Mobile Social Networks

Jon Crowcroft &Eiko Yoneki&Narseo Vallina Rodriguez

jon.crowcroft@cl.cam.ac.uk,eiko.yoneki@cl.cam.ac.uk

Systems Research Group
University of Cambridge Computer Laboratory

I. Spread of Infectious Diseases

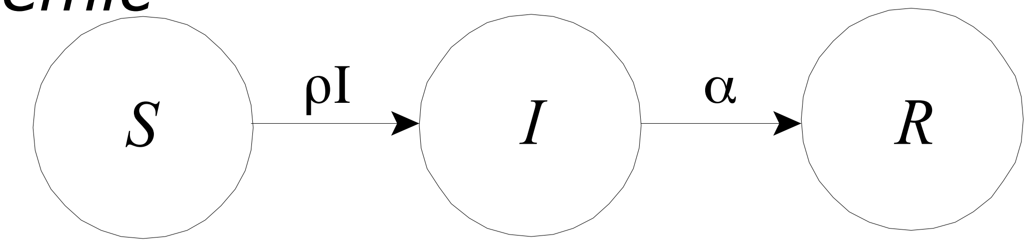
- Thread to public health: e.g., 

, SARS, AIDS
- Current understanding of disease spread dynamics
  - Epidemiology: Small scale empirical work
  - Physics/Math: Mostly large scale abstract/simplified models
- Real-world networks are far more complex

  - Advantage of real world data
  - Emergence of wireless technology for proximity data (tiny wireless sensors, mobile phones...)
  - Post-facto analysis and modelling yield insight into human interactions

- **Model realistic infectious disease epidemics and predictions**

Susceptible, Infected, Recovered: the SIR Model of an Epidemic



What is a Mathematical Model?

a mathematical description of a scenario or situation from the real-world

focuses on specific quantitative features of the scenario, ignores others

a simplification, abstraction, “cartoon”

involves hypotheses that can be tested against real data and refined if desired

one purpose is improved understanding of real-world scenario

e.g. celestial motion, chemical kinetics

The SIR Epidemic Model

First studied, Kermack & McKendrick, 1927.

Consider a disease spread by contact with infected individuals.

Individuals recover from the disease and gain further immunity from it.

S = fraction of *susceptibles* in a population

I = fraction of *infecteds* in a population

R = fraction of *recovereds* in a population

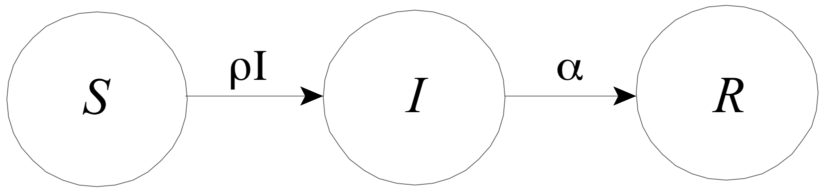
$$S + I + R = 1$$

The SIR Epidemic Model (Cont'd)

- *Differential equations* (involving the variables *S*, *I*, and *R* and their rates of change with respect to time *t*) are

$$\frac{dS}{dt} = -\rho S I, \quad \frac{dI}{dt} = \rho S I - \alpha I, \quad \frac{dR}{dt} = \alpha I$$

- An equivalent *compartment diagram* is



Parameters of the Model

- ρ = the *infection rate*
- α = the *removal rate*
- The *basic reproduction number* is obtained from these parameters:

$$N_R = \rho / \alpha$$

- This number represents the average number of infections caused by one infective in a totally susceptible population. As such, an epidemic can occur only if $N_R > 1$.

Vaccination and Herd Immunity

If only a fraction S_0 of the population is susceptible, the *reproduction number* is $N_R S_0$, and an epidemic can occur only if this number exceeds 1.

Suppose a fraction V of the population is vaccinated against the disease. In this case, $S_0=1-V$ and no epidemic can occur if

$$V > 1 - 1/N_R$$

The *basic reproduction number* N_R can vary from 3 to 5 for smallpox, 16 to 18 for measles, and over 100 for malaria [Keeling, 2001].

Case Study: Boarding School Flu

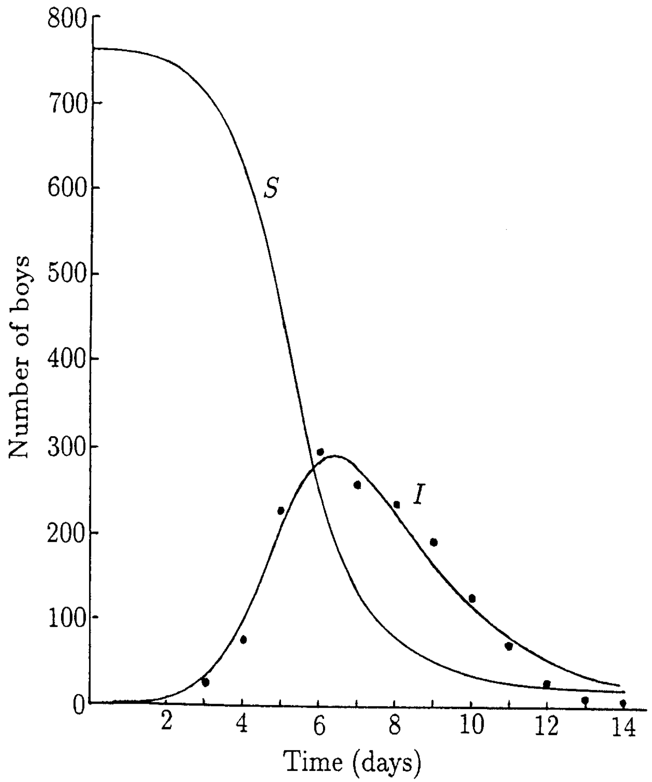
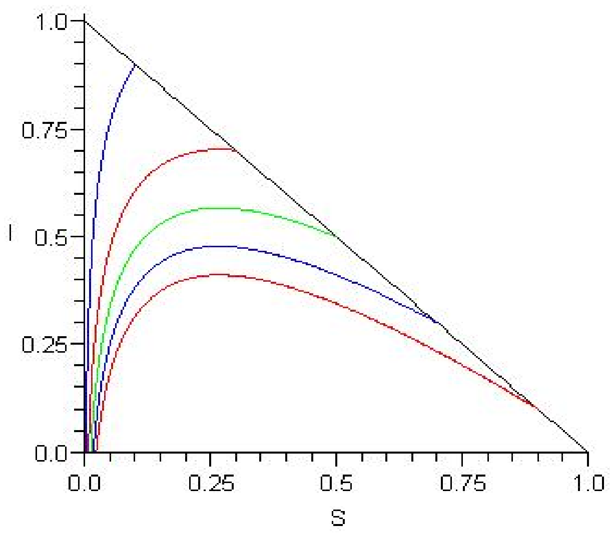
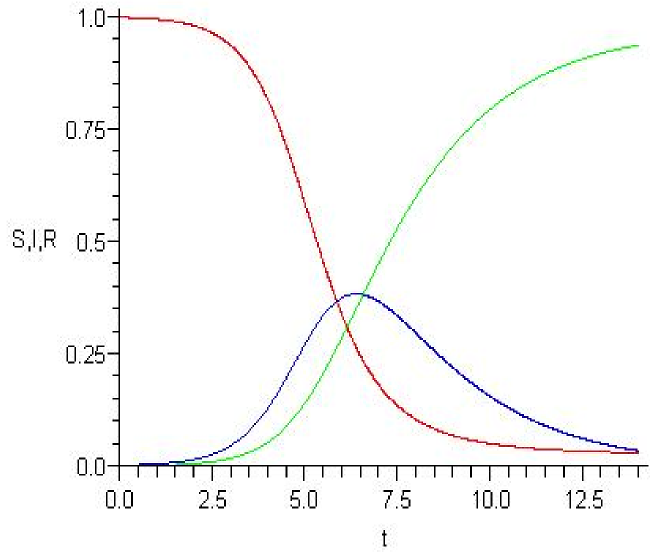


Fig. 19.3. Influenza epidemic data (•) for a boys boarding school as reported in British Medical Journal, 4th March 1978. The continuous curves for the infectives (I) and susceptibles (S) were obtained from a best fit numerical solution of the SIR system (19.1)–(19.3): parameter values $N = 763$, $S_0 = 762$, $I_0 = 1$, $\rho = 202$, $r = 2.18 \times 10^{-3}$/day. The conditions for an epidemic to occur, namely $S_0 > \rho$ is clearly stisfied and the epidemic is severe since R/ρ is not small.

Boarding School Flu (Cont'd)

- In this case, time is measured in days, $\rho = 1.66$, $\alpha = 0.44$, and $R_N = 3.8$.

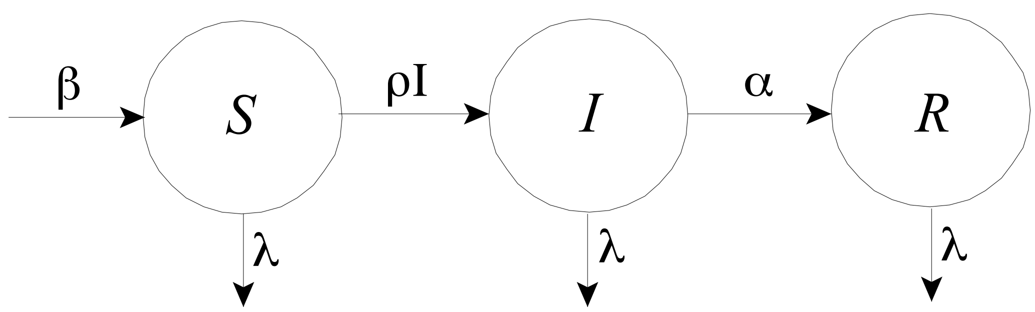




Flu at Hypothetical Hospital

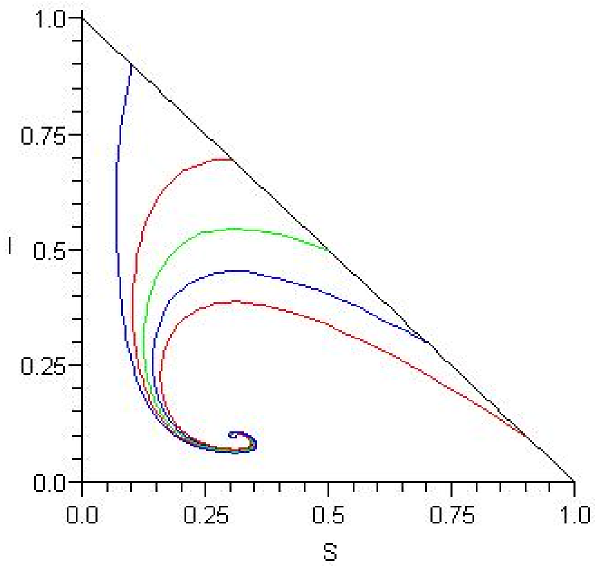
- In this case, new susceptibles are arriving and those of all classes are leaving.

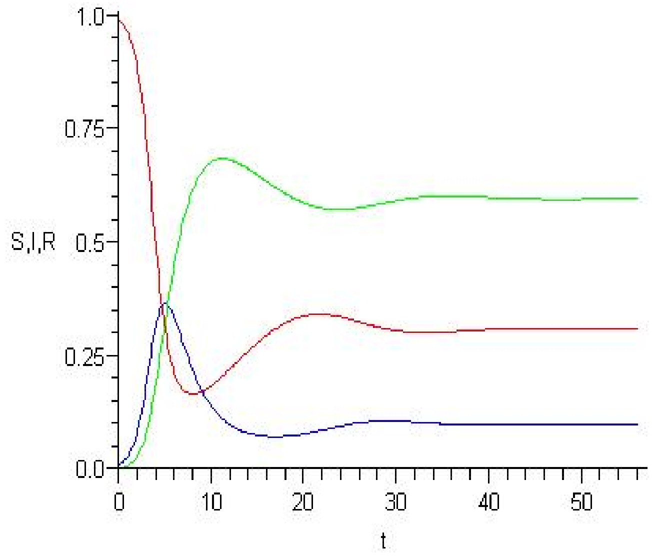
$$\frac{dS}{dt} = \beta - \rho SI - \lambda S, \quad \frac{dI}{dt} = \rho SI - \alpha I - \lambda I, \quad \frac{dR}{dt} = \alpha I - \lambda R$$



Flu at Hypothetical Hospital (Cont'd)

- Parameters ρ and α are as before. New parameters $\beta = \lambda = 1/14$, representing an average turnover time of 14 days. The disease becomes *endemic*.





Case Study: Bombay Plague, 1905-6

- The *R* in SIR often means *removed* (due to death, quarantine, etc.), not *recovered*.

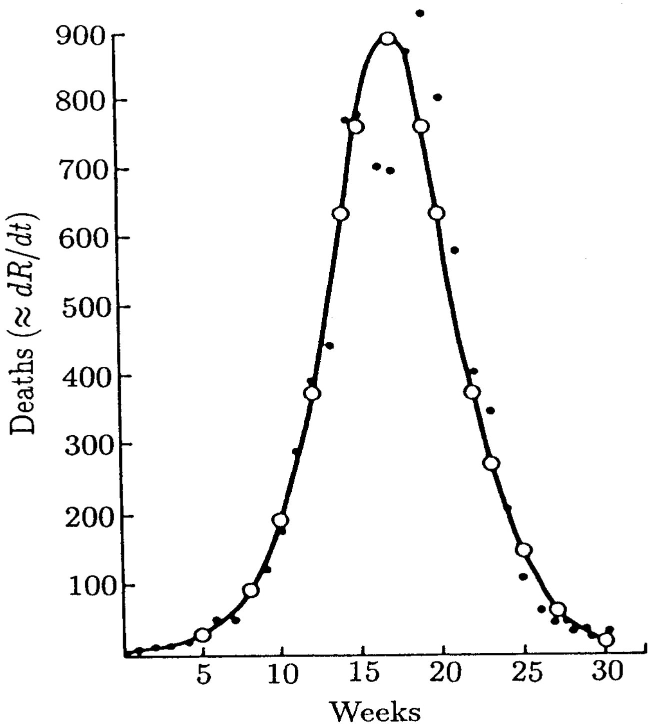


Fig. 19.2. Bombay plague epidemic of 1905–6. Comparison between the data (•) and theory (∘) from the (small) epidemic model and where the number of deaths is approximately dR/dt given by (19.16). (After Kermack and McKendrick 1927)

Eyam Plague, 1665-66

Raggett (1982) applied the SIR model to the famous Eyam Plague of 1665-66.

http://www.warwick.ac.uk/statsdept/staff/WSK/Courses/ST333/eyam.html

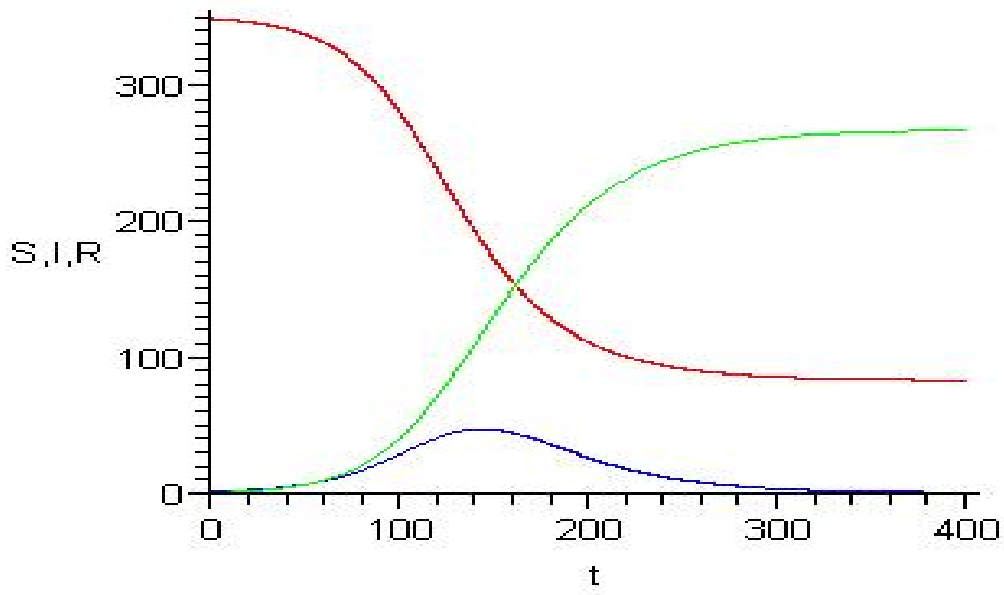
It began when some cloth infested with infected fleas arrived from London. George Vicars, the village tailor, was the first to die.

Of the 350 inhabitants of the village, all but 83 of them died from September 1665 to November 1666.

Rev. Wm. Mompesson, the village parson, convinced the villagers to essentially quarantine themselves to prevent the spread of the epidemic to neighboring villages, e.g. Sheffield.

Eyam Plague, 1665-66 (Cont'd)

- In this case, a rough fit of the data to the SIR model yields a basic reproduction number of R_N=



Enhancing the SIR Model

Can consider additional populations of disease vectors (e.g. fleas, rats).

