Digital Epidemiology: Understanding Epidemic Spread using Human Contact Networks

Eiko Yoneki
eiko.yoneki@cl.cam.ac.uk
http://www.cl.cam.ac.uk/~ey204

Systems Research Group
University of Cambridge Computer Laboratory

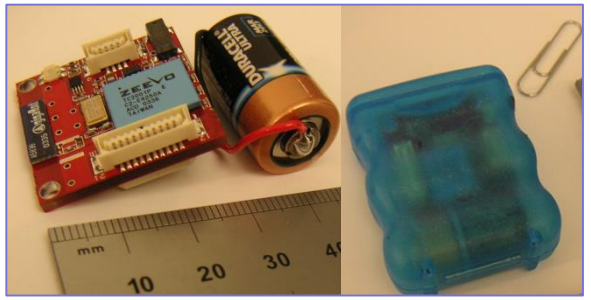
Opportunistic Networks

- Pocket Switched Networks: Devices carried by people, thus ‘do what users do’

- Pocket switched networks involve consumer devices and incorporate social aspects and opportunistic communication

Measure Human Contact Networks

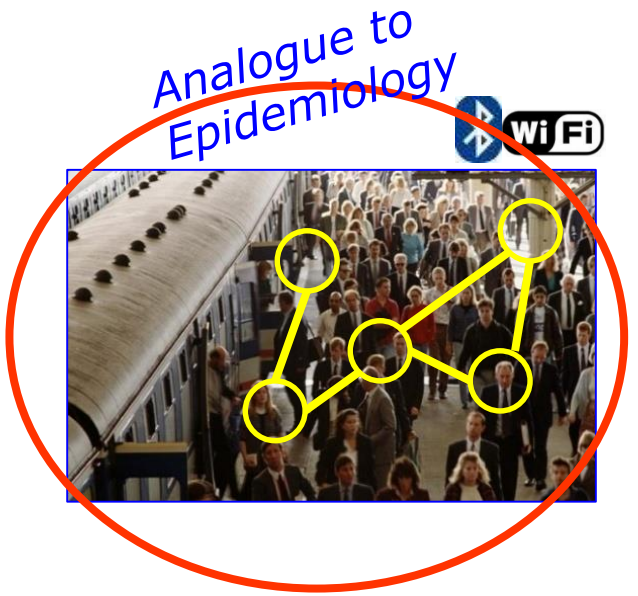
- Sensors
  - Bluetooth Intel iMote
  - 802.15.4 + (magnet, gyroscope)
- RFID Tags
  - UHF Tag Alien ALN-9640 - "Squiggle®" Inlay
  - OpenBeacon active RFID Tag
- Mobile Phones
  - Bluetooth
  - GPS, Google latitude
- GPS Logger
- Online Social Networks
  - Twitter, Facebook, Foursquare…



Spread of Infectious Diseases

- Thread to public health: e.g., 

SARS, Respiratory and other close-contact infectious diseases
- Current understanding of disease spread dynamics
  - Epidemiology: small scale empirical work
- Real-world networks are far more complex
  - Advantage of real world data
  - Emergence of wireless technology for proximity data
  - Post-facto analysis and modelling yield insight into human interactions



