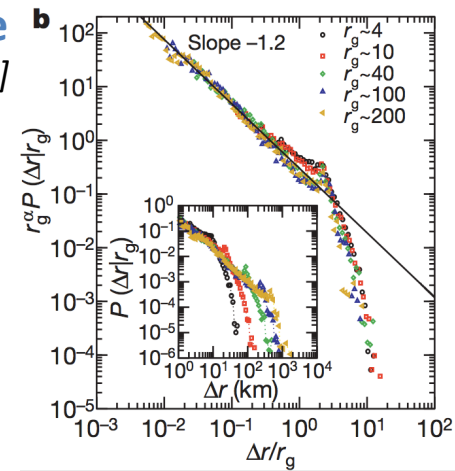
limited number of mobility motifs
*[source: Schneider et al.,
J.R.Soc. Interface, 10(84), 2013]*

entropy of human mobility

entropy of random process

entropy of time-oblivious human mobility

predictability versus radius of gyration

elevate predictability
*[source: Song et al.,
Science, 327(5968), 2010]*

Mobility analysis (III)

- **Examples: transportation systems**
  - Estimating traffic and mobility flows

Mobility analysis

Human mobility

Validation

Transportation systems

Fundamental laws

Models

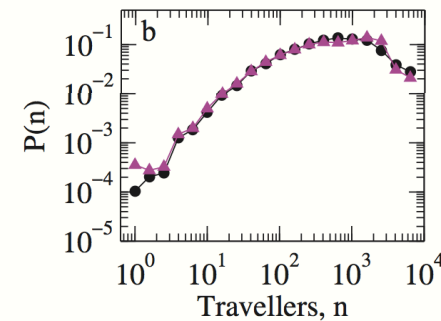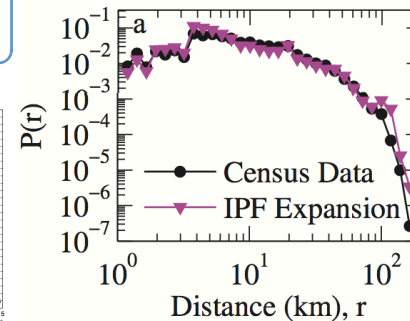*traffic congestion due to accidents (ground truth)*

detection of traffic jams
[source: Janecek et al., ACM UbiComp, 2012]
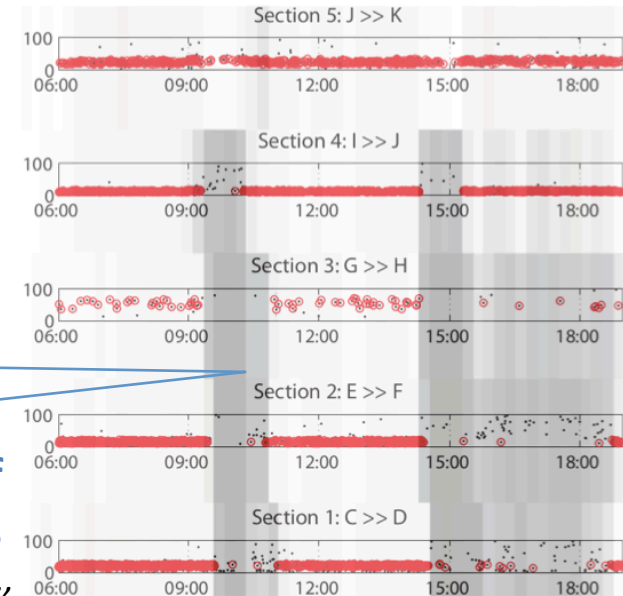
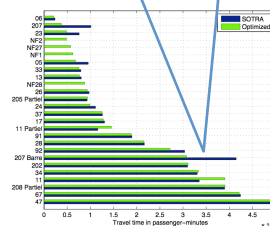**inference of O-D matrices**
[source: Yang et al., NetMob, 2013]

spared wait time

optimization of transportations
[source: Berlingerio et al., NetMob D4D, 2013]

Mobility analysis (IV)