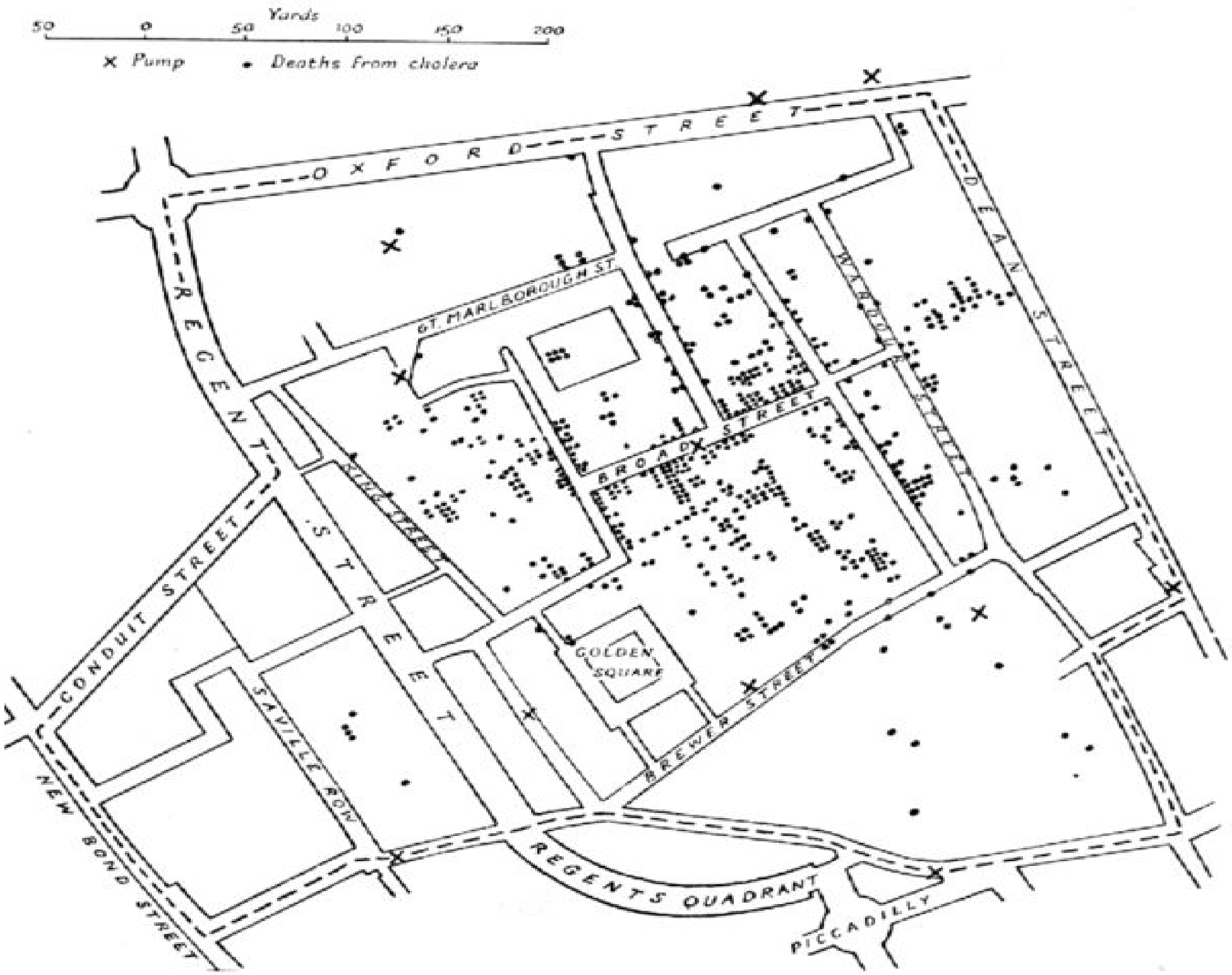
Can consider an *exposed* (but not yet infected) class, the SEIR model.

SIRS, SIS, and double (gendered) models are sometimes used for sexually transmitted diseases.

Can consider biased mixing, age differences, multiple types of transmission, geographic spread, etc.

Enhancements often require more compartments.

Geo-mapping,, Snow's Ghost Map



We meet, we connect, we communicate

We *meet* in real life in the real world

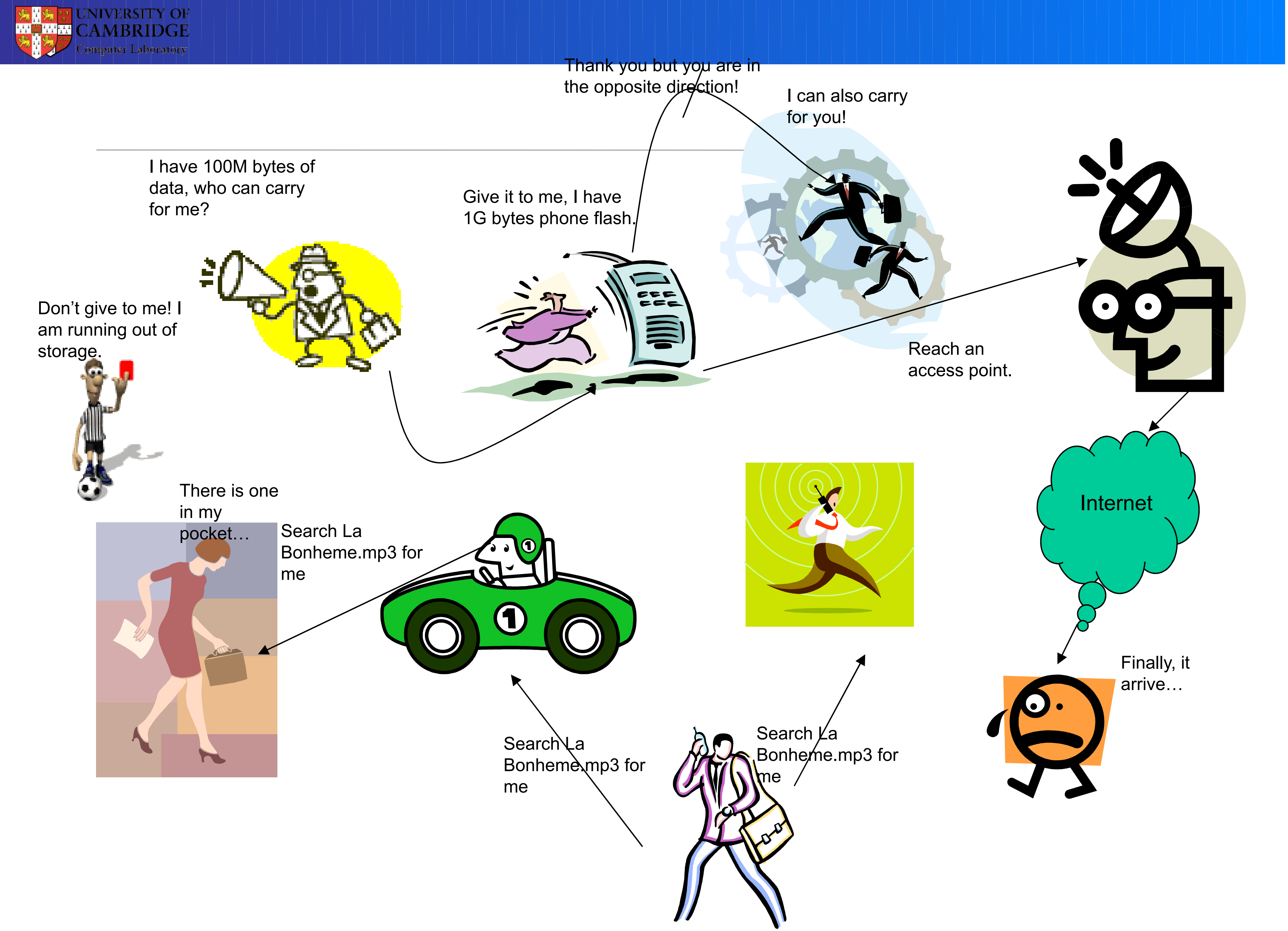
We use text *messages*, phones, IM

We make *friends* on facebook, Second Life

How are these **related**?

How do they **affect** each other?

How do they **change** with new technology?

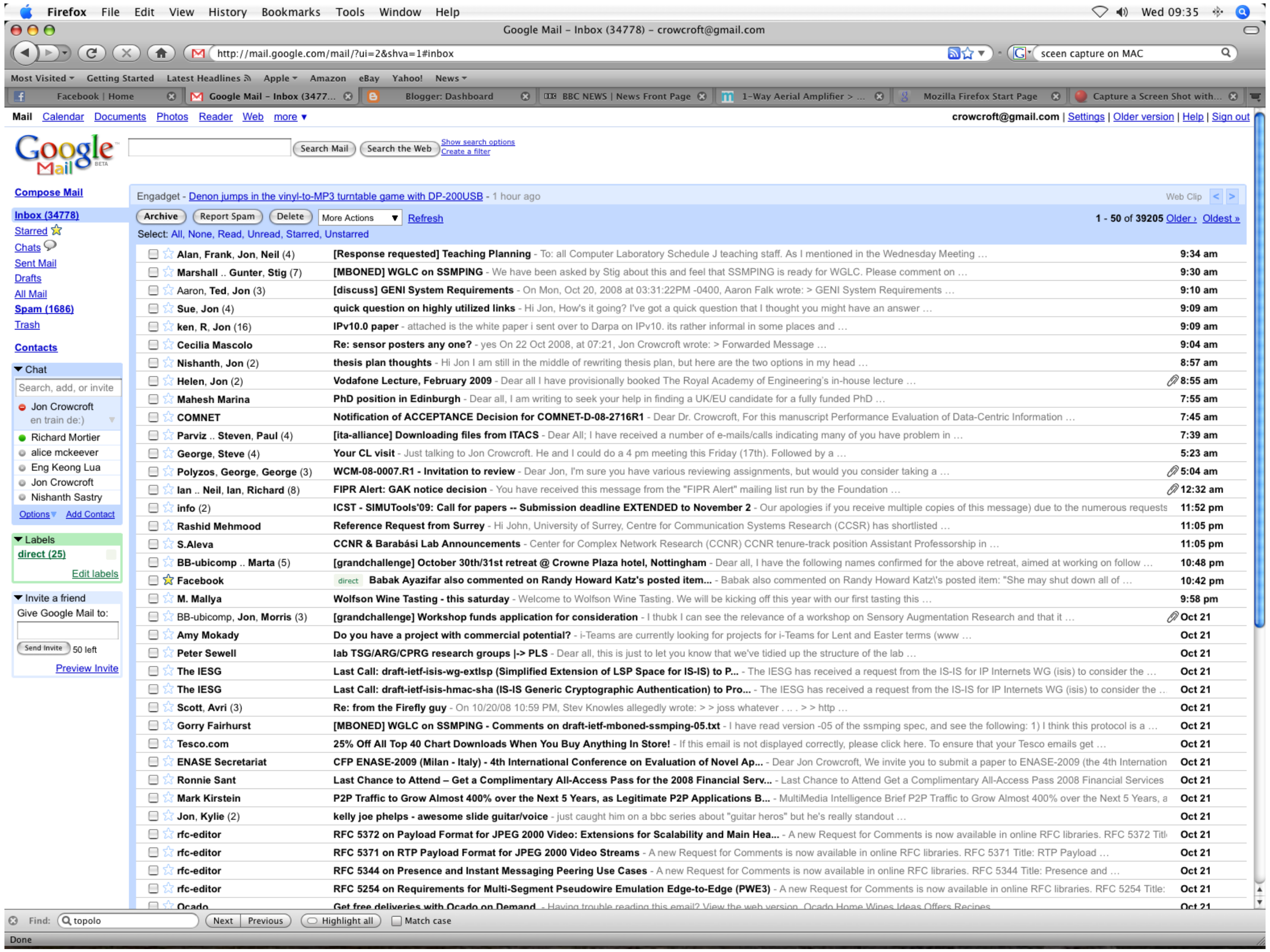


UNIVERSITY OF CAMBRIDGE Computer Laboratory
Thank you but you are in the opposite direction!
I can also carry for you!
I have 100M bytes of data, who can carry for me?
Give it to me, I have 1G bytes phone flash.
Don't give to me! I am running out of storage.
Reach an access point.
Internet
There is one in my pocket…
Search La Bonheme.mp3 for me
Search La Bonheme.mp3 for me
Search La Bonheme.mp3 for me
Finally, it arrive…

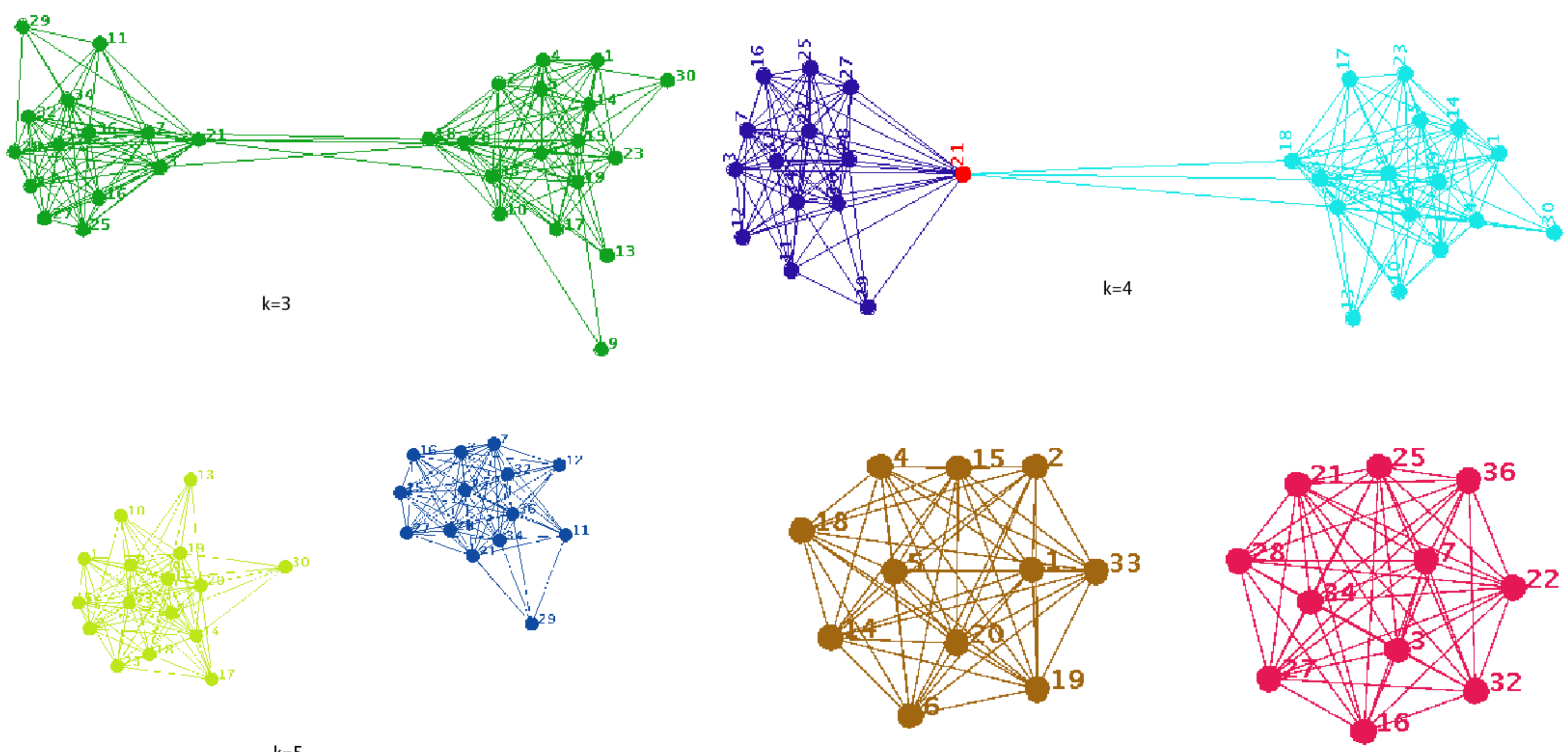
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

My facebook friendswheel

My email statistics!



Cliques and Communities



k=3

k=4

k=5

k=10

We are still learning about this!

There are big problems understanding this

Data?

Privacy?

Usefulness?

Spread of Infectious Diseases

- Thread to public health: e.g., 

, SARS, AIDS
- Current understanding of disease spread dynamics
  - Epidemiology: Small scale empirical work
  - Physics/Math: Mostly large scale abstract/simplified models
- Real-world networks are far more complex

  - Advantage of real world data
  - Emergence of wireless technology for proximity data (tiny wireless sensors, mobile phones...)
  - Post-facto analysis and modelling yield insight into human interactions

- **Model realistic infectious disease epidemics and predictions**

The FluPhone Project

- Understanding behavioural responses to infectious disease outbreaks
- Proximity data collection using mobile phone from general public in Cambridge

https://www.fluphone.org

FluPhone UNIVERSITY OF CAMBRIDGE UNIVERSITY OF LIVERPOOL E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL MRC Medical Research Council

Main page | Information | Help | Contact us   Email: Password: Log in Forgotten your password?