Modelling realistic infectious disease spread/prediction

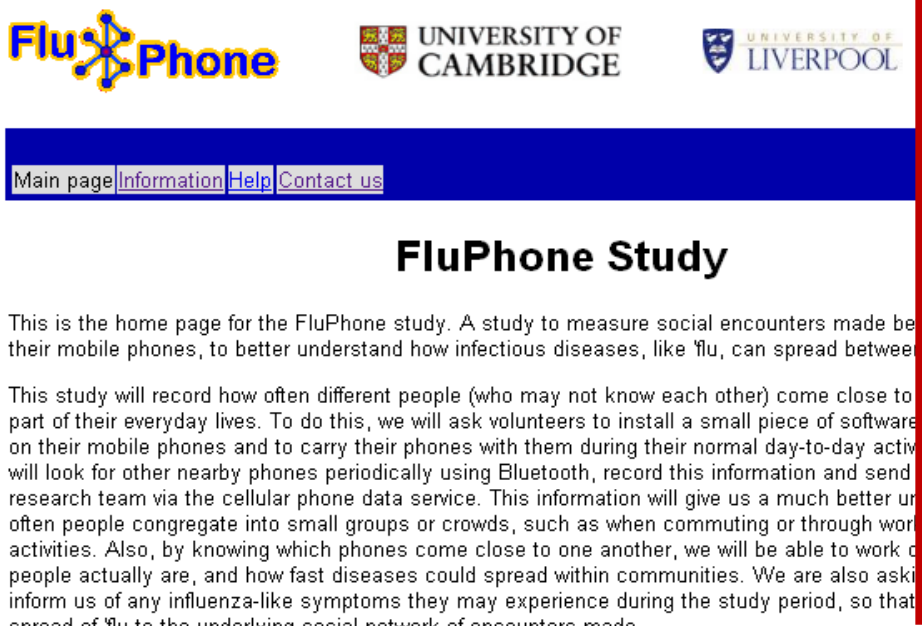
FluPhone Project

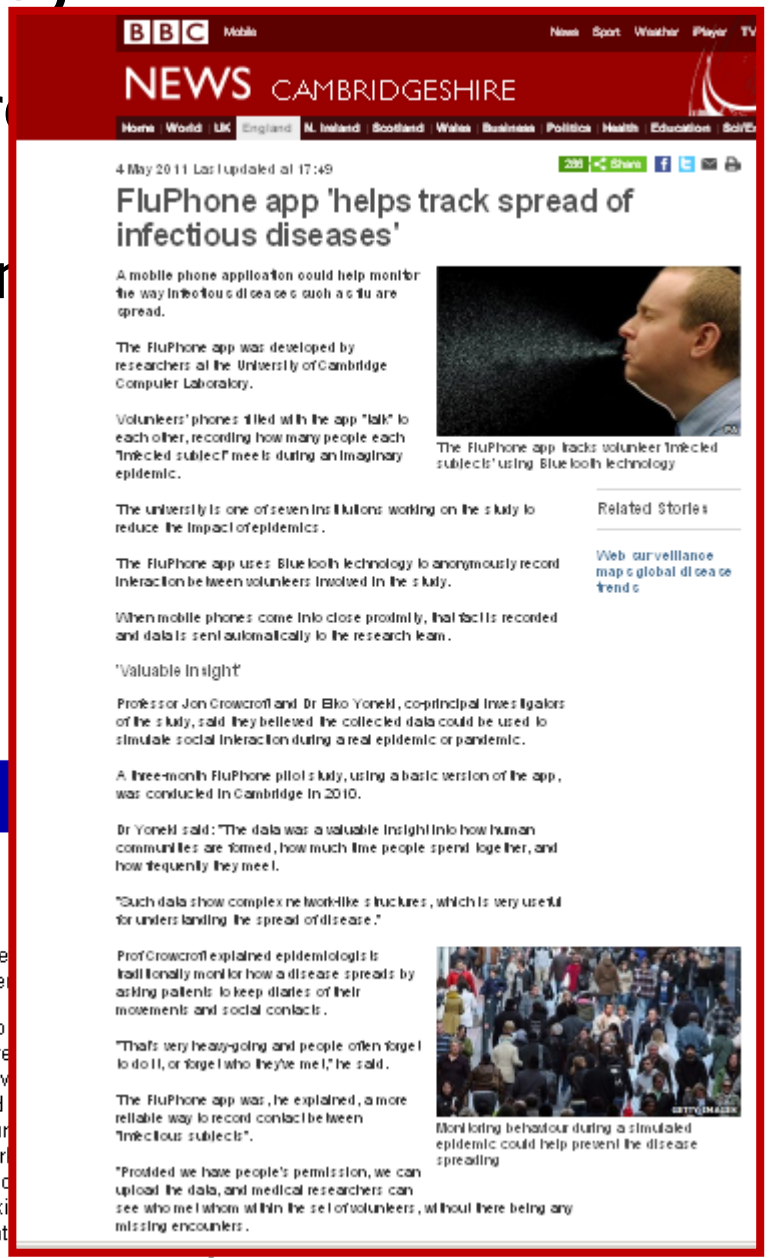
- Scan Bluetooth devices every 2 minutes
- Symptom Survey

