Mobile networks

Applications of Google matrix to directed networks and Big Data

Défi MASTODONS APPLIGOOGLE

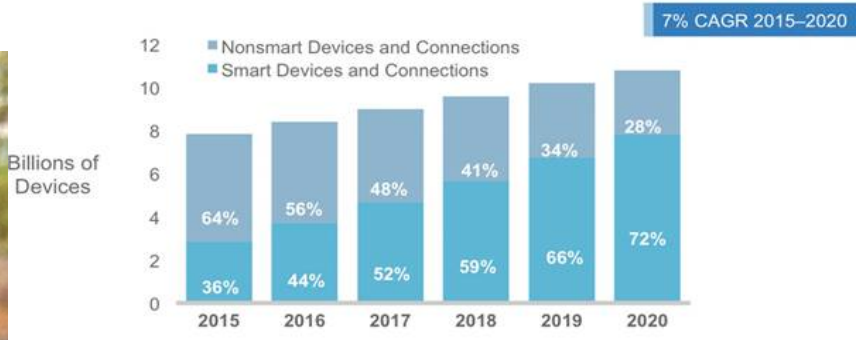
Katia Jaffrès-Runser
University of Toulouse, INPT-ENSEEIHT,
IRIT lab, IRT Team

May 17, 2016

The smartphone phenomenon

- Multiple sensing and communication capabilities
  - Sensors, camera, GPS, microphone
  - 3G, WiFi, Bluetooth, etc.
  - Storage capabilities (several Gbytes)
  - Computing power

[Cisco VNI Global Mobile Data Traffic Forecast (2015-2020)]



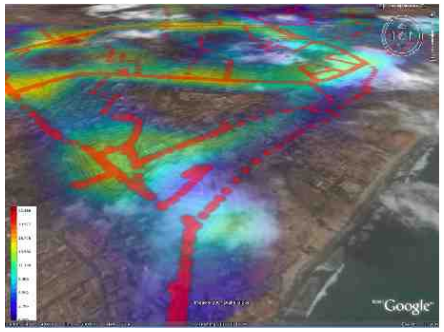
Mobile Traffic is growing constantly

- Global **mobile data traffic grew 74% in 2015** :
  - Global mobile data traffic reached 3.7 exabytes per month
  - A user generates 495 MBytes/month in 2015 up to **3.3 GBytes/month in 2020**
  - 4.8 billion of users in 2015 up to **5.5 billion in 2020**

[Cisco VNI Global Mobile Data Traffic Forecast (2015-2020)]

+

=



Next Big Networking Challenge: meet traffic demand !

1. If data is not delay sensitive:
   - e.g. Videos, Application / system updates, music, podcasts, etc.

Leverage opportunistic encounters to route or flood **delay tolerant** data hop by hop

Benefit: Reduce downloads from infrastructure wireless network

2. If several connectivity options exist:
   - e.g. 3G/4G, WiFi, Femto cells

Offload / Pre-fetch data using
the ‘best‘ available connectivity, at the best time and location

Benefit: **Load balancing** between available infrastructures

Crowdsourcing (part of) this huge network

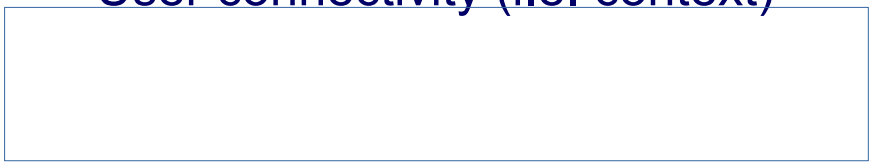
- This huge network of users is constantly active.
  - The **context** each user is evolving is changing all the time
  - The **content** each user is consuming / sending is evolving as well

- To provide the next intelligent data communications, ***we need to understand how this network evolves***

- How is this big dynamic network evolving?
  - Getting network traces
  - Model the interactions of this dynamic network to capture its evolution