- ## Examples: validation

  - Limited set of (controversial) results
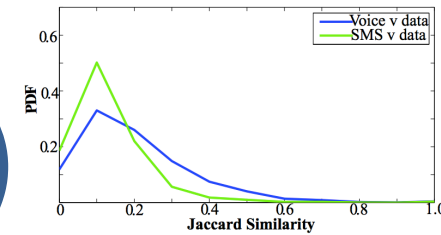
different home/work location detection when using diverse mobile traffic data
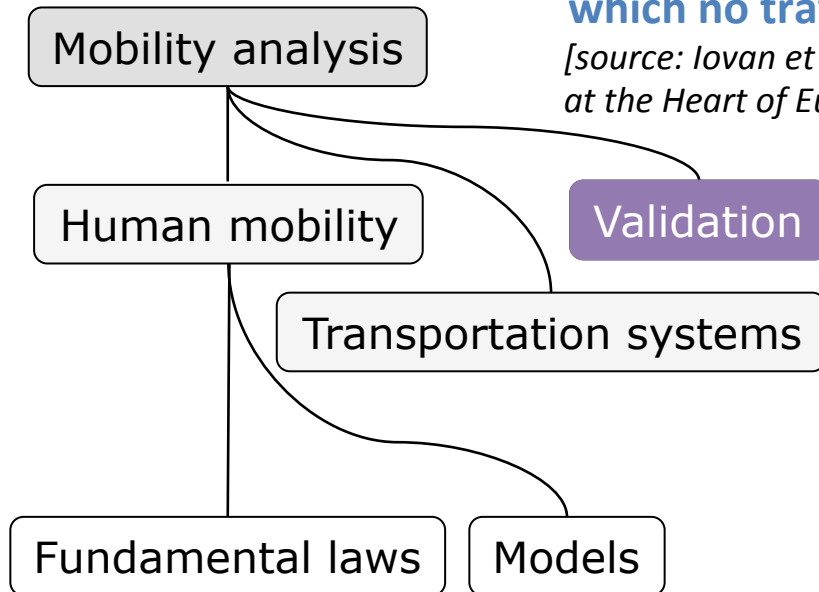*[source: Ranjan et al., ACM MC2R, 2012]*

minimum sampling frequency beyond which no travel distance variation occurs
[source: Iovan et al., Geographic Information Science at the Heart of Europe 2012]