FluPhone Study

This is the home page for the FluPhone study. A study to measure social encounters made between people, using their mobile phones, to better understand how infectious diseases, like 'flu, can spread between people.

This study will record how often different people (who may not know each other) come close to one another, as part of their everyday lives. To do this, we will ask volunteers to install a small piece of software (called FluPhone) on their mobile phones and to carry their phones with them during their normal day-to-day activities. The software will look for other nearby phones periodically using Bluetooth, record this information and send it back to the research team via the cellular phone data service. This information will give us a much better understanding of how often people congregate into small groups or crowds, such as when commuting or through work or leisure activities. Also, by knowing which phones come close to one another, we will be able to work out how far apart people actually are, and how fast diseases could spread within communities. We are also asking participants to inform us of any influenza-like symptoms they may experience during the study period, so that we can match the spread of 'flu to the underlying social network of encounters made.

If you wish to take part in this study, please read the study information below, and then click **here** to start the registration process.

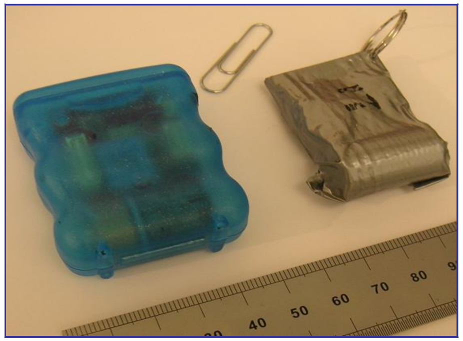
News:

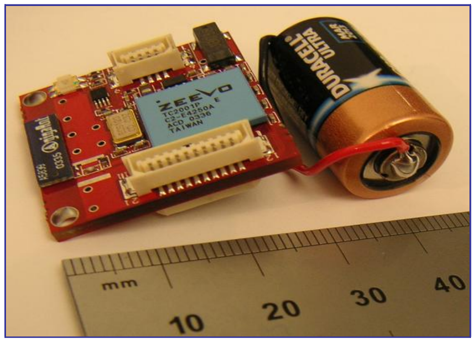
- The pilot study within the university will start on the April 1st, 2010
- The webpage is up!

Various Data Collection

- Flu-like symptoms
- Proximity detection by Bluetooth
- Environmental information (e.g. in train, on road)
- Feedback to users
  - (e.g. How many contacts past hours/days)
- Towards potential health-care app
- Extending Android/iPhone platforms

iMote





Sensor Board or Phone or ...

- iMote needs disposable battery
  - Expensive
  - Third world experiment
- Mobile phone
  - Rechargeable
  - Additional functions (messaging, tracing)
  - Smart phone: location assist applications
- Provide device or software

Phone Price vs Functionality

- ~<20 GBP range
  - Single task (no phone call when application is running)
- ~>100 GBP
  - GPS capability
  - Multiple tasks – run application as a background job
- Challenge to provide software for every operation system of mobile phone
- FluPhone
  - Mid range Java capable phones (w/ Blutooth JSR82 –Nokia)
  - Not yet supported (iPhone, Android, Blackberry)

Experiment Parameters vs Data Quality

- Battery life vs Granularity of detection interval
- Duration of experiments
  - Day, week, month, or year?
  - Data rate
- Data Storage
  - Contact /GPS data <50K per device per day (in compressed format)
  - Server data storage for receiving data from devices
  - Extend storage by larger memory card
- Collected data using different parameters or methods → aggregated?

Proximity Detection by Bluetooth

nly ~=15% of devices Bluetooth on

canning Interval

- 5 mins phone (one day battery life)

luetooth inquiry (e.g. 5.12 seconds) gives >90% chance of finding device

omplex discovery protocol

- Two modes: discovery and being discovered

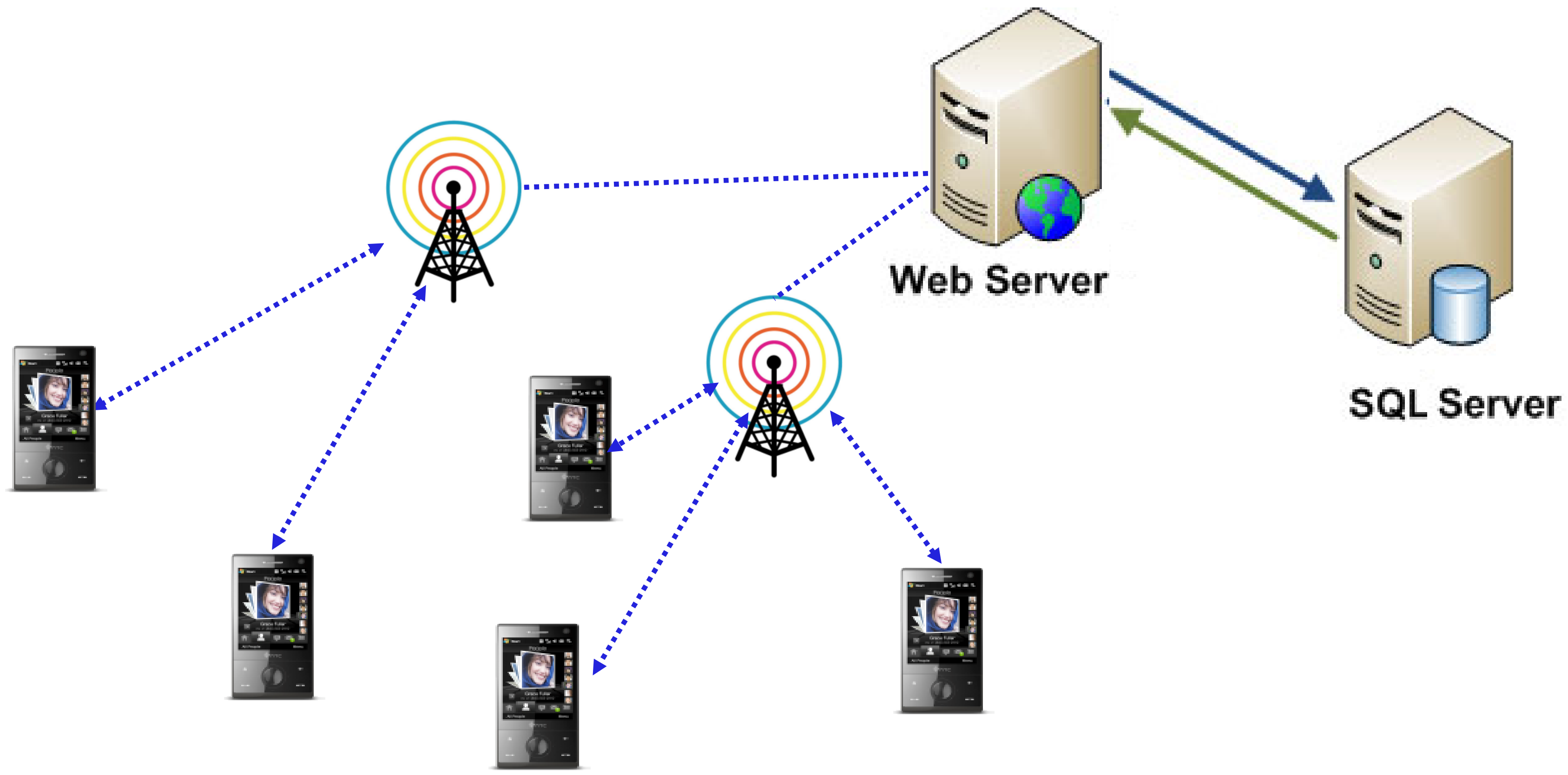
Make sure to produce reliable data!

Data Retrieval Methods

- Retrieving collected data:
  - Tracking station
  - Online (3G, SMS)
  - Uploading via Web
  - via memory card
- Incentive for participating experiments
- Collection cycle: real-time, day, or week?

FluPhone Server

- Via GPRS/3G FluPhone server collects data



Security and Privacy

- Current method: Basic anonymisation of identities (MAC address)
- FluPhone server – use of HTTPS for data transmission via GPRS/3G

- Anonymising identities may not be enough?
  - Simple anonymisation does not prevent to be found the social graph

- Ethic approval tough!
  - ~40 pages of study protocol document for FluPhone project – took several months to get approval

Currently No Location Data

- Location data necessary?
  - Ethic approval gets tougher
  - Use of WiFi Access Points or Cell Towers
  - Use of GPS but not inside of buildings
- Infer location using various information
  - Online Data (Social Network Services, Google)
  - Us of limited location information – Post localisation


Bluetooth Big Brother uses mobiles and laptops to track thousands of Britons

By DAVID DERBYSHIRE
Last updated at 8:15 AM on 22nd July 2008

Comments (14) | Add to My Stories

A Big Brother network of hidden scanners is monitoring hundreds of thousands of Britons without their knowledge, it emerged yesterday.

Scientists track people walking around cities, using the Bluetooth signals from their mobiles, laptops and handheld computers.

Scanners in bars, offices and universities register nearby Bluetooth devices and send the information to a central database.

Scanner Location in Bath

Consent

FluPhone | UNIVERSITY OF CAMBRIDGE | UNIVERSITY OF LIVERPOOL | E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL | MRC Medical Research Council

Main page | Information | Help | Contact us

Email: Password: Log in
Forgotten your password?

Consent

This page asks for you to agree to be a participant in the study. It is important that you understand and agree with the consent and study information.

See here for information about this study.

You must read this software disclaimer.

If you are under 16 years old, we require that your parent or carer provide their consent for you to take part in this study. You must be over 12 years old to take part in this study.

To participate in this study, you must:

- have read, understand and agree with the consent and study information;
- have access to a compatible mobile phone;
- have read, understand and agree with the Fluphone software disclaimer;
- agree to download and run the FluPhone software application on your mobile phone;
- agree for the Fluphone application to use the Bluetooth function of your mobile phone;
- agree for the Fluphone application to send the data it collects to the study team via your network connection , and that this may incur a cost to the billpayer of the phone;
- be willing to allow your data and the information that the FluPhone software collects on your phone to be used for scientific research by the study team.

Please note: You have to fill in all fields.

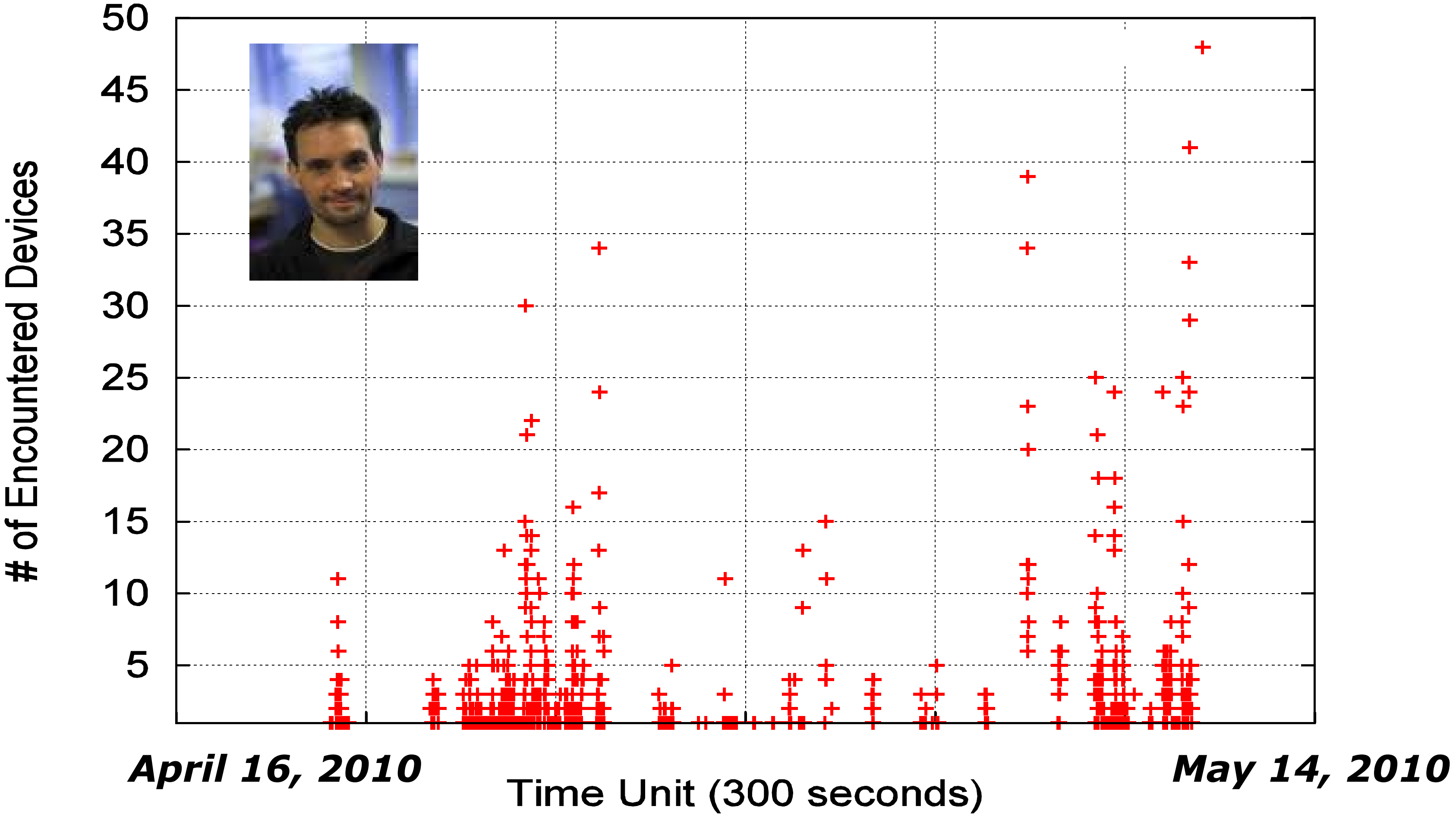
1. Please click the appropriate one:	○ I am under 16 years old, but over 12 years old, and have parental or carer consent for me to take part in this study. ○ I am over 16 years old and wish to take part in this study.
2. Please click the appropriate one:	☐ I have the permission of the bill payer to use this mobile in the study.
	Submit

Study Status

- Pilot study (April 21 ~ May 15)
  - Computer Laboratory
  - Very few participants – people do not worry flu in summer
- University scale study (May 15 ~ June 30)
  - Advertisement (all departments, 35 colleges, student union, industry support club, Twitter, Facebook...)
  - Employees of University of Cambridge, their families, and any residents or people who work in Cambridge
  - Issues
    - Limited phone models are supported
    - Slightly complex installation process
    - Motivation to participate...

Encountered Bluetooth Devices

- A FluPhone Participant Encountering History



UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Existing Human Connectivity Traces

- Existing traces of contact networks
- ..thus far not a large scale

Experimental data set	MIT	UCSD	CAM	INFC06	BATH
Device	Phone	PDA	iMote	iMote	PC
Network type	Bluetooth	WiFi	Bluetooth	Bluetooth	Bluetooth
Duration (days)	246	77	11	3	5.5
Granularity (seconds)	300	600	120	120	Continuous
Number of Experimental Devices	97	274	36	78	7431

- Let's use Cambridge trace data to demonstrate what we can do with FluPhone data...