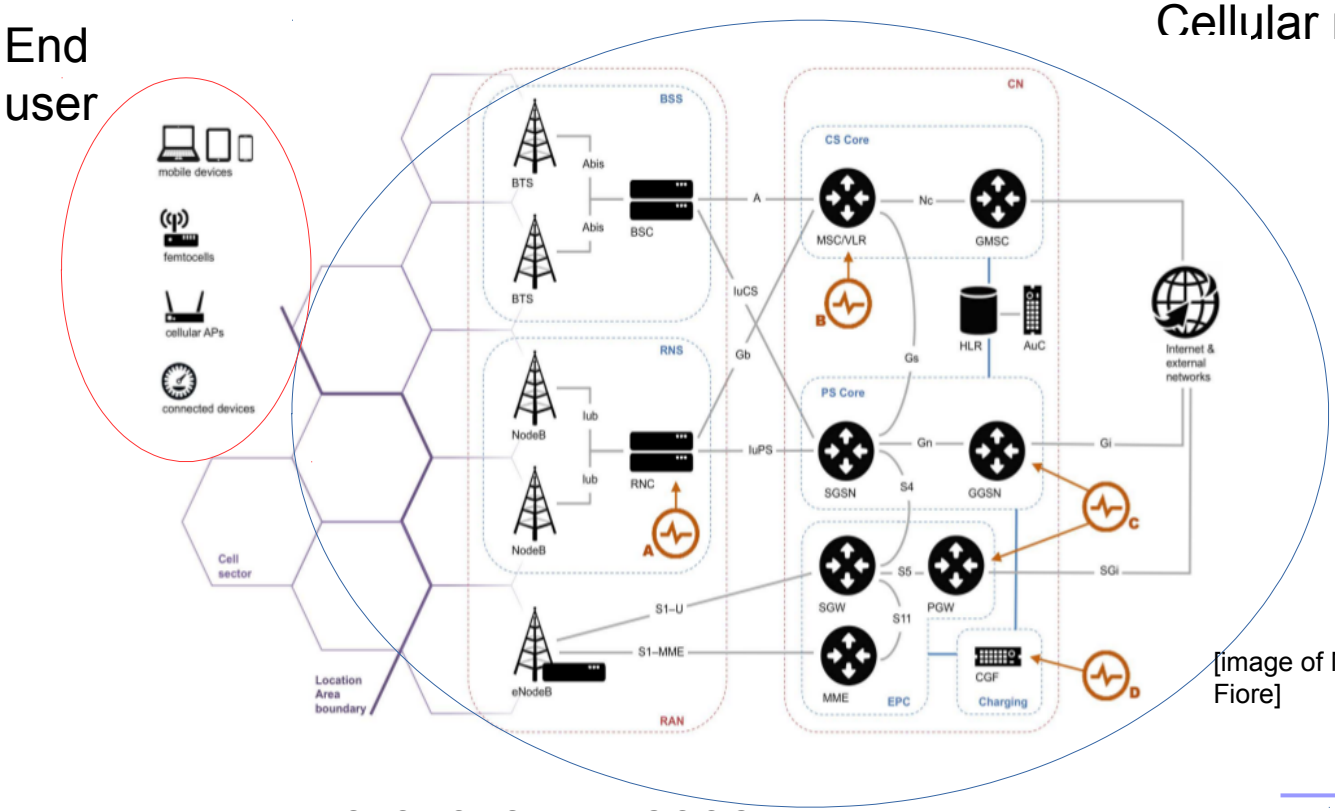
- The goal is to forecast

User demands for content
User connectivity (i.e. context)



Getting network traces

- Two sources of network traces
  - Cellular **network operator** monitoring (Talk of Marco last year)
  - **Monitoring end user activity** on the smartphone



[image of Marco Fiore]

Network operator traces

- Different 'qualities' of traces
  - Depends on place of measurement in the core network
    - Basic call logs, text messages
    - Data sessions for some
    - Location of users: typically at cell level
- Large scale
    - In the order of millions of users (up to 20M)
    - But not very easy to access such data

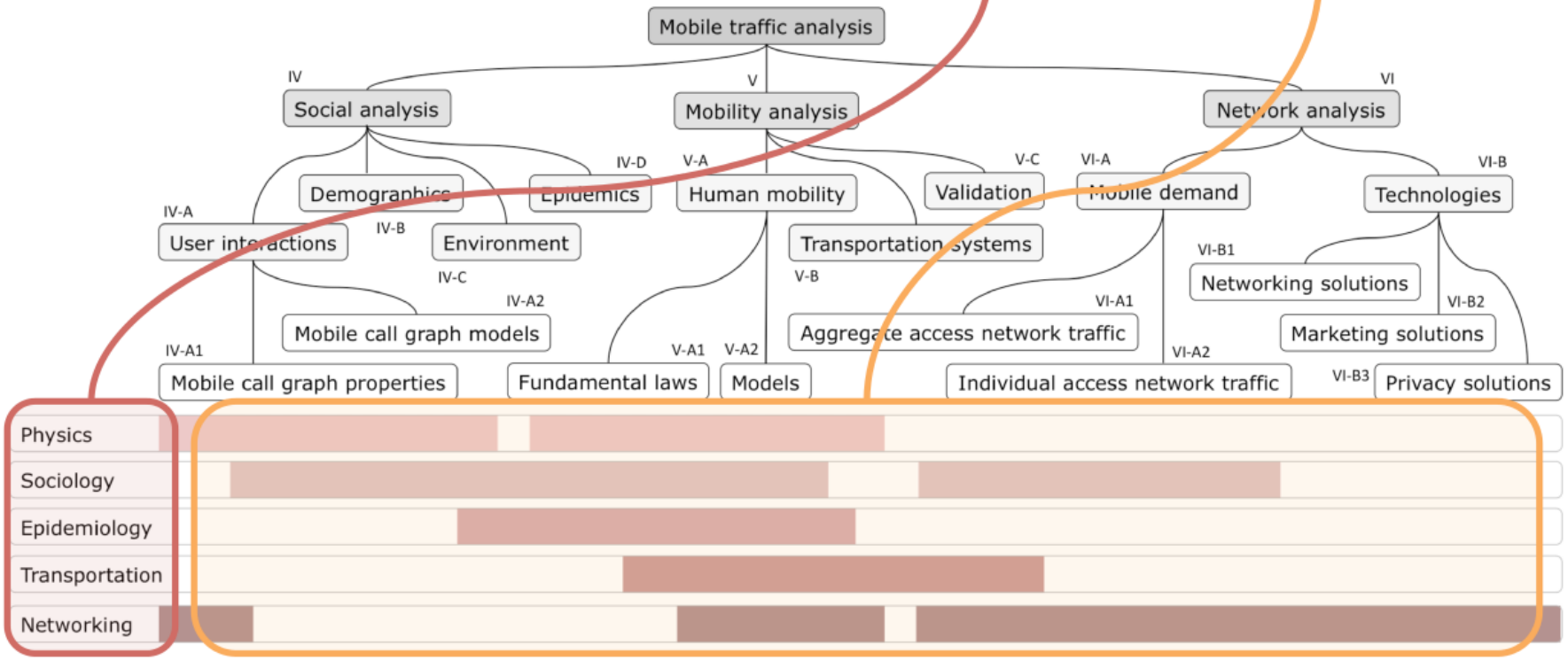
- Not a fine-grained view of user context and content
  - We want to understand and forecast the **content** and **context** of a user