50 samples/day

Mobility analysis

Human mobility

Validation

Transportation systems

Fundamental laws

Models

captures longer trips

covers larger populations
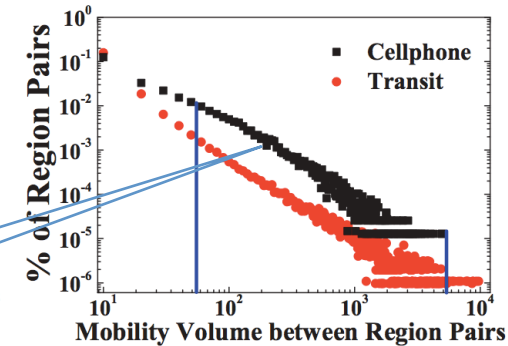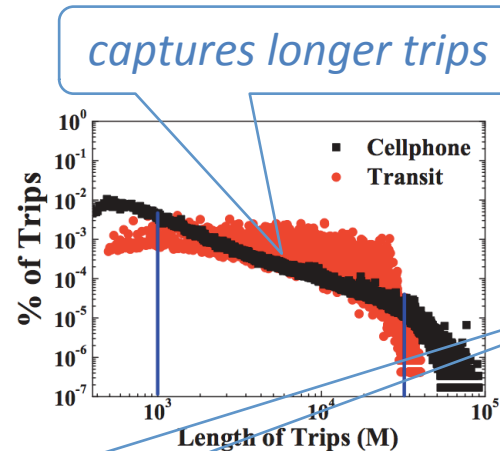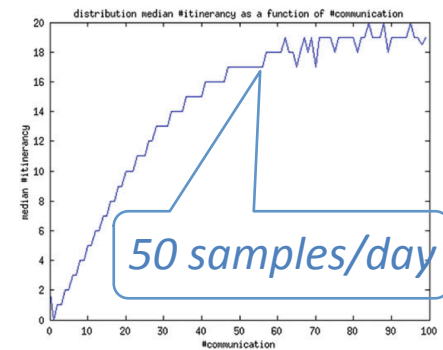
mobile traffic data is more informative than transit data collected by authorities
[source: Zhang et al., ACM MobiCom 2014]
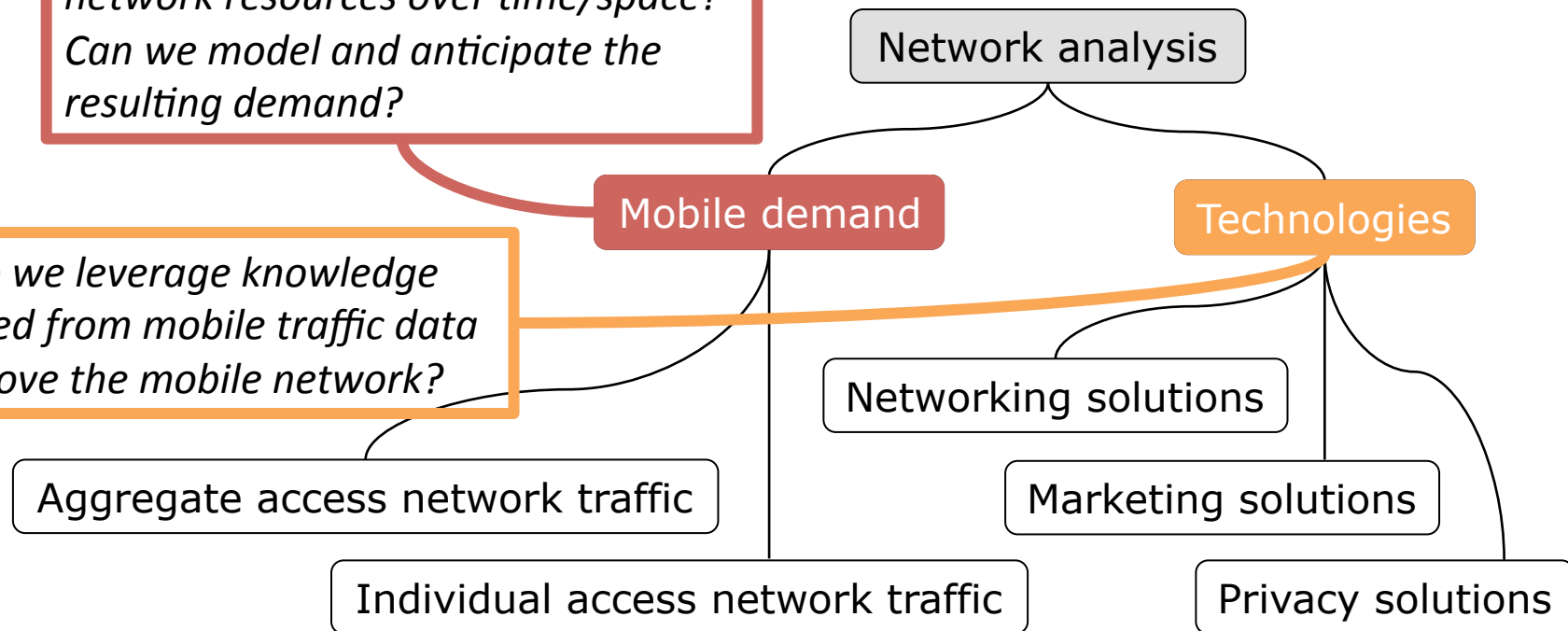
Networking analysis (I)

- **Adapting network design to the mobile demand**
  - Networking, sociology

How do mobile subscribers consume network resources over time/space? Can we model and anticipate the resulting demand?