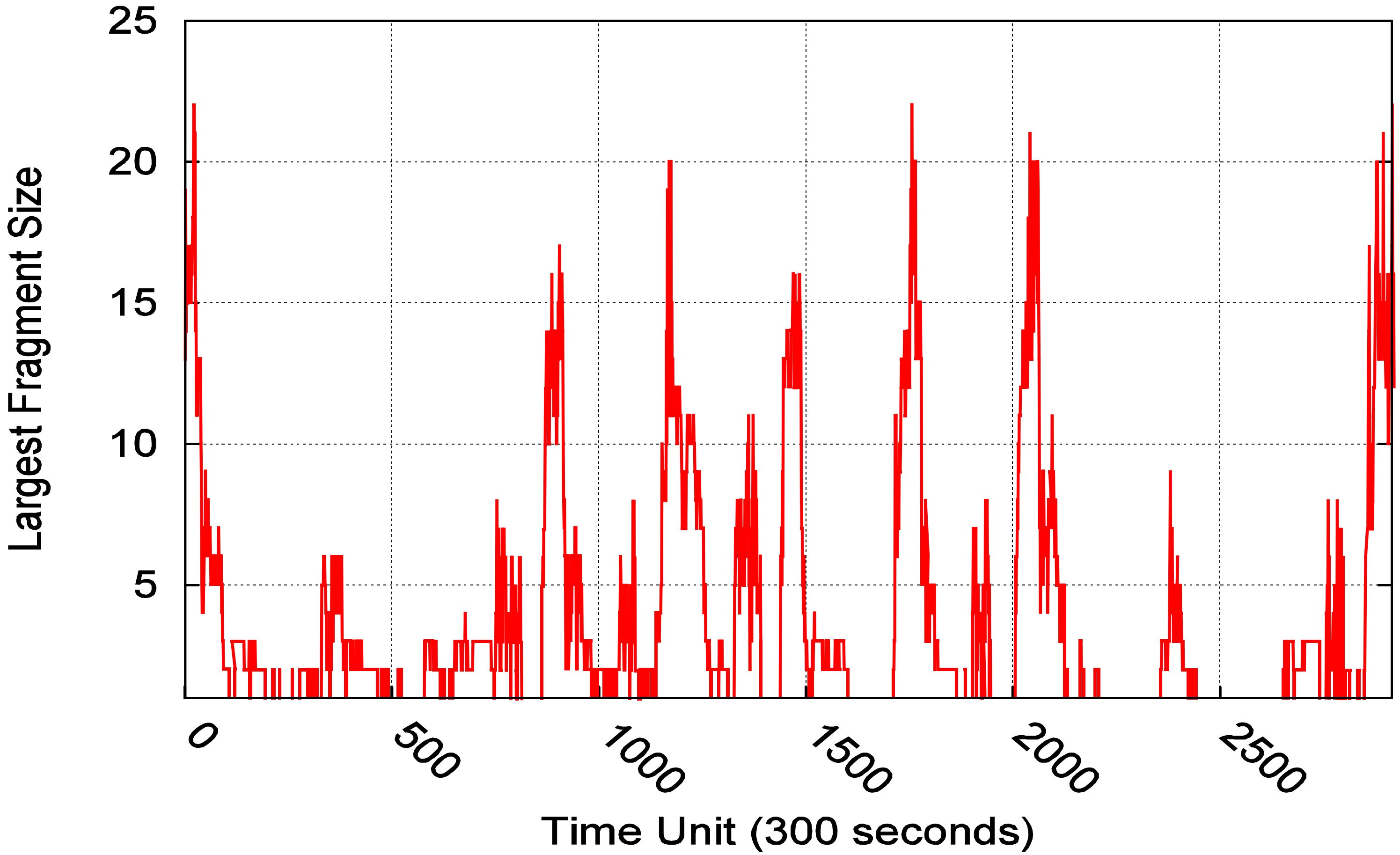
Analyse Network Structure and Model

- Network structure of social systems to model dynamics
- Parameterise with interaction patterns, modularity, and details of time-dependent activity
  - Weighted networks
  - Modularity
  - Centrality (e.g. Degree)
  - Community evolution
  - Network measurement metrics
  - Patterns of interactions

Publications at:
http://www.haggleproject.org
http://www.social-nets.eu/

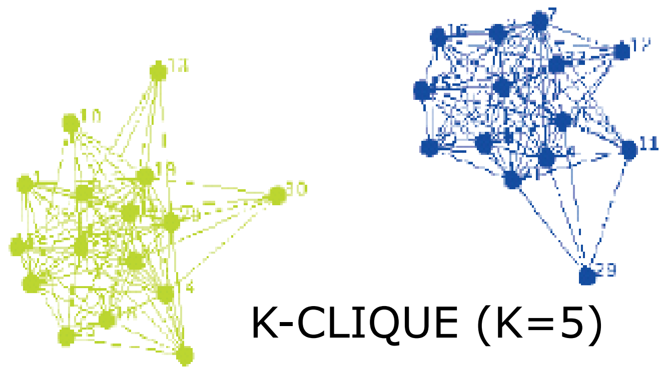
Regularity of Network Activity

- Cambridge Data (11 days by undergraduate students in Cambridge): Size of largest fragment shows network dynamics

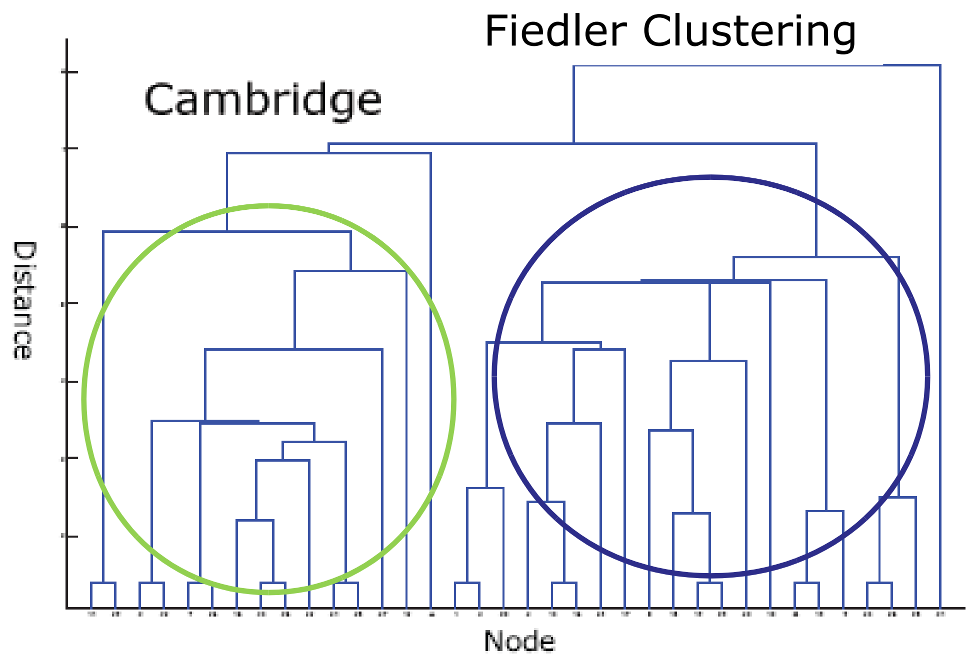


Uncovering Community

- Contact trace in form of weighted (multi) graphs
  - Contact Frequency and Duration
- Use community detection algorithms from complex network studies
  - K-clique, Weighted network analysis, Betweenness, Modularity, Fiedler Clustering etc.



K-CLIQUE (K=5)



Simulation of Disease – SEIR Model

Four states on each node:

SUSCEPTIBLE→**E**XPOSED→**I**NFECTED→**R**ECOVERD

Parameters

p: exposure probability

a: exposed time (incubation period)

t: infected time

Diseases

D1 (SARS): p=0.8, a=24H, t=30H

D2 (FLU): p=0.4, a=48H, t=60H

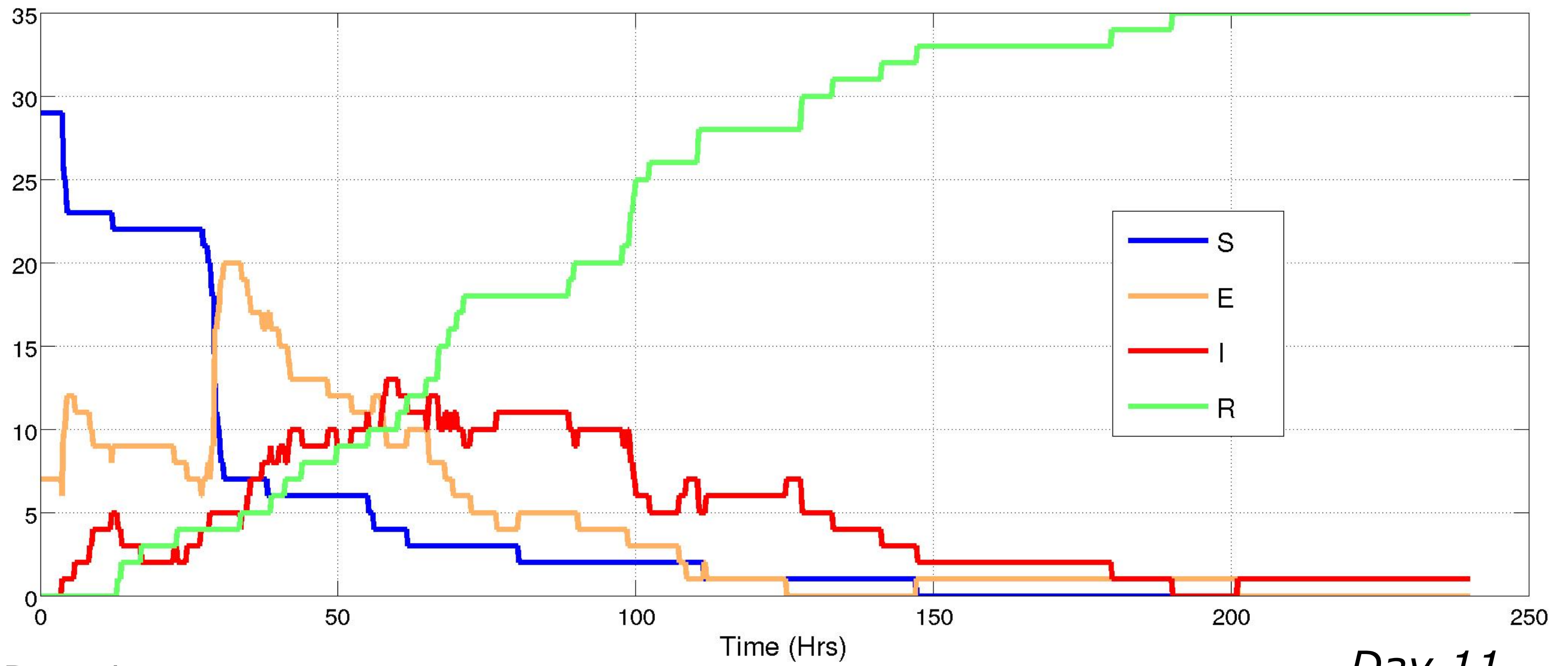
D3 (COLD): p=0.2, a=72H, t=120H

Seed nodes

Random selection of 20% of nodes (=7) among 36 nodes

SARS

- Exposure probability = 0.8
- Exposed time = 24H (average)
- Infected time =30H (average)

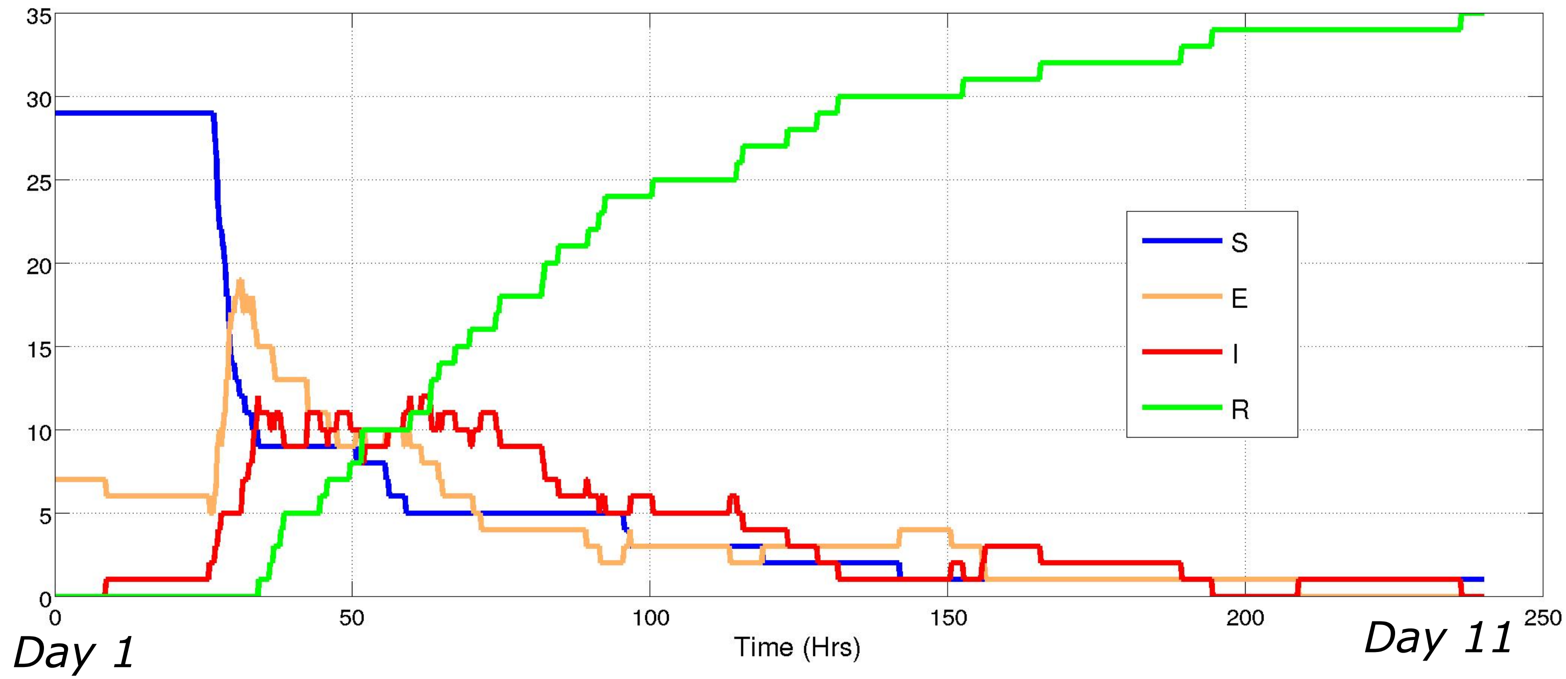


Day 1

Day 11

Flu

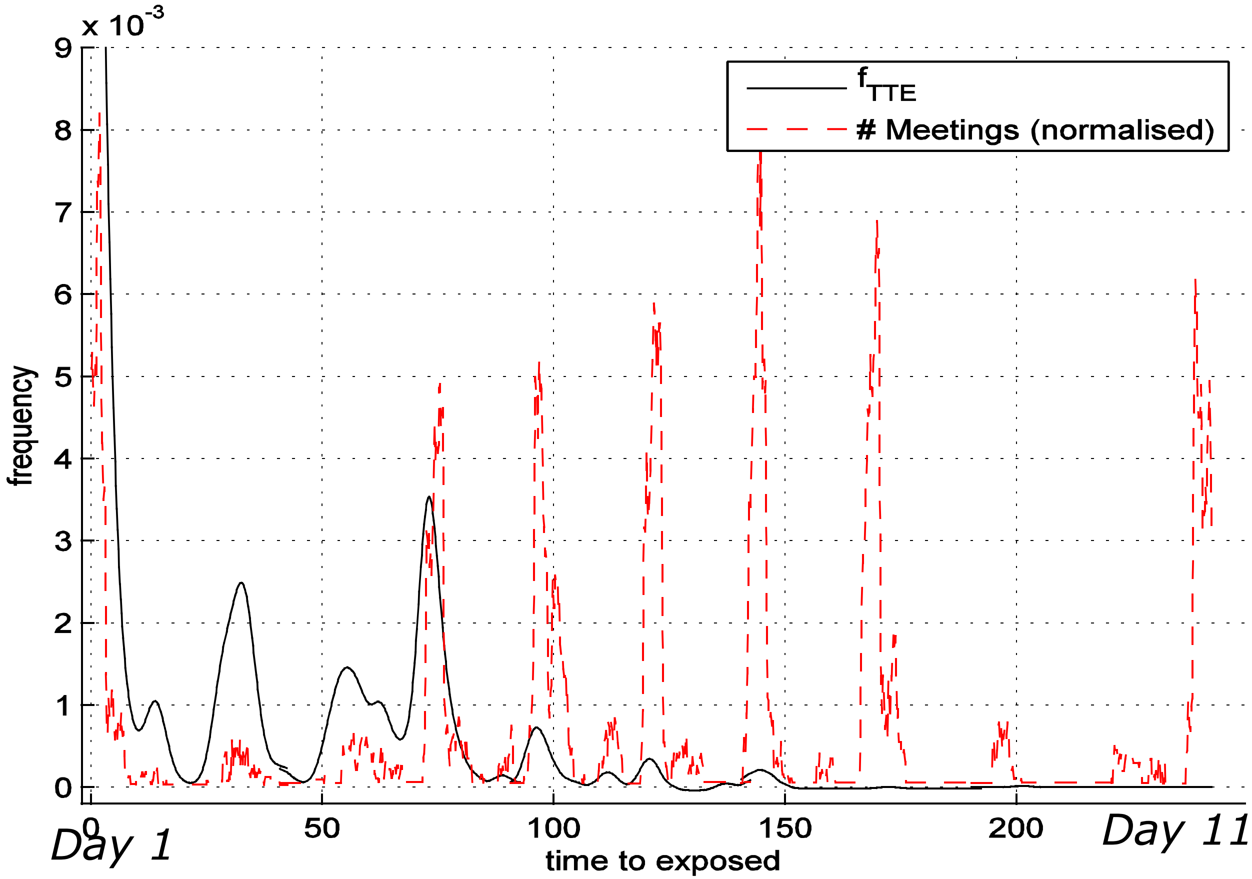
- Exposure probability = 0.4
- Exposed time = 48H (average)
- Infected time = 60H (average)