Network model : call graph

- These networks can be modeled in various ways
  - Vertices : users
  - Edges :
    - Directed / undirected
    - Weighted or not
- It depends on how the call records have been modeled
  - Undirected edge if a call has connected 2 users
  - Directed edge if a text message was sent by A to B
- Time dimension
  - Can be captured using weights on edges (persistence)

Network operator traces

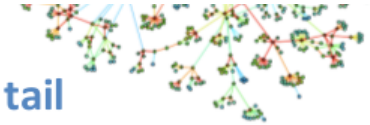
- **A highly multidisciplinary research field**
  - Mobile traffic data is used in many domains (with overlaps)



[Slide of Marco Fiore]

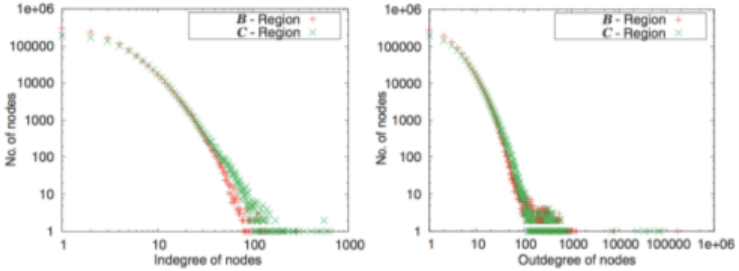
Network operator traces

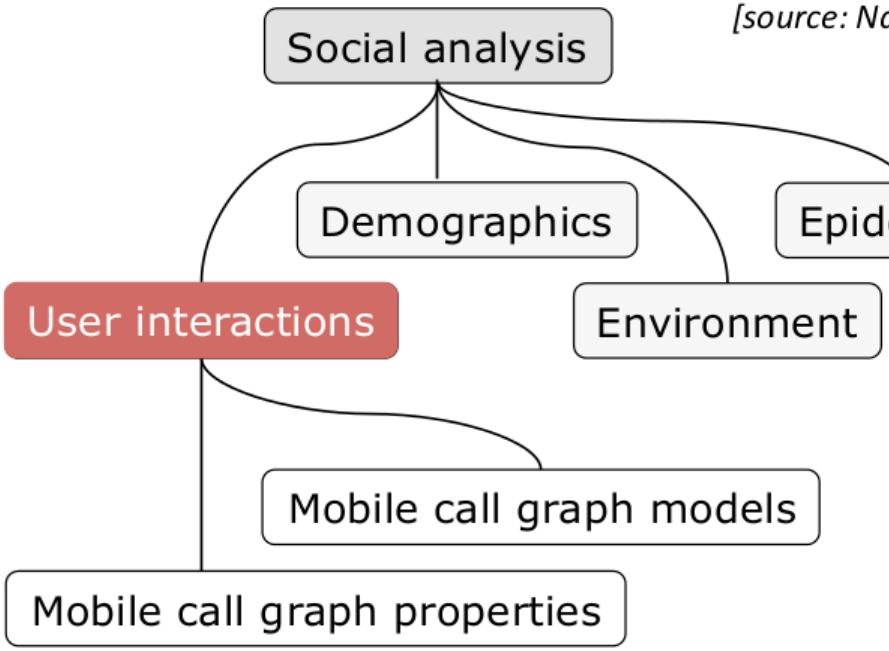
– Mobile call graph analysis

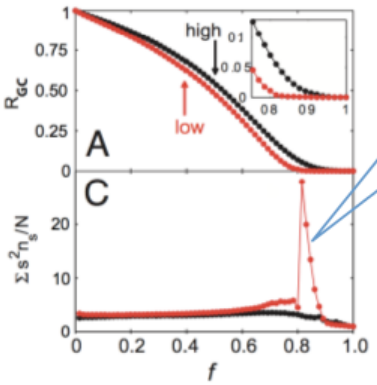


power-law degree distribution tail

[source: Nanavati et al., ACM CIKM, 2006]







network collapse only if low-weight ties are removed first!

strength of weak ties

[source: Onnela et al., PNAS, 104(18), 2007]

[Slide of Marco Fiore]

Crowdsourcing Mobile app

Goal : Sample user context and content data

- Runs in background on **volunteer** phone users
  - Monitors different sensors periodically (5 mins)

- Should be **seamless** with respect to regular phone usage
  - **Upload data** to our servers before memory is full
    - Full memory = no reactivity
    - But : does not ruin the 3G data plan !