How do we leverage knowledge extracted from mobile traffic data to improve the mobile network?
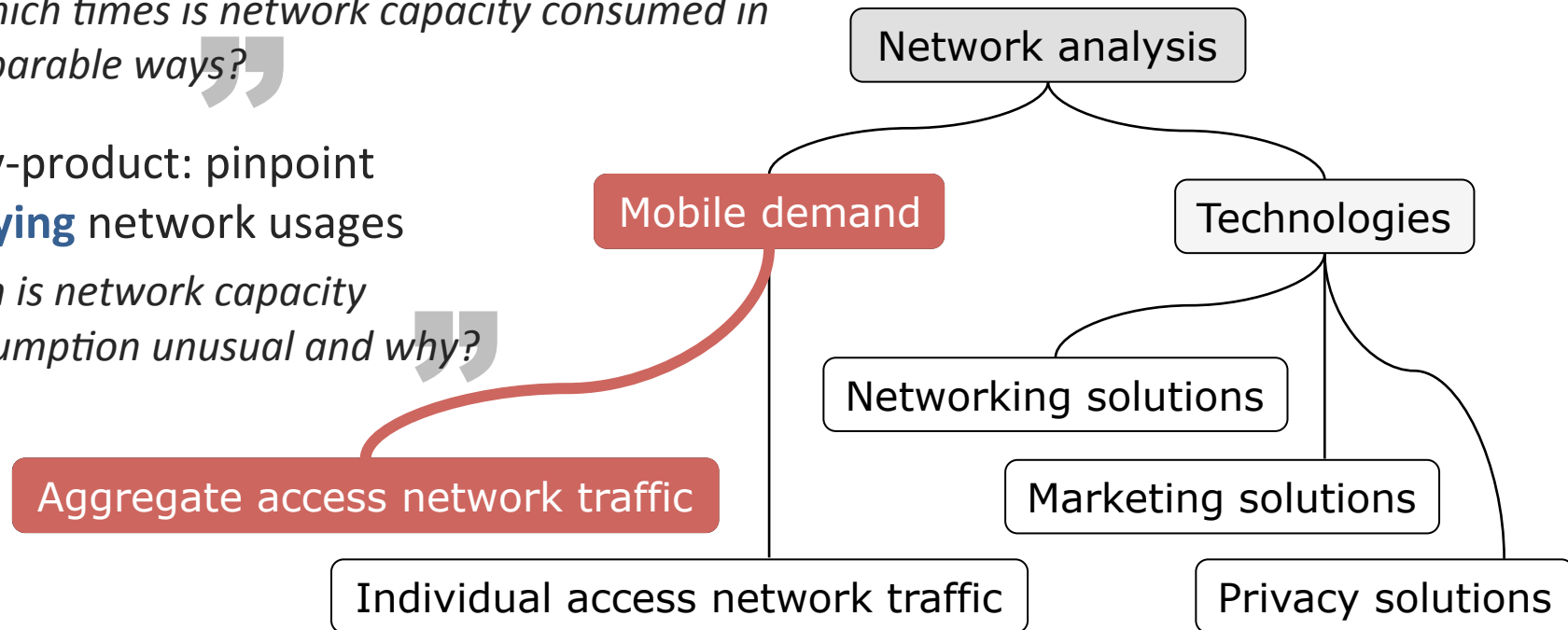
Network analysis

Mobile demand

Technologies

Aggregate access network traffic

Networking solutions

Marketing solutions

Individual access network traffic

Privacy solutions

Networking analysis (II)

- **Example: aggregate access network traffic**
  - Mobile traffic demand profiling