UNIVERSITY OF CAMBRIDGE Computer Laboratory

Time to Exposure vs #of Meetings

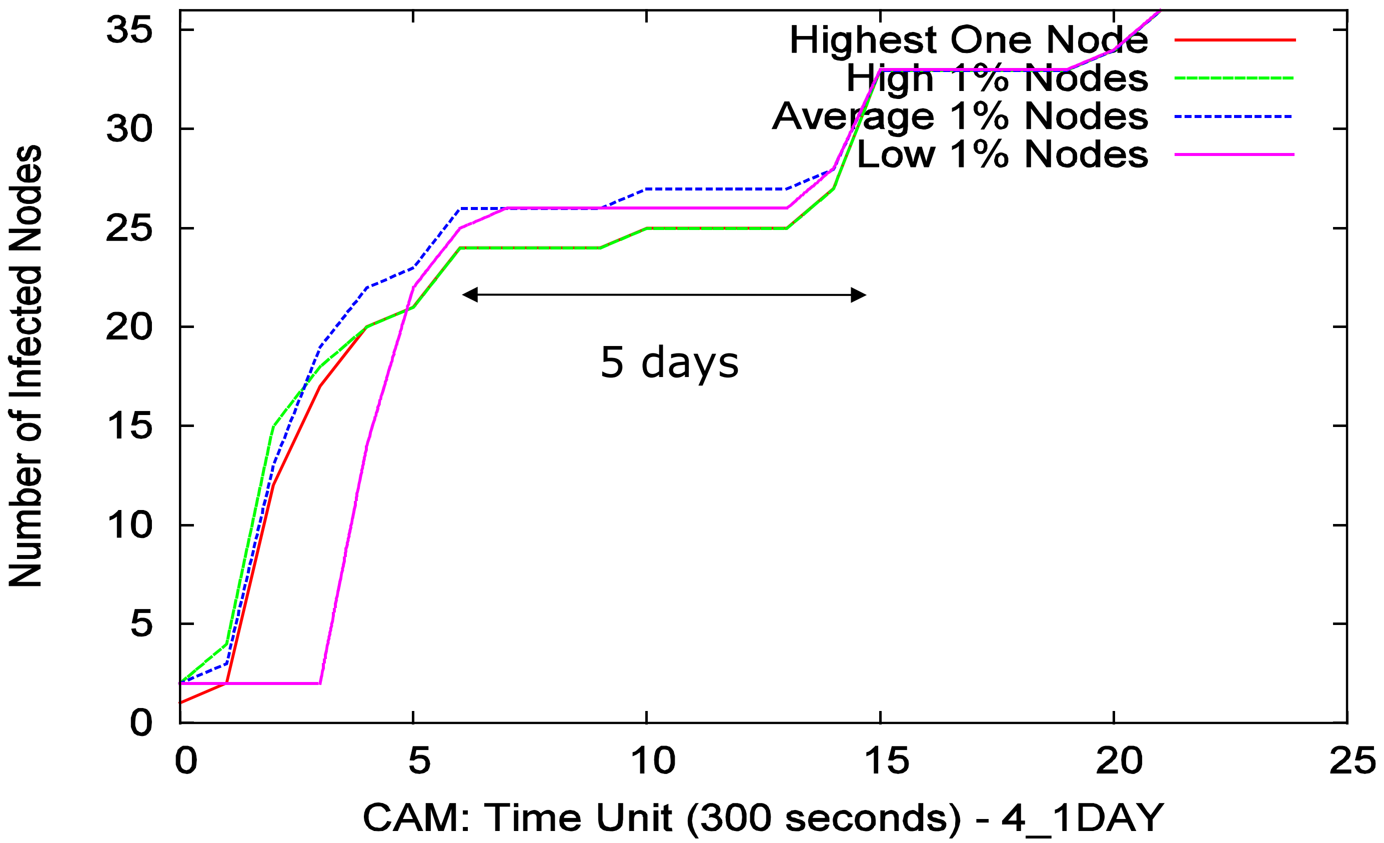
- Distribution of time to infection (black line) is strongly influenced by the time dependent adjacency matrices of meetings



Day 1
Day 11

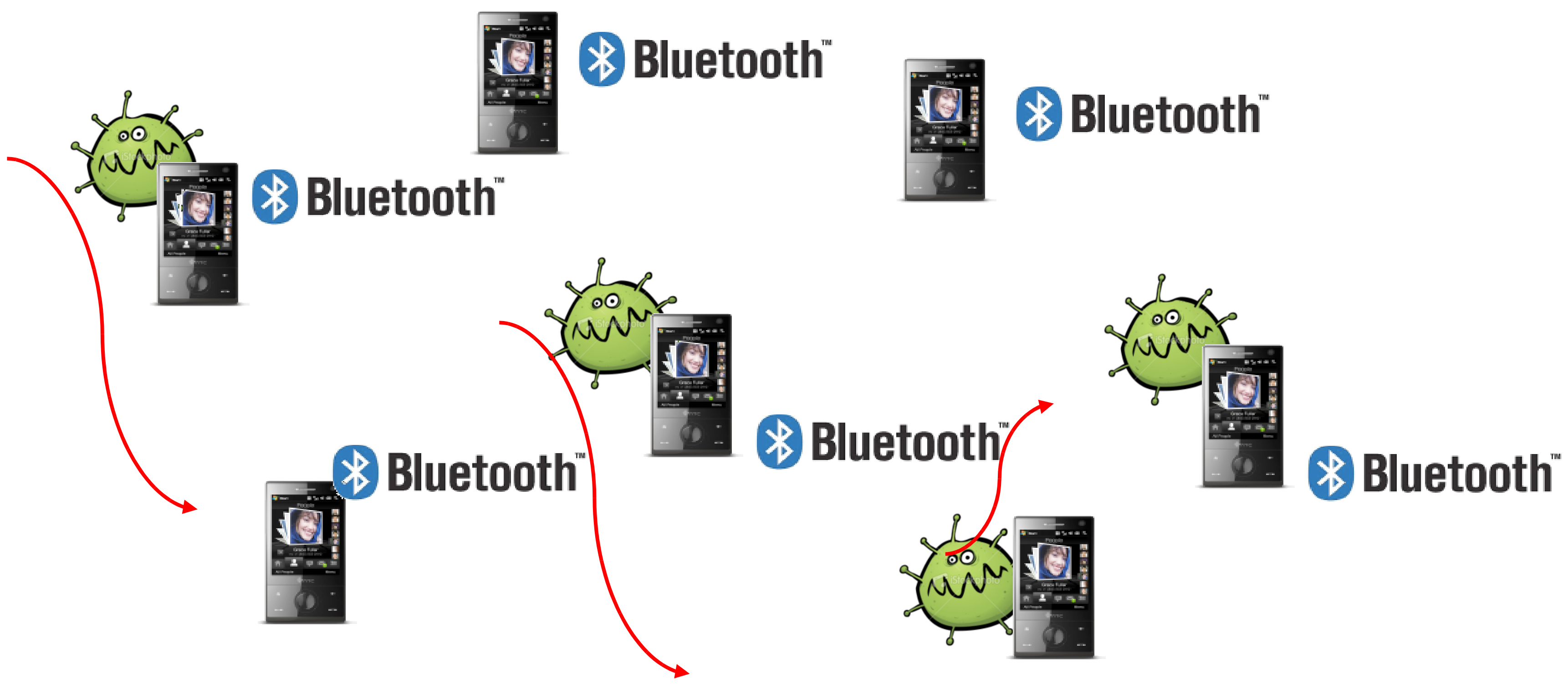
D0: Simple Epidemic (3 Stages)

- First Rapid Increase: Propagation within Cluster
- Second Slow Climbing
- Reach Upper Limit of Infection



Virtual Disease Experiment

- Spread virtual disease via Blutooth communication in proximity radio range
- Integrate SAR, FLU, and COLD in SIER model
- Provide additional information (e.g. Infection status, news) to observe behavioural change



The FluPhone Project

http://www.cl.cam.ac.uk/research/srg/netos/fluphone/

https://www.fluphone.org

Email: flu-phone@cl.cam.ac.uk

FluPhone | UNIVERSITY OF CAMBRIDGE | UNIVERSITY OF LIVERPOOL | E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL | MRC Medical Research Council

Main page | Information | Help | Contact us

Email: Password: Log in
Forgotten your password?

FluPhone Study

This is the home page for the FluPhone study. A study to measure social encounters made between people, using their mobile phones, to better understand how infectious diseases, like 'flu, can spread between people.

This study will record how often different people (who may not know each other) come close to one another, as part of their everyday lives. To do this, we will ask volunteers to install a small piece of software (called FluPhone) on their mobile phones and to carry their phones with them during their normal day-to-day activities. The software will look for other nearby phones periodically using Bluetooth, record this information and send it back to the research team via the cellular phone data service. This information will give us a much better understanding of how often people congregate into small groups or crowds, such as when commuting or through work or leisure activities. Also, by knowing which phones come close to one another, we will be able to work out how far apart people actually are, and how fast diseases could spread within communities. We are also asking participants to inform us of any influenza-like symptoms they may experience during the study period, so that we can match the spread of 'flu to the underlying social network of encounters made.

If you wish to take part in this study, please read the study information below, and then click **here** to start the registration process.

For more detailed information about the study, please visit the study information page.

News:

- The pilot study within the university will start on the April 1st, 2010
- The webpage is up!

ErdOS

Enabling opportunistic resources sharing in mobile Operating Systems

Narseo Vallina-Rodríguez

Jon Crowcroft

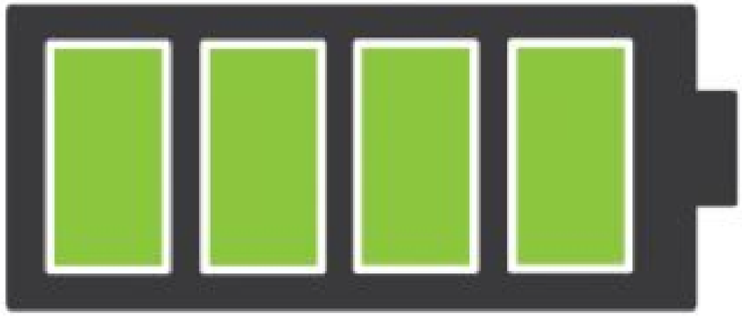
University of Cambridge

MUM 2010, Cyprus

Motivation


WiFi
Bluetooth
GSM/GPRS/3G
Camera
Accelerometer
GPS
CPU (1 GHz)
Storage (>2 GB)
11:03 PM
Google
Maps
Gallery
YouTube
Phone
Gmail
Browser
Helping Grandpa Get His Tech On
New York Times
61°

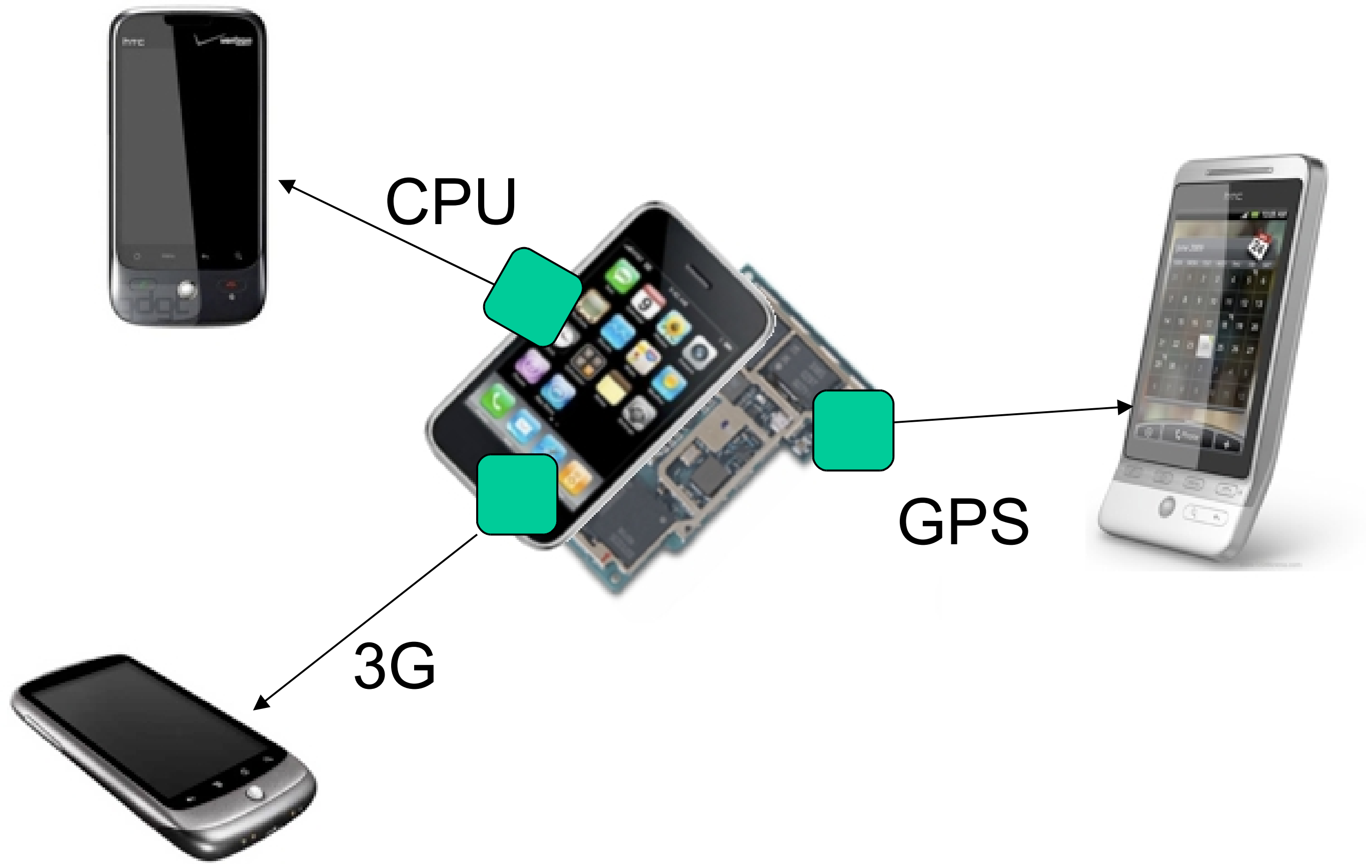
Motivation



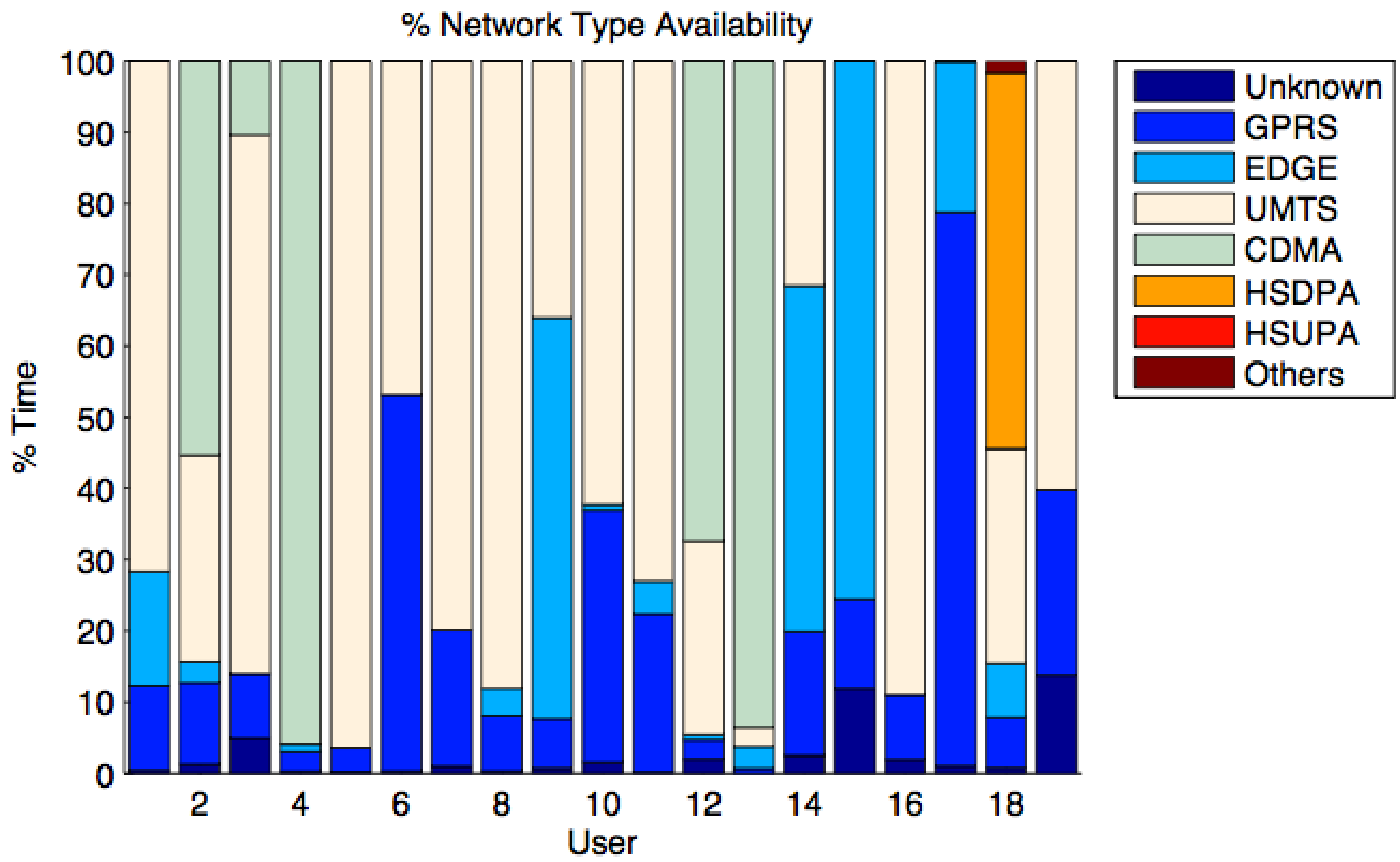
"Energy is still the main limitation in mobile systems"