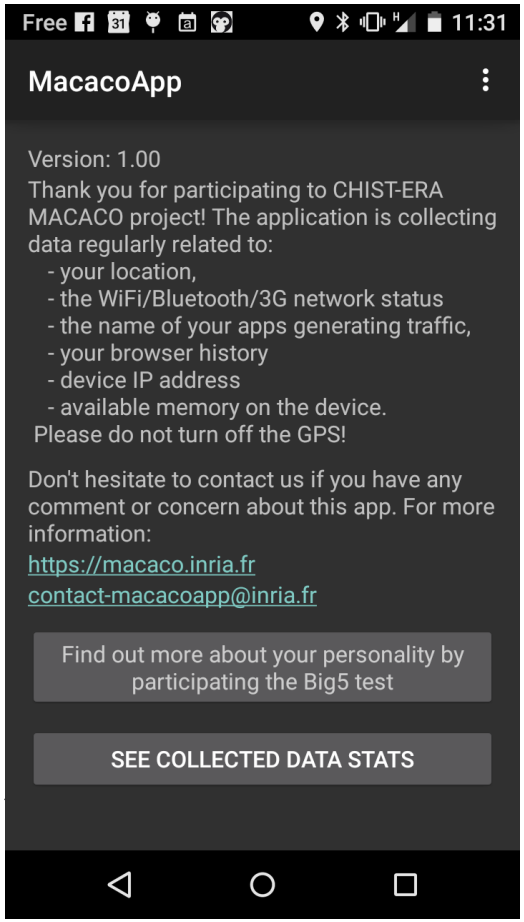
Favor uploads on WiFi

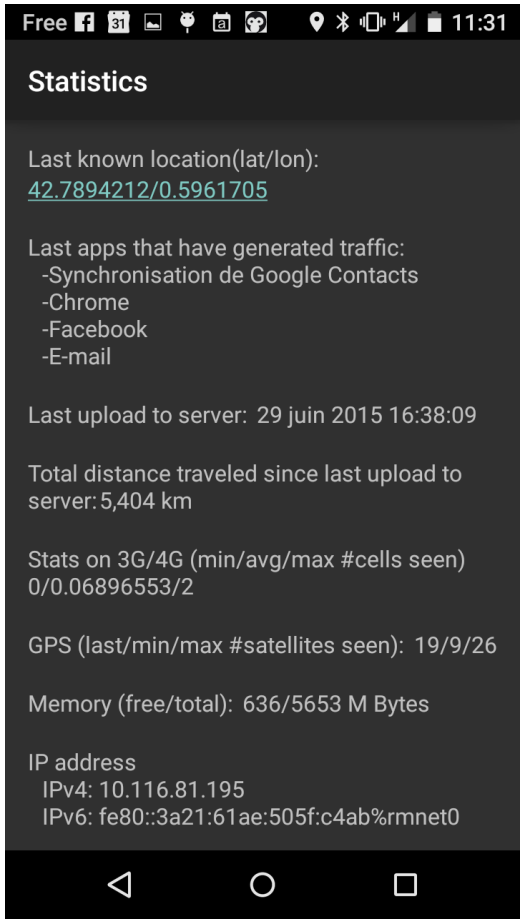
- **Energy** constraint !!
  - Monitoring all sensors is costly

CHIST-ERA macaco project

Mobile context-Adaptive CAching for COntent-centric networking

www.macaco.inria.fr





App Store

macaco

Macaco App

Measured data every 5 minutes :

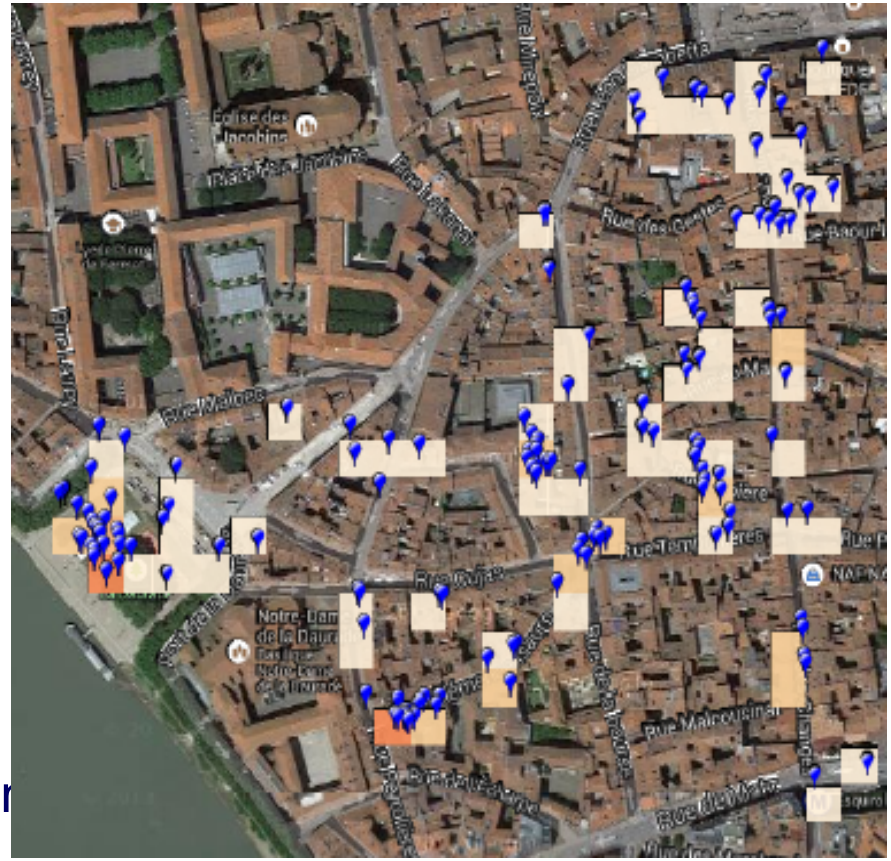
- Context data
  - Location (GPS, Internet)
  - WiFi connectivity
  - Bluetooth connectivity
  - Cellular network towers
  - Battery discharge
  - Accelerometer

- Content data
  - Name of applications

that have generated traffic

  - Volume of traffic
  - Browser history
  - Name of applications running Nar

    of applications



Main issue: getting volunteers :-)

- **Privacy** issues (discussion with CNIL)
  - Keep data within project partners,
  - Have data anonymized (hashed IMEI - location)
  - Limit storage duration of non-anonymized data use
  - Option to remove its own data from the collection

- **Energy efficient app design**
  - Keep the volunteers using the app

- Provide **a motivation** for participating
  - Added value of the app (e.g. visualize its own data, game, …)
  - Financial retribution (voucher)
  - Lottery
  - For the greater good :-) …

Energy aware design

- Energy hungry sensors:
  - **GPS** localisation
    - Unavailable indoors
    - Useless if no motion -> DETECT MOVEMENT
  - **Bluetooth** scan
    - Use Low-Energy bluetooth
    - Useful to detect available opportunistic communications
  - **Accelerometer**
    - Reduce the sampling duration and interval