1. identify **profiles** of typical user demand

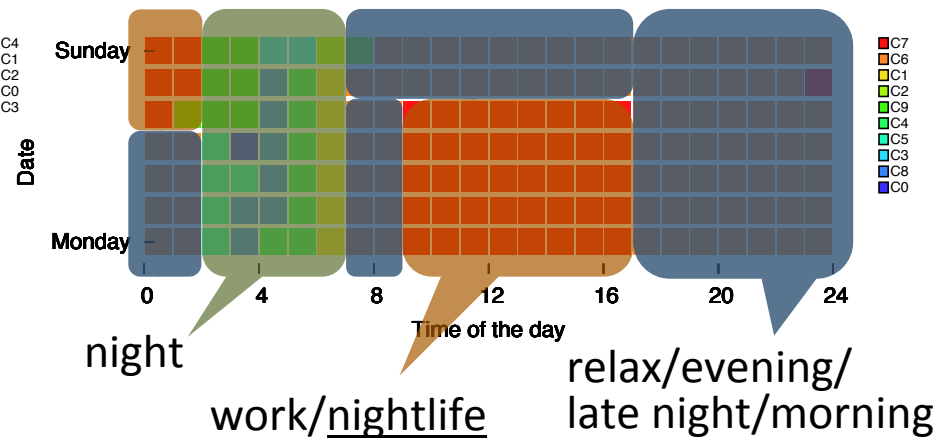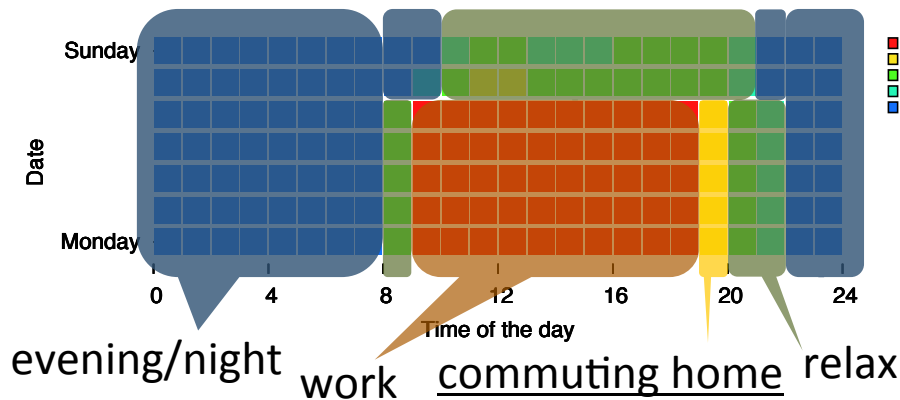*"at which times is network capacity consumed in comparable ways?"*

2. by-product: pinpoint **outlying** network usages

"when is network capacity consumption unusual and why?"

Network analysis

Mobile demand

Technologies

Networking solutions

Marketing solutions

Aggregate access network traffic

Individual access network traffic

Privacy solutions

Networking analysis (III)

- **Example: aggregate access network traffic**
  - Milan case study

evening/night   work   commuting home   relax

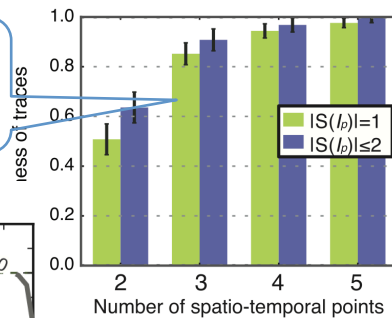night   work/nightlife   relax/evening/ late night/morning

mobile demand outliers

Halloween night, Milan patron saint day, opening of opera season at "La Scala" theater, Christmas holidays and events, several public holidays, several football matches of AC Milan and Inter Milan, one collection probe outage, etc.
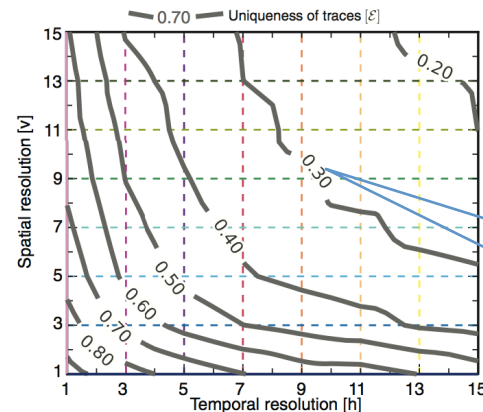
Networking analysis (IV)

- **Example: privacy solutions**
    - k-anonymizing mobile traffic datasets

few random points pinpoint a subscriber

generalization does not help

elevate uniqueness + poor anonymizability
[source: De Montjoye et al., Sci. Rep. 3(1376), 2013]