Motivation

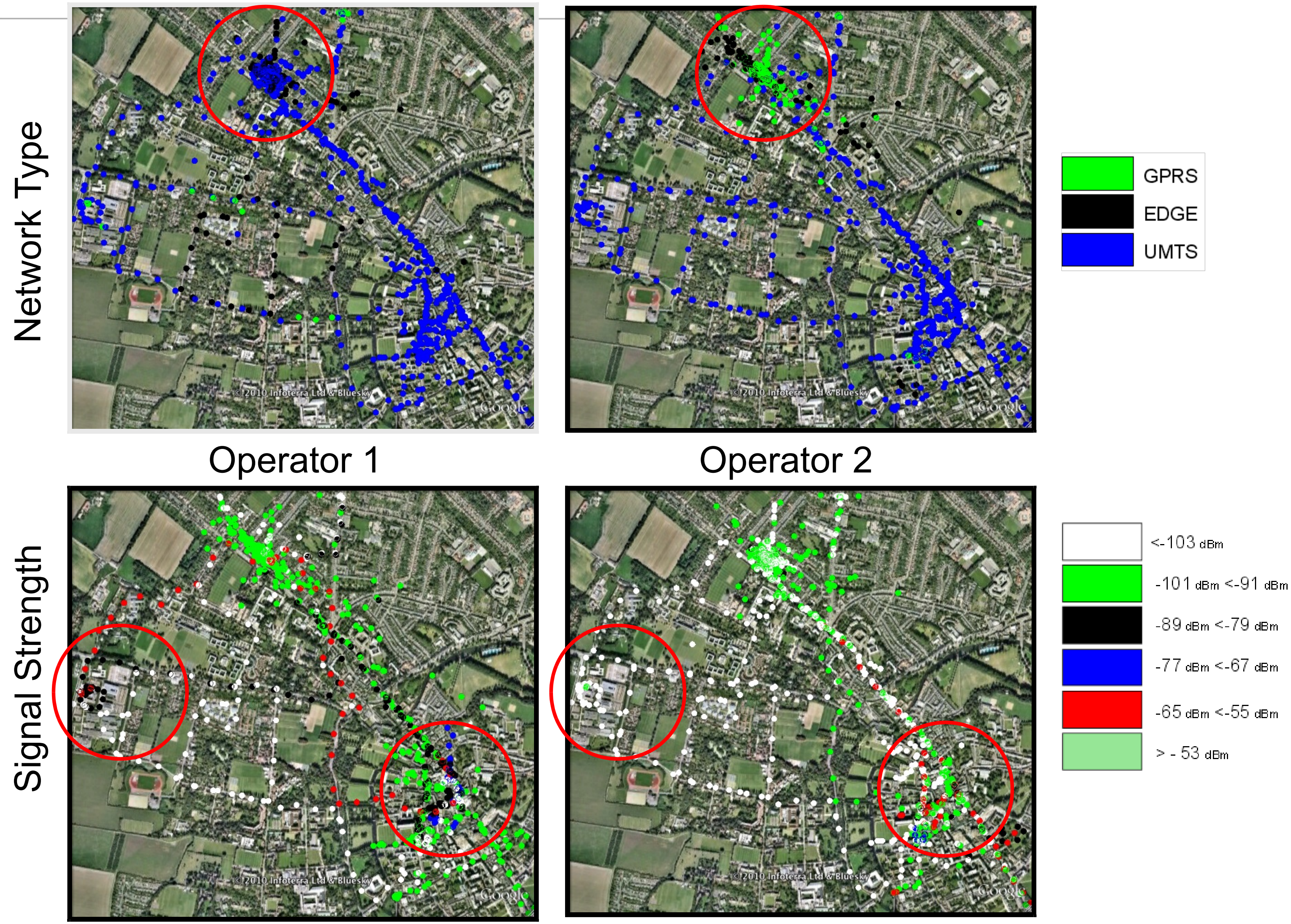
Motivation



Motivation


% Network Type Availability
% Time
100
90
80
70
60
50
40
30
20
10
0
2
4
6
8
10
12
14
16
18
User
Unknown
GPRS
EDGE
UMTS
CDMA
HSDPA
HSUPA
Others

Motivation


Network Type
Signal Strength
Operator 1
Operator 2
GPRS
EDGE
UMTS
<-103 dBm
-101 dBm <-91 dBm
-89 dBm <-79 dBm
-77 dBm <-67 dBm
-65 dBm <-55 dBm
> - 53 dBm
©2010 Infoterra Ltd & Bluesky
Google

Motivation

Why not sharing mobile resources opportunistically with other users?

II. ErdOS

Social energy-aware OS

Access co-located resources opportunistically

Customised proactive resources management

Social connections provide access control

Dataset Description

18 Android OS users

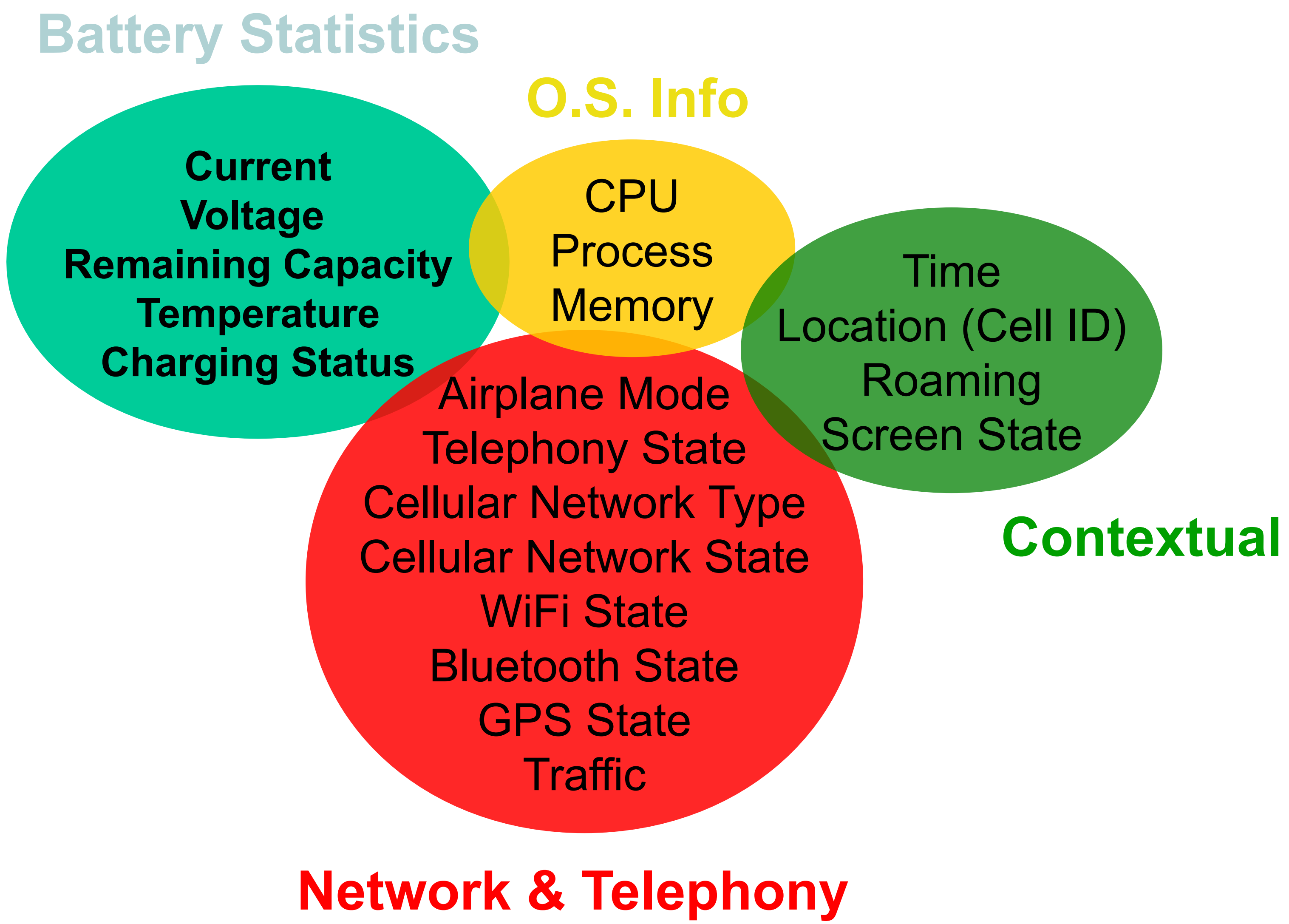
1-2 weeks

Resources Tracker

"Exhausting battery statistics". Mobiheld 2010

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Dataset Description

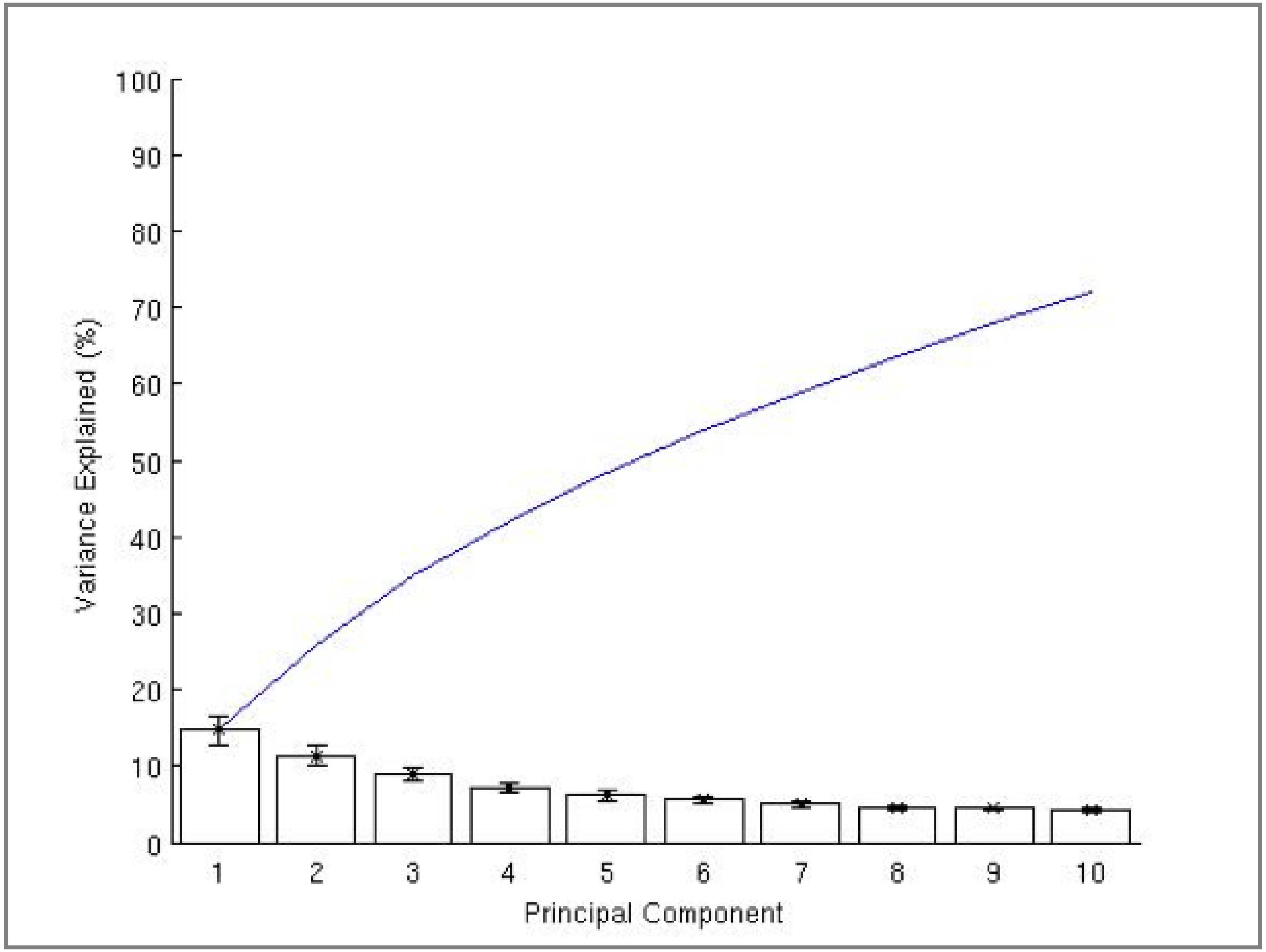


Usage Analysis Tools

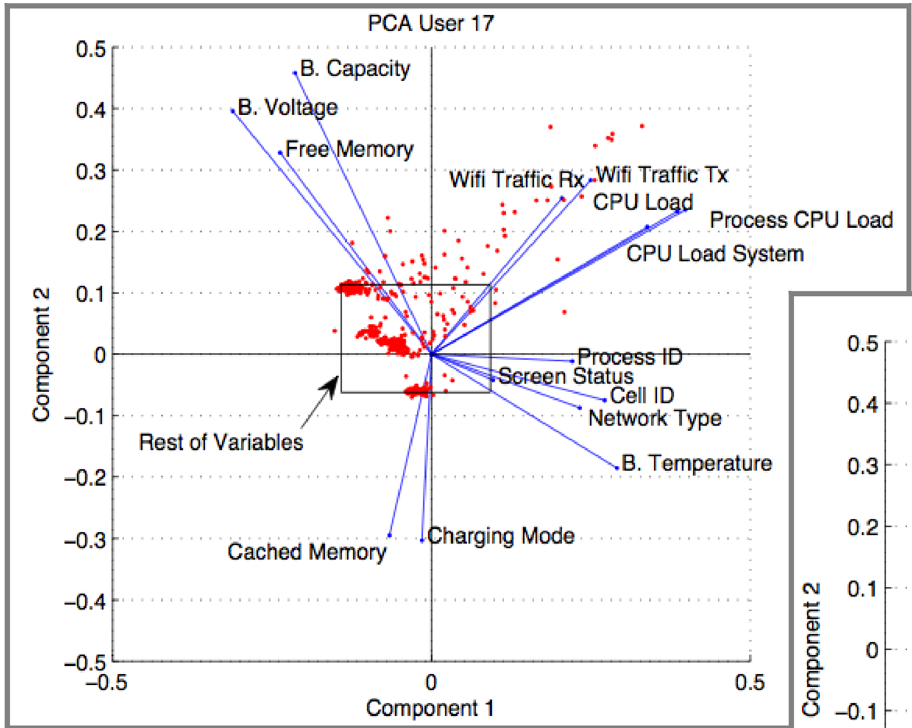
Principal Component Analisys (PCA):

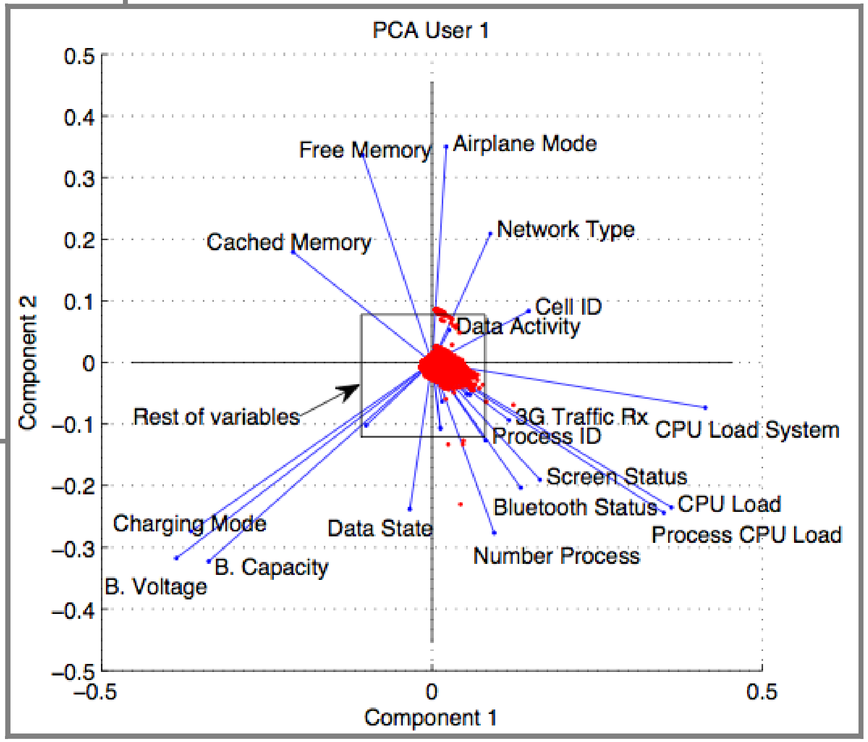
Transforms a number of possibly correlated variables into a smaller number of uncorrelated ones called Principal Components