Current Macaco data statistics

- Collected with MacacoApp in Europe (Jan. 2016)
  - 99 devices
  - ~1,650,000 Measurements

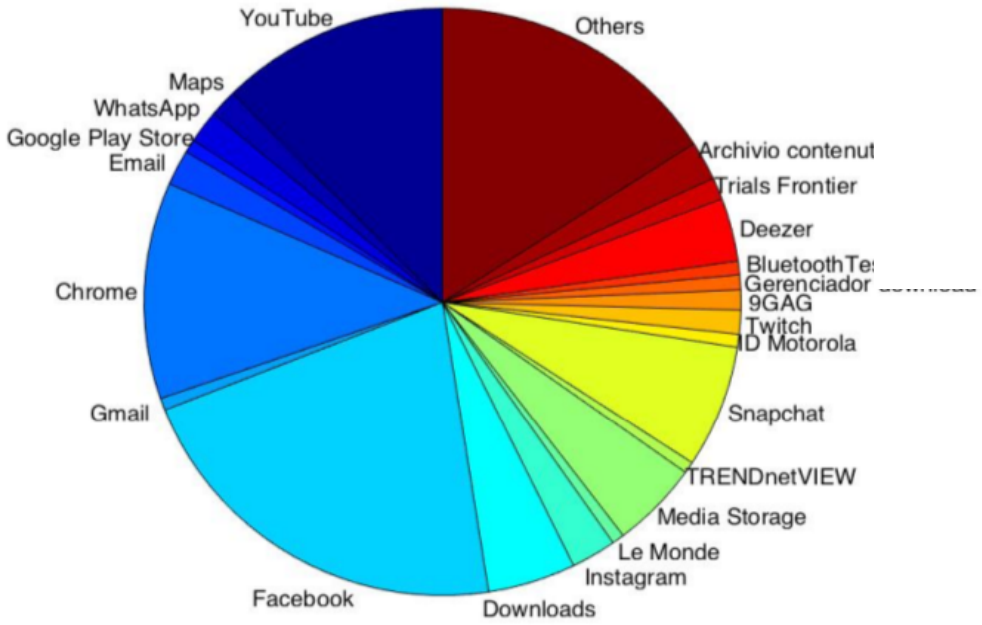
- 2 collection campaigns in China

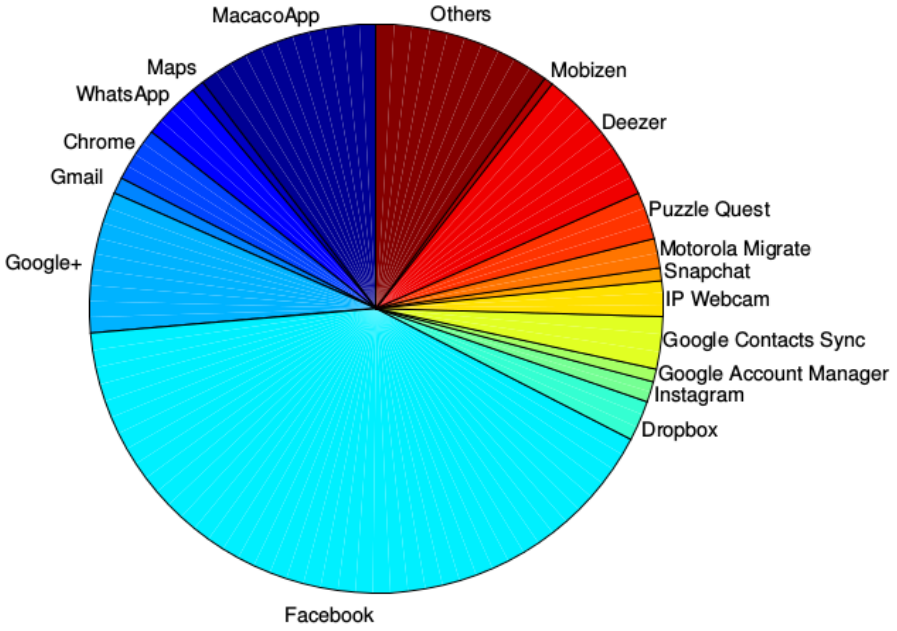
Table 1 Data collection statistics

	Europe	China 1st compaign	China 2nd compaign
Number of phones	98	28	10
Number of measurements	1,603,013	35,941	17,686
Period	18 months (2014-07-13 to 2016-01-20)	1 month (2015-06-24 to 2015-07-27)	4 months (2015-09-30 to 2016-01-12)

Current Macaco data statistics

Total download data:
207,816MB

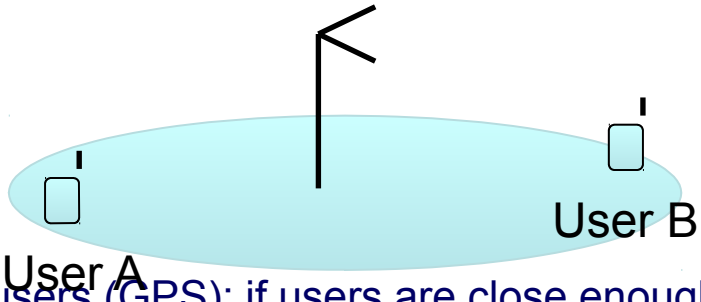




Total upload data:
46,908MB

How to exploit such datasets?