Network analysis

Mobile demand

Technologies

Networking solutions

Marketing solutions

Aggregate access network traffic

Individual access network traffic
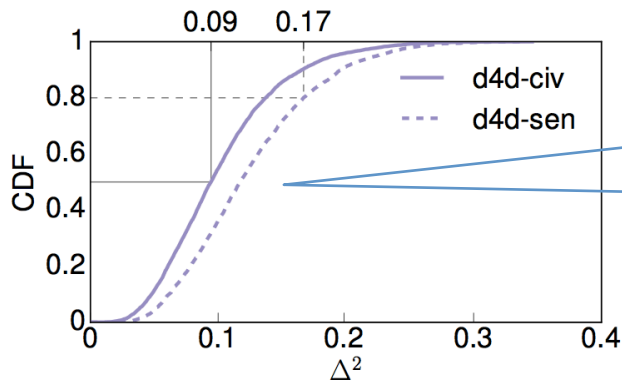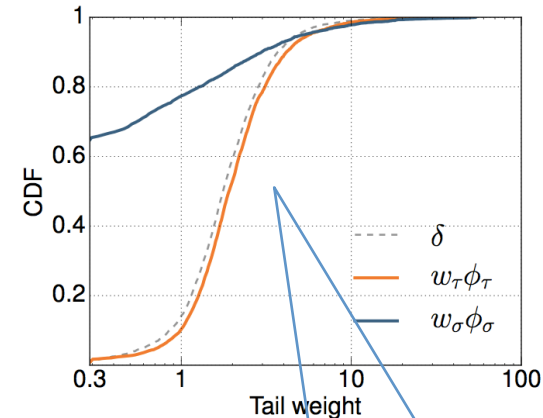
Privacy solutions

Networking analysis (V)

- **Example: privacy solutions**
  - Ivory Coast and Senegal use cases
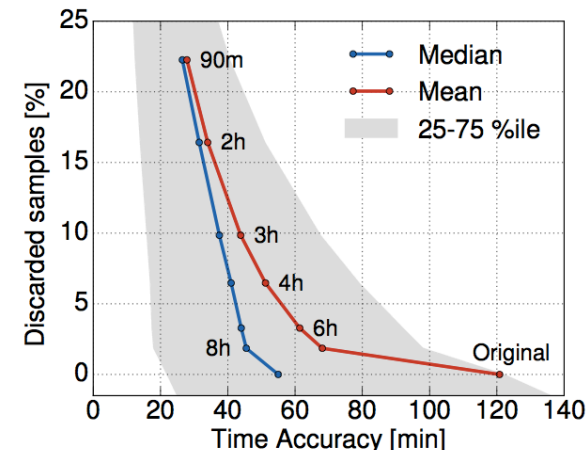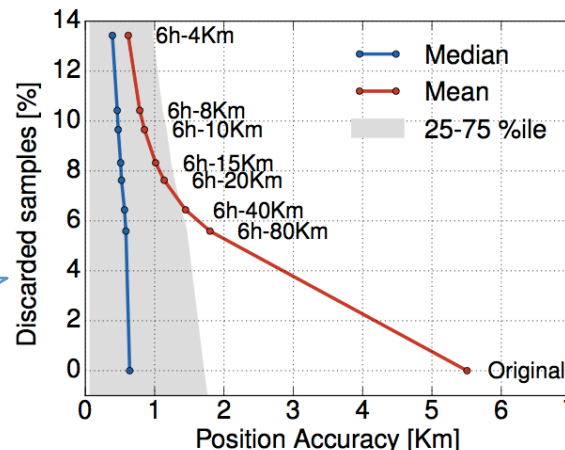
1. measuring "anonymizability" of users in a dataset

for most users, it is easy to find other users that have similar patterns "on average"...

...but, with a long tail of temporally diverse samples!

2. specialized generalization

anonymized dataset accuracy (space and time) under suppression

? Reading list (and questions)

(Additional) Readings

- ## Surveys
  - V.D. Blondel, A. Decuyper, G. Krings, "A survey of results on mobile phone datasets analysis," arXiv:1502.03406, 2015
  - F. Calabrese, L. Ferrari, V.D. Blondel, "Urban Sensing Using Mobile Phone Network Data: A Survey of Research," ACM Computing Surveys 47(2), 2014
  - D. Naboulsi, M. Fiore, S. Ribot, R. Stanica, "Large-scale Mobile Traffic Analysis: a Survey," Inria Technical Report HAL-01132385, 2015

- ## My research activities
  - D. Naboulsi, R. Stanica, M. Fiore, "Classifying Call Profiles in Large-scale Mobile Traffic Datasets", IEEE Infocom, 2014
  - M. Gramaglia, M. Fiore, "On the anonymizability of mobile traffic datasets," NetMob D4D, 2015

- ## Reference community conference
  - NetMob, every year at MIT, Boston, MA, USA