Principal Component Analysis



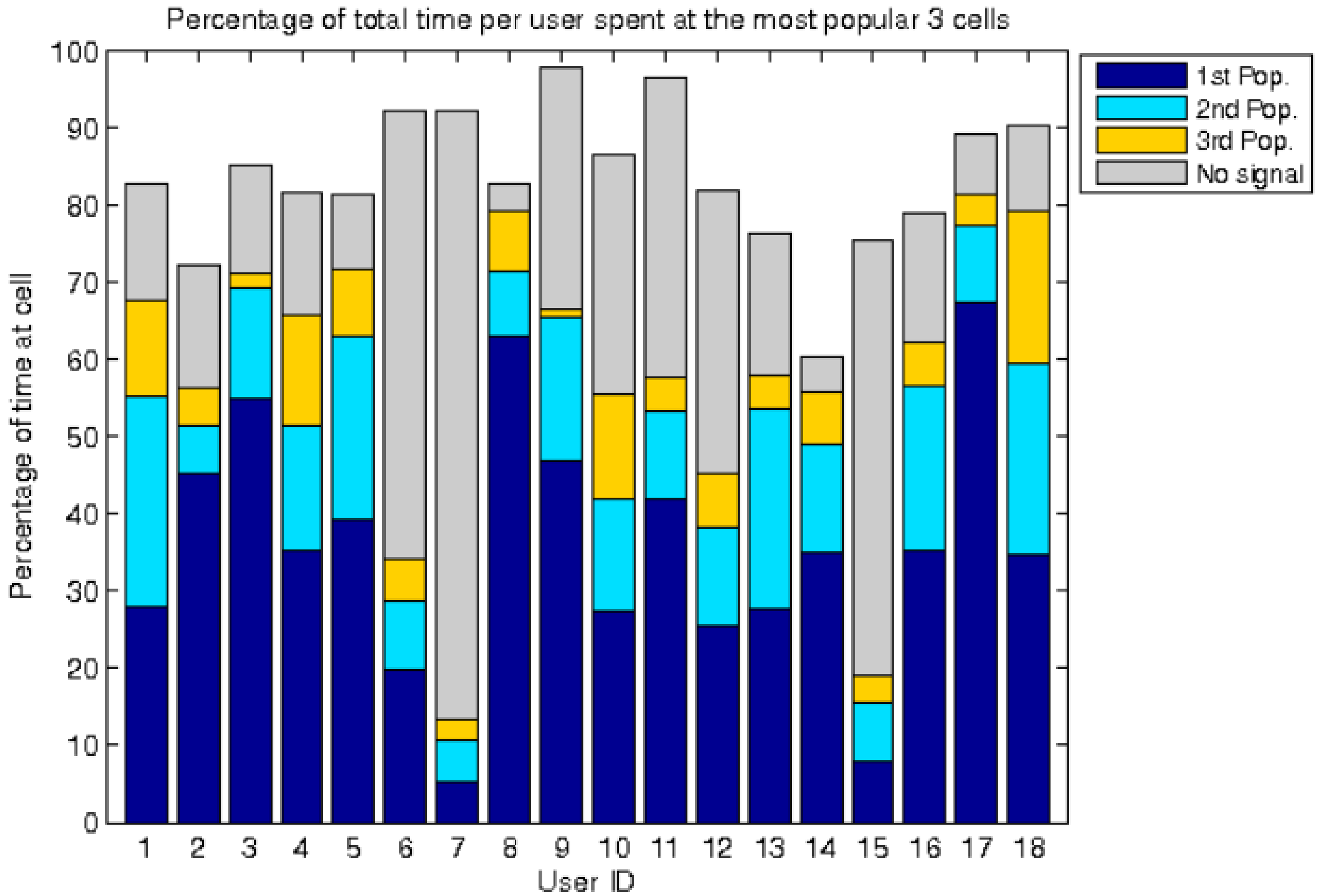
Principal Component Analysis



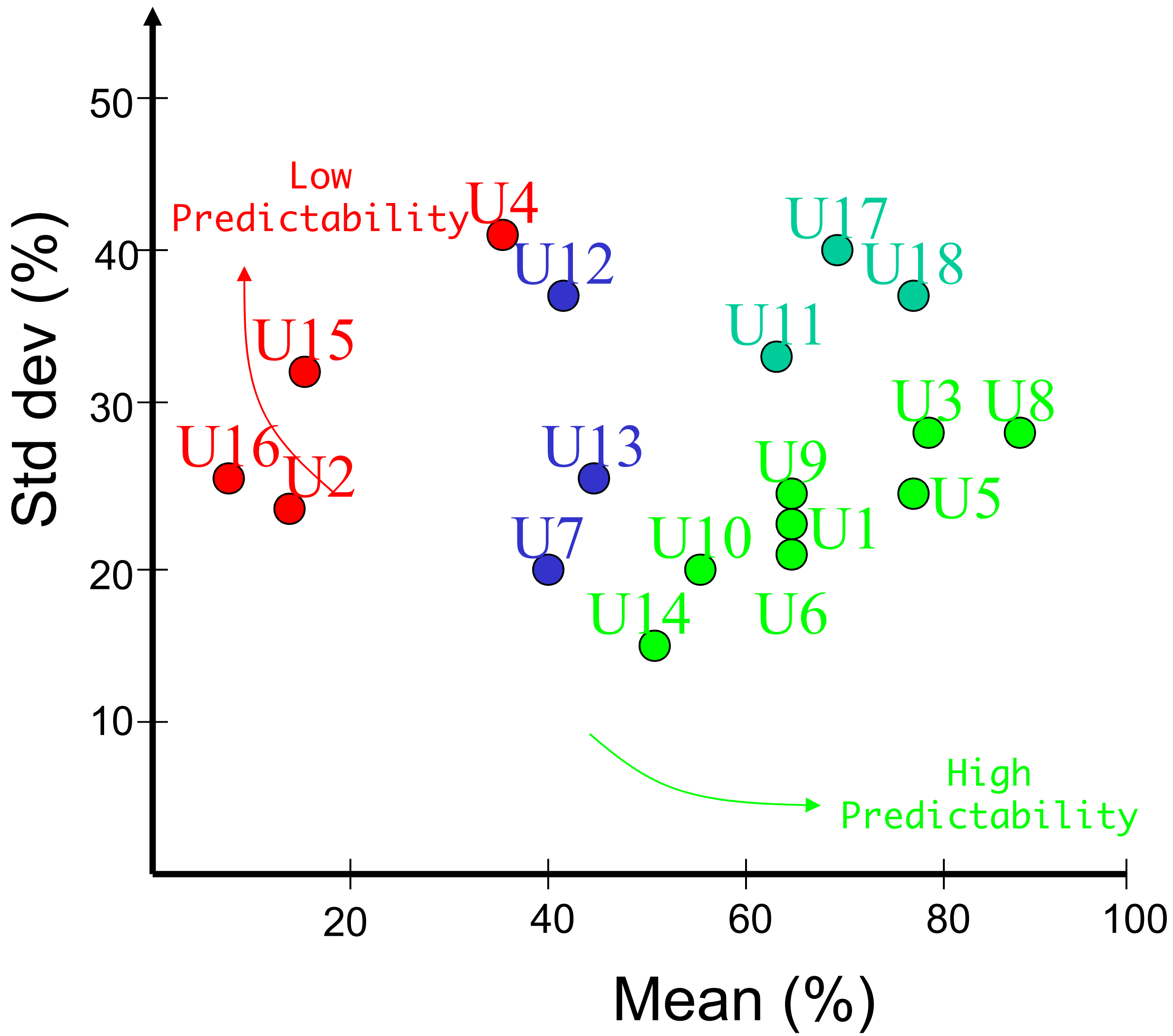


UNIVERSITY OF CAMBRIDGE
Computer Laboratory

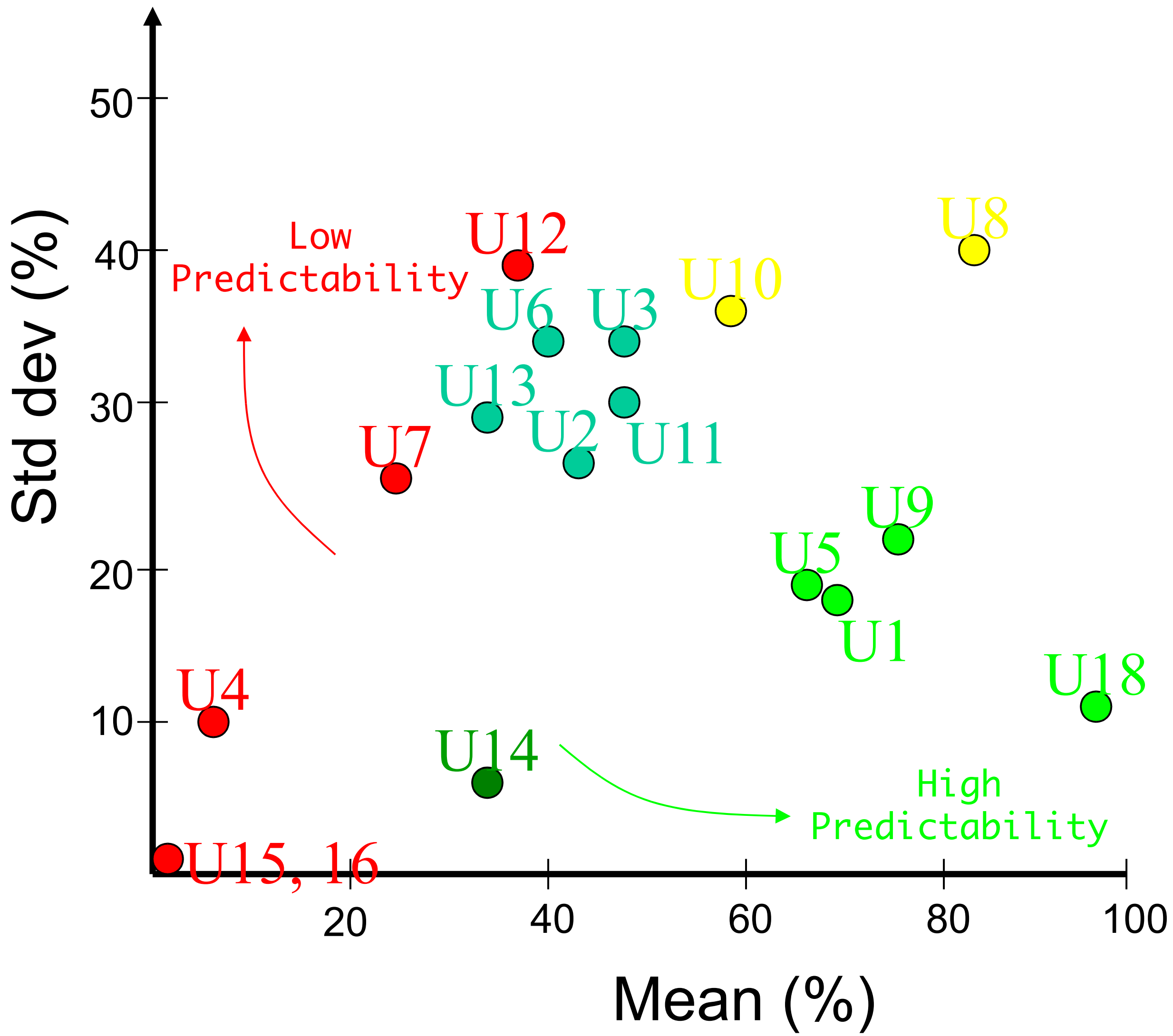
Context importance



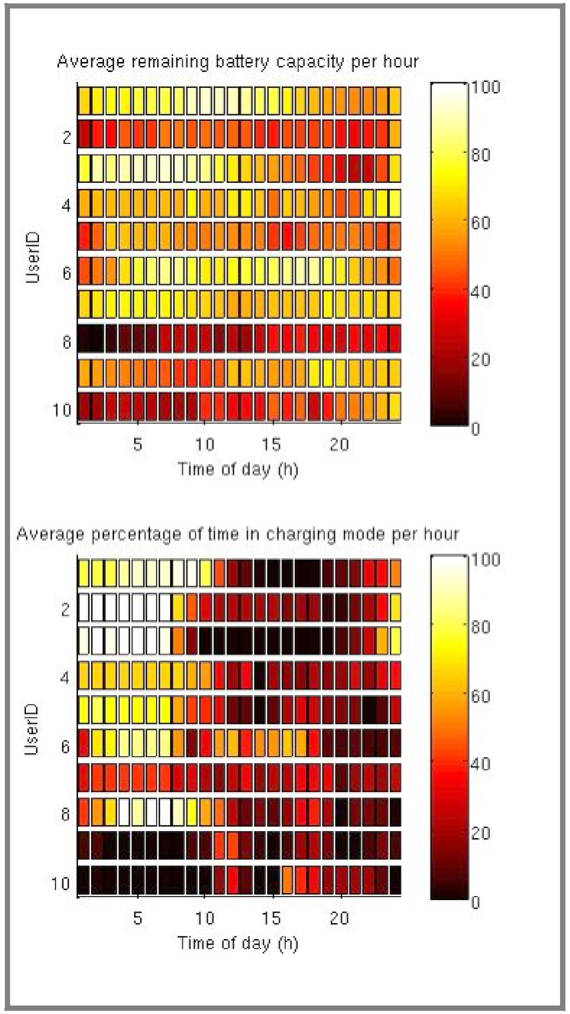
Spatial context: Screen usage

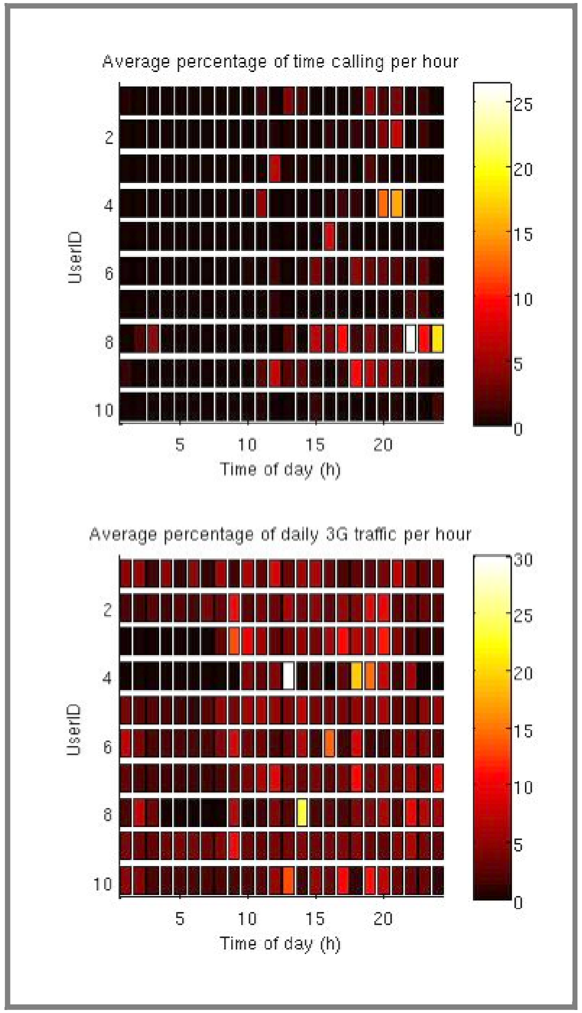


Spatial context: Cellular traffic

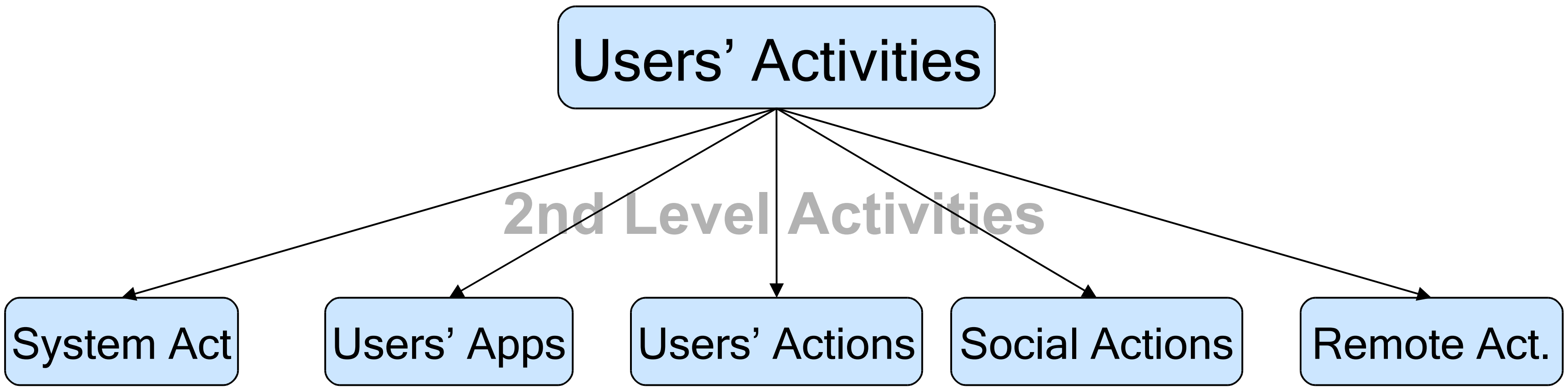


Temporal context: Daily usage

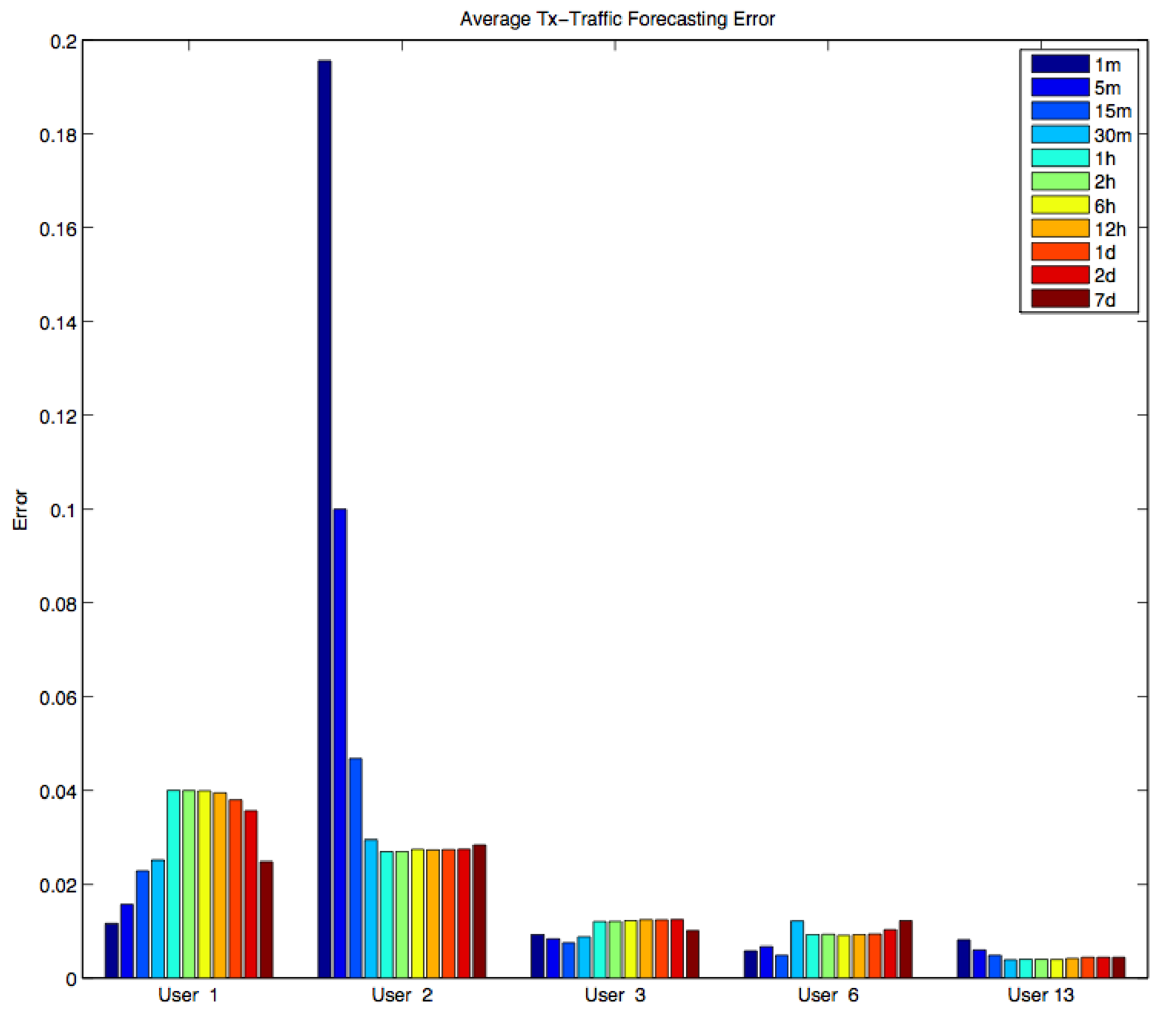




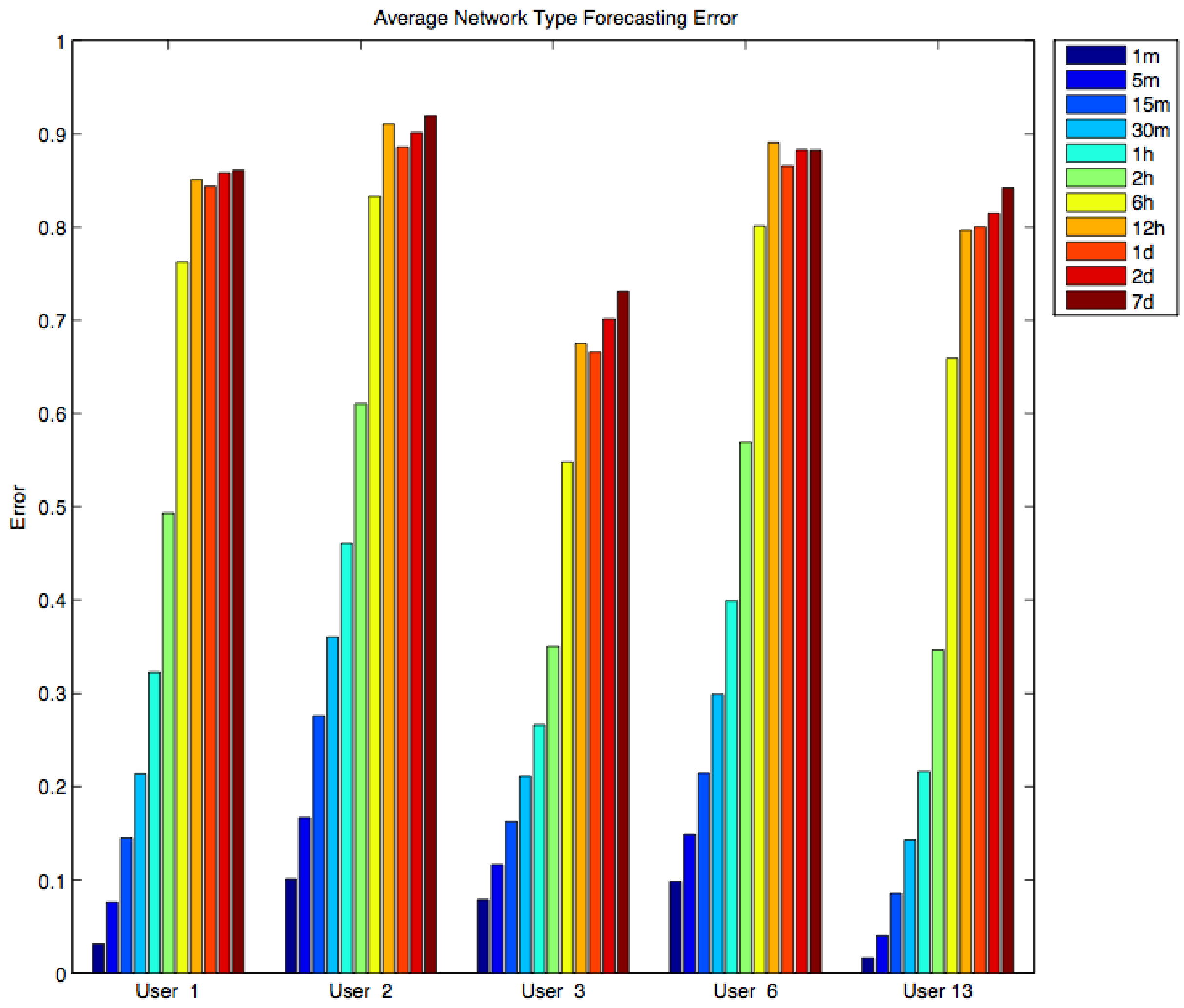
Resources Allocations: *Activities*



Forecasting Resources Demands



Forecasting Resources State

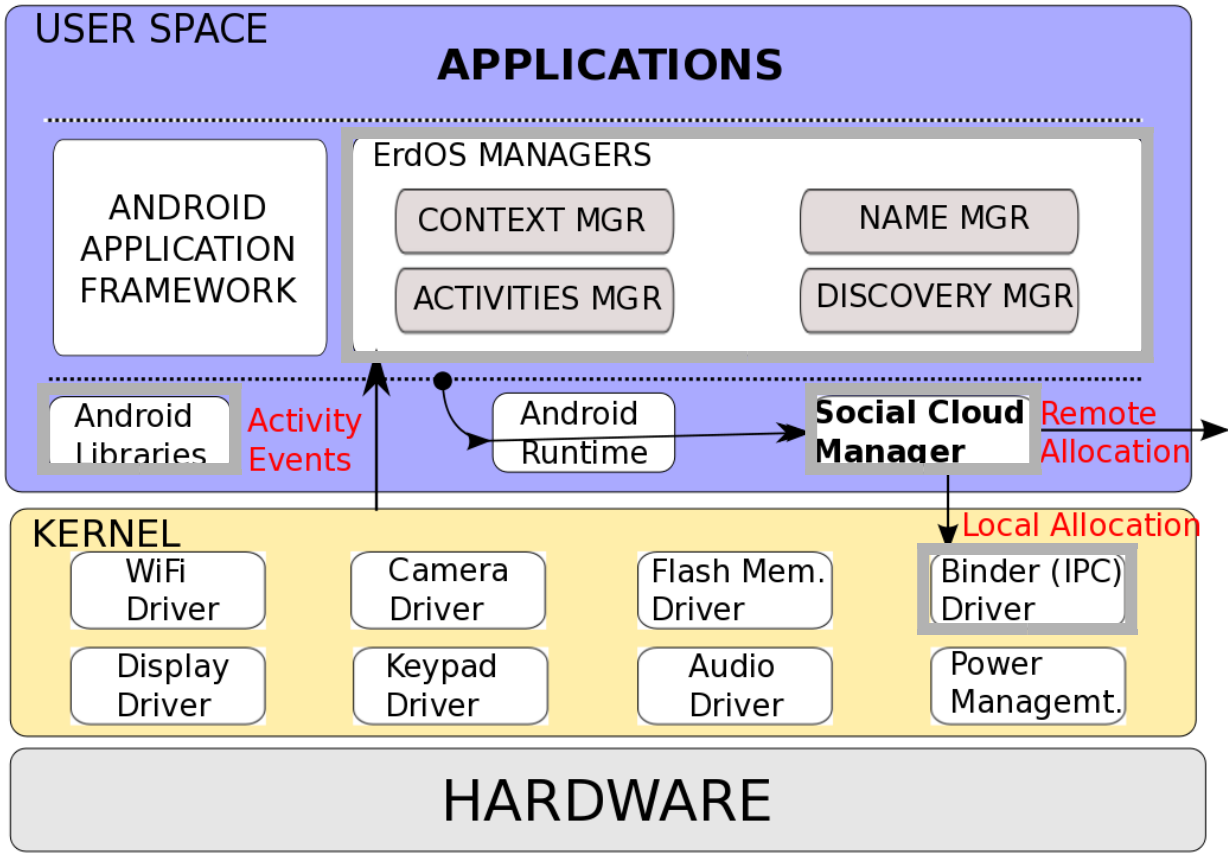


Access Control

Social links facilitate access control and security

- Unix-like permissions are made automatically based on users' social networks
- Proximity reduces privacy and security issues
- OSNs can help to exchange public keys

Architecture



Related work

- Resource allocation and energy-aware OS
  - **ECOSystem**. Zeng et al. ACM ASPLOS, 2002
  - **Quanto**. Stoica et al. USENIX 2008
  - **CinderOS**. Rumble et al. MOBIHELD 2009

- Mobile usage and energy demand
  - Falaki et al. ACM Mobisys 2010
  - Oliver, ACM HotPlanet 2010
  - Balasubramanian et al. ACM IMC 2010
  - Rice et al. ACM PerCOM 2010

Conclusions

- Energy is a primary target for optimization in mobile handsets
  - Benefits in QoS and energy savings by accessing resources opportunistically
  - Social links can be used for access control policies
- Applications and users' behavior generate complex dynamics and interdependencies among resources
  - Energy allocation and resources control must be customized to each user and handset
  - Pro-active resources management aided by contextual information

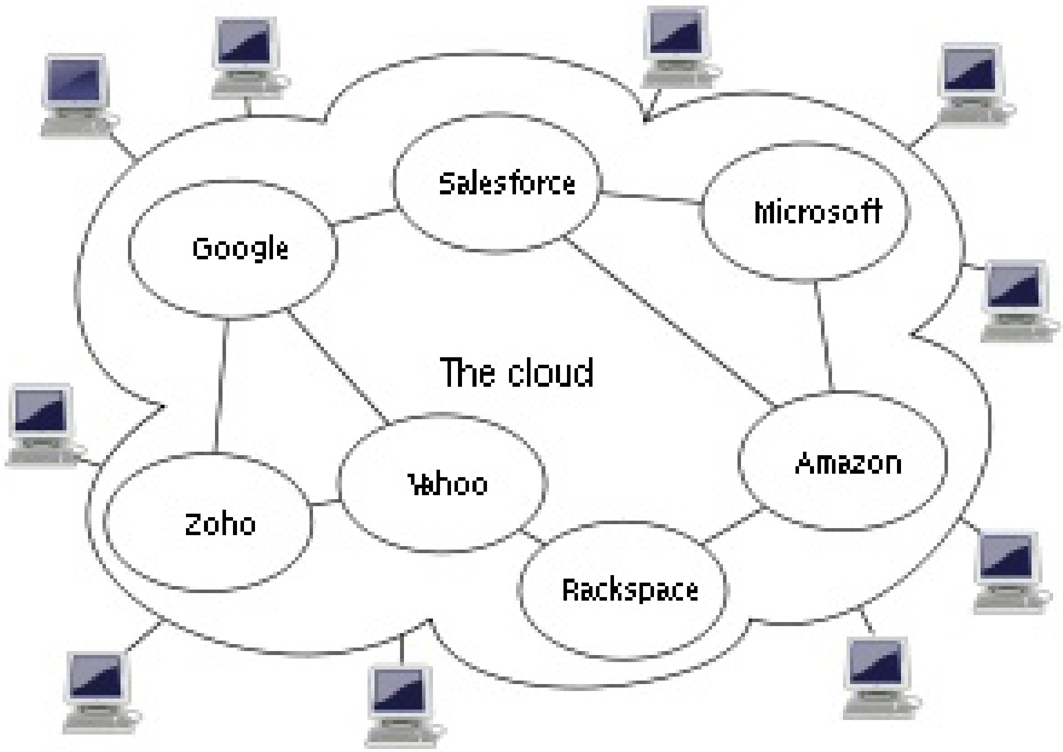
Future Work

- Finishing implementation as an Android OS extension
  - Performance/Scalability evaluation
- Demonstrate benefits of sharing different resources (Cellular Nets, GPS, CPU)