- Other open (less precise) datasets exist (cf. Crawdad http://crawdad.cs.dartmouth.edu/)
- Different types of temporal contact measurements
  - Measure a direct link between User A and B (e.g. Bluetooth, WiFi Direct connectivity)
  - Assume a link exists between User A and User B if they are connected to the same WiFi / cell access point

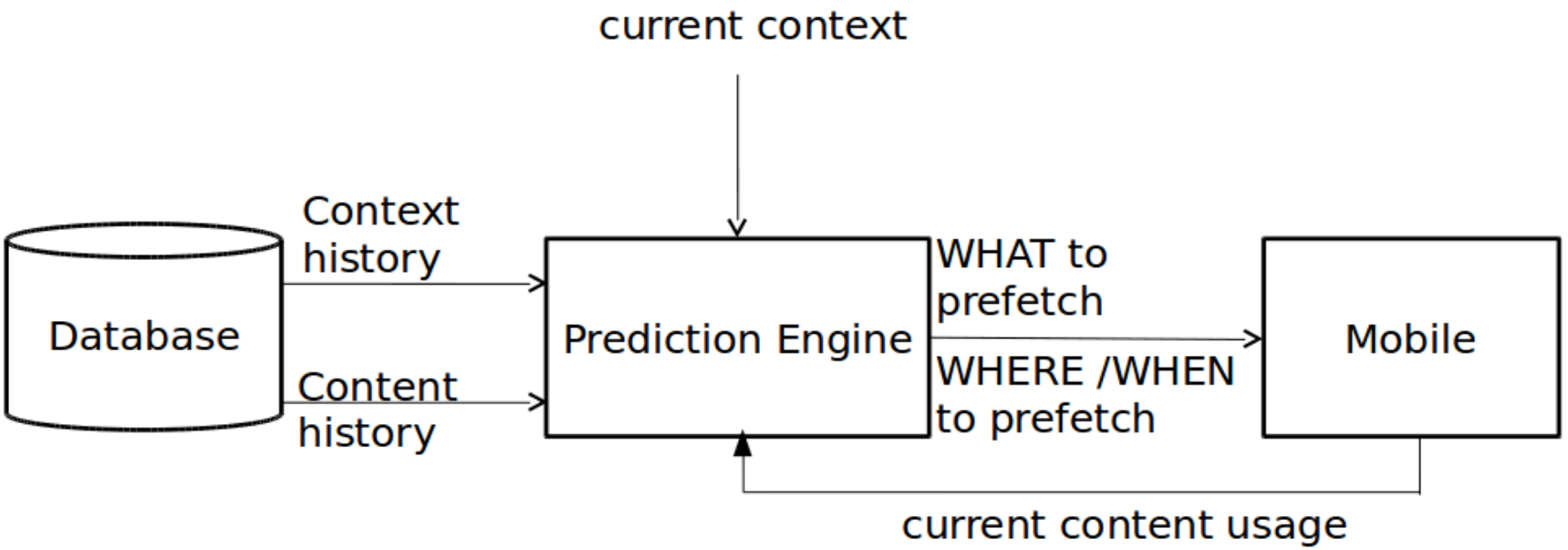


- Measure location of users (GPS): if users are close enough, assume they are connected