- Resources Discovery Protocols
- Research on lighter forecasting techniques
  - Cloud Computing?
- Security evaluation
- Incentive schemes?

III
Droplets:- Condensing the Cloud

From the Cloud…

The cloud has its risks…



Centralisation of PII

What if

- Provider goes broke
- Lose all your family photos
- Assets sold to another (unknown) provider

In a large organisation,

- There will always be someone bad
- Who can datamine
- Identity theft
- And worse

To Droplets

At the other extreme…

A fully decentralised approach…

Can obviate cloud risks
- But introduces complexity
- Management overhead (p2p/manet/dtn)
- Availability/resilience
- Total data loss if device stolen

Can we compromise
- Between extreme centralisation…
- And extreme decentralisation?

Via extraction

Firstly, we need to pull/push data

From/to the cloud…

…and condensation

Replication/Decentralisation are necessary, but not sufficient…

Need to encrypt data

- Both in Cloud
- And in Mist

The mist is a collection of droplets -

- small objects with key/capability
- and auditor

What about cloud business models?

- Use privacy preserving advertising (MPI)
- Use k-anonymity and threshold security
- Use differential privacy for market research

Implementation Details

Platform	*Google AppEngine*	*VM (e.g., on EC2)*	*Home Computer*	*Mobile Phone*
Storage	moderate	moderate	high	low
Bandwidth	high	high	limited	low
Accessibility	always on	always on	variable	variable
Computation	limited	flexible, plentiful	flexible, limited	limited
Cost	free	expensive	cheap	cheap
Reliability	high	high	medium (failure)	low (loss)

To Conclude…

Use contributed resources are fine
- Home hub, phone, etc

But need to unify with cloud
- Do so at API level

Have *both* decentralised and central

Advantages of both
- Low latency access to home/pocket
- High resilience in cloud
- No loss of privacy if bad cloud/pick pocket

Questions?

Thanks!

Email: jac22@cl.cam.ac.uk

http://www.cl.cam.ac.uk/~nv240/erdos.html