MACACO : adds content dimension to context for offloading and prefetching

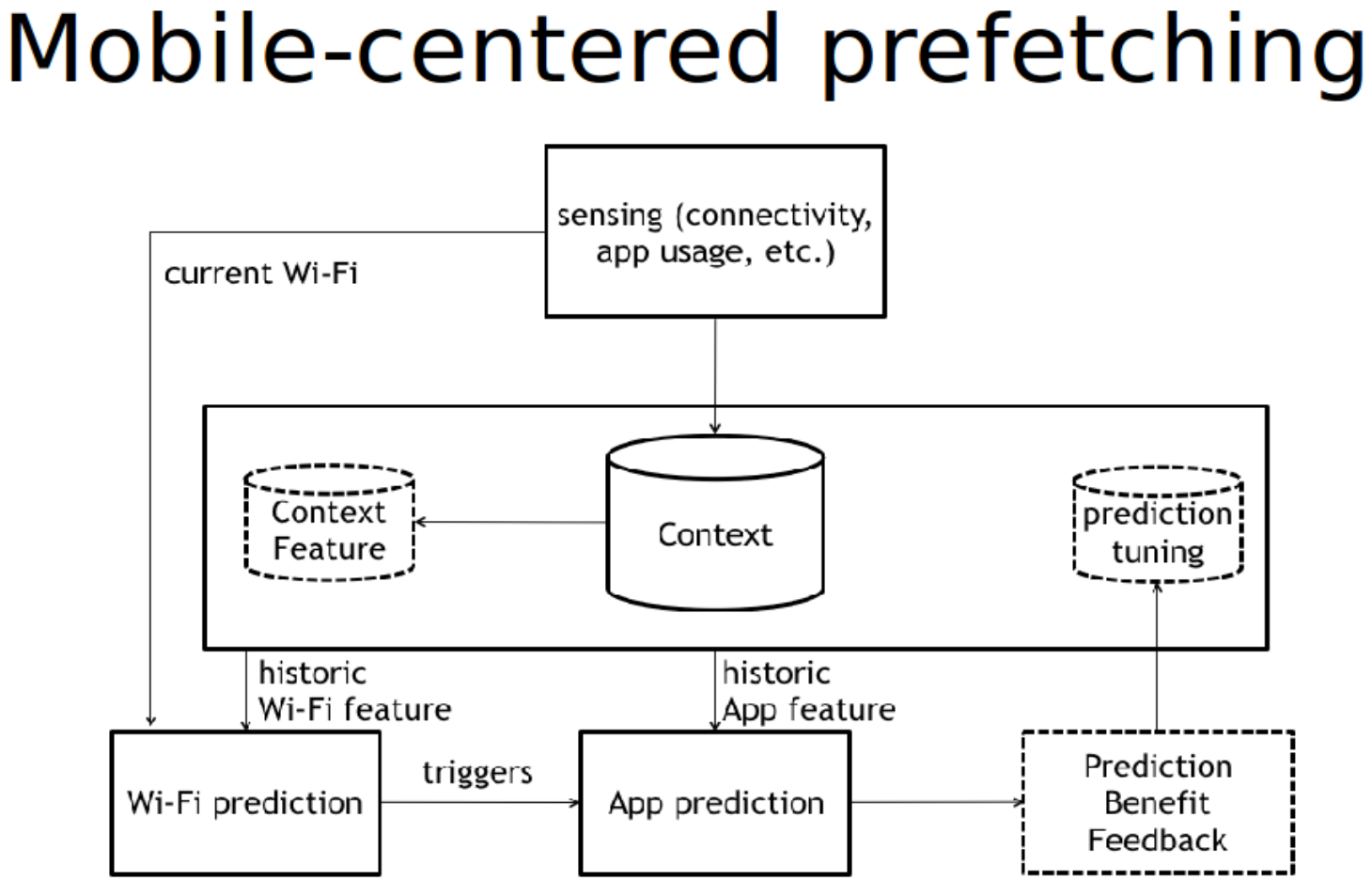
Prefetching data

- General application architecture:



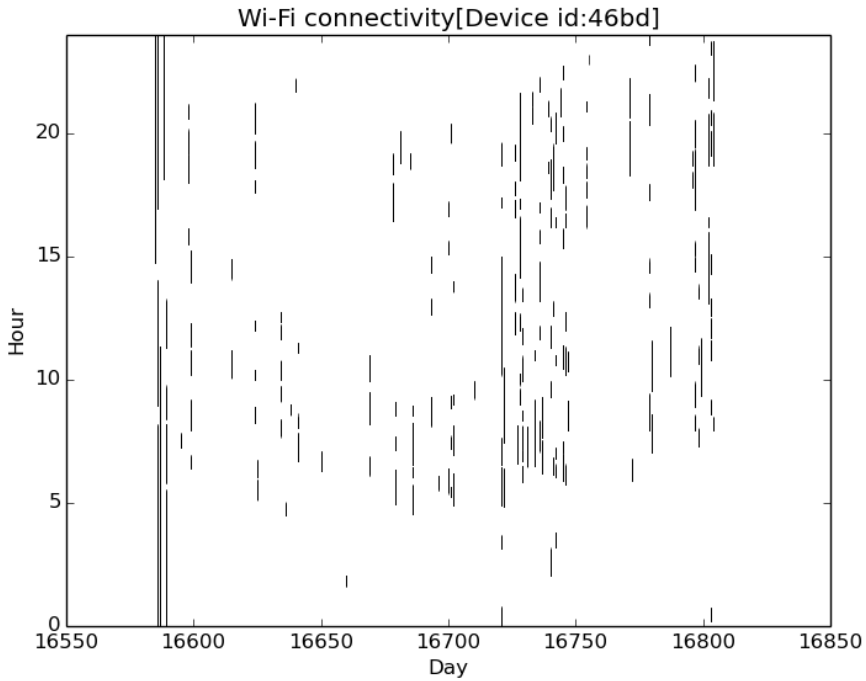
Can be declined onto two types of network architecture :
client/server or mobile-centered

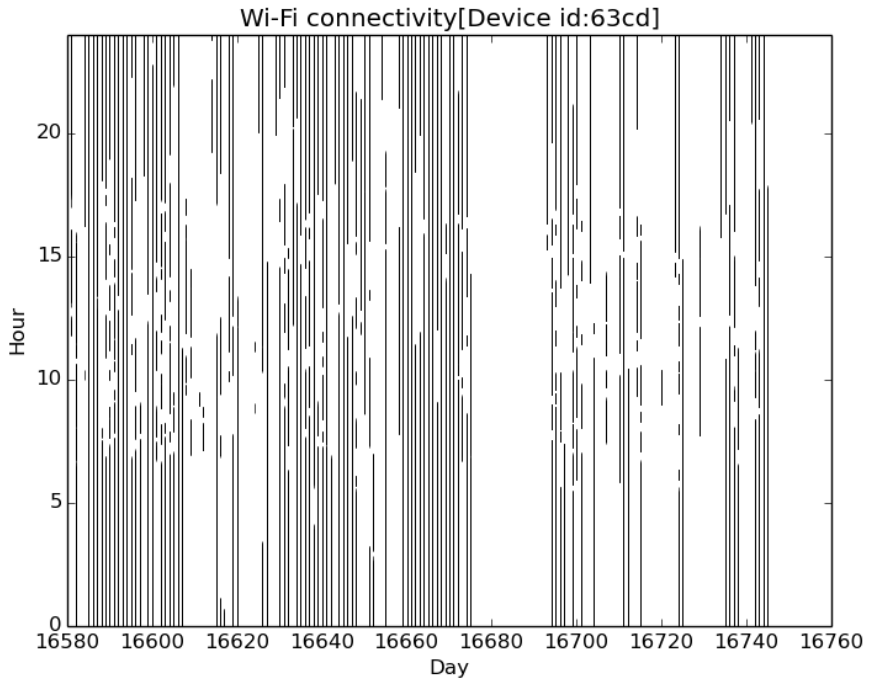
Mobile-centered prefetching



Network models for context (user centered)

When to prefetch data ?





Not all devices have the same opportunities …

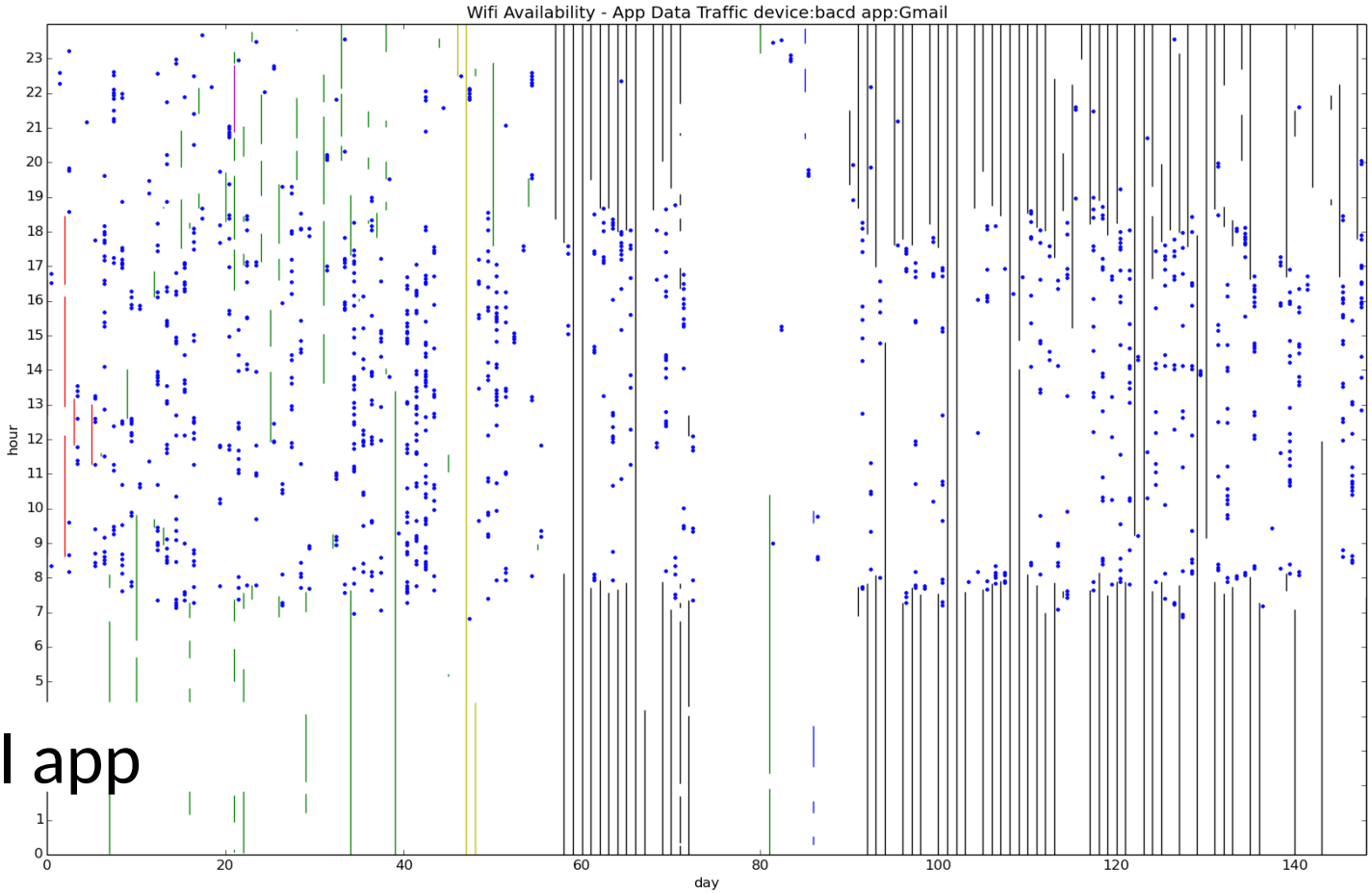
Network models for context (user centered)

- What are the important access points in my life?
  - Monitor the disconnection time distribution
- To select important access points for a user
  - Vertices :
    - Access points I have seen
    - Access points I have used (upload/download)
  - Edges
    - Directed : AP1 → AP2 means AP2 is seen after AP1
    - Weighted : proportion of time this edge has occurred (integrate dynamics)
    - Account for traffic at vertices?

Network gives information on
Context/connectivity of users

Network models for content (user centered)

What app should be prefetched?



Gmail app

Network models for content (user centered)

- For a specific user's near future content prediction
  - Vertices :
    - Applications installed on the phone
    - Applications uploading / downloading data
  - Edges
    - Directed : App1 → App2 means App2 is launched after App1
    - Weighted : proportion of time this edge has occurred (integrate dynamics)

Network gives information on content access of users

Client / serveur prefetching

- Profile users / applications using network wide information
  - Categories of users with similar context / content
  - Categories of apps that can benefit from prefetching

- A mobile node sends its current content and context to a serveur
  - The server sends back prefetching order (which apps, when/where) using its network-wide profiling
  - Leverages this data to update its profiling

Network models for context (all users)

- Vertices :
  - devices only : peer-to-peer connectivity
  - devices AND access points (APs) : devices connected through APs
- Edges
  - Undirected : symmetric channel
  - Directed : upload / download of data
  - Unweighted : binary decision on connectivity
  - Weighted : communication success probability / link availability over time

Network gives information on connectivity of users

Network models for content (all users)

- Study the network wide usage of applications
- Vertices :
  - App X installed on device A
    - Nb of devices * Nb of Apps
- Edges :
  - An edge exists if device A is connected to device B, and both are running App X
    - Subnetwork of users using App X
  - Undirected edges
  - Incorporate traffic volume up/download?

Retrieve information on
Network wide usage of apps

Future works

- Modeling efforts to be made to capture the potential of our data collection
- Use google matrix analysis, among others, to analyse our network models
- Size of these networks is not very big, but there are long measurements (several months)
  - Analyse the dynamics over time