FluPhone Project

- Understanding behavioural r
  disease outbreaks
- Proximity data collection usir
  general public in Cambridge

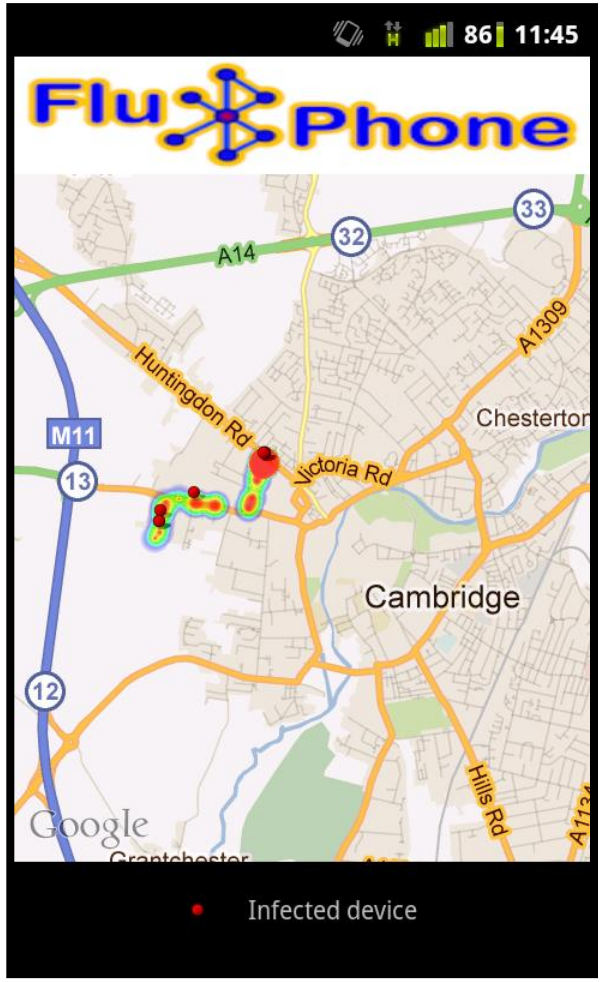
https://www.fluphone




BBC Mobile
NEWS CAMBRIDGESHIRE
4 May 2011 Last updated at 17:49
FluPhone app 'helps track spread of infectious diseases'
A mobile phone application could help monitor the way infectious diseases such as flu are spread.
The FluPhone app was developed by researchers at the University of Cambridge Computer Laboratory.
Volunteers' phones filled with the app "talk" to each other, recording how many people each "infected subject" meets during an imaginary epidemic.
The FluPhone app tracks volunteer 'infected subjects' using Bluetooth technology
The university is one of seven institutions working on the study to reduce the impact of epidemics.
The FluPhone app uses Bluetooth technology to anonymously record interaction between volunteers involved in the study.
When mobile phones come into close proximity, that fact is recorded and data is sent automatically to the research team.
'Valuable insight'
Professor Jon Crowcroft and Dr Eiko Yoneki, co-principal investigators of the study, said they believed the collected data could be used to simulate social interaction during a real epidemic or pandemic.
A three-month FluPhone pilot study, using a basic version of the app, was conducted in Cambridge in 2010.
Dr Yoneki said: "The data was a valuable insight into how human communities are formed, how much time people spend together, and how frequently they meet.
"Such data show complex network-like structures, which is very useful for understanding the spread of disease."
Prof Crowcroft explained epidemiologists traditionally monitor how a disease spreads by asking patients to keep diaries of their movements and social contacts.
"That's very heavy-going and people often forget to do it, or forget who they've met," he said.
The FluPhone app was, he explained, a more reliable way to record contact between "infectious subjects".
"Provided we have people's permission, we can upload the data, and medical researchers can see who met whom within the set of volunteers, without there being any missing encounters.
Related Stories
Web surveillance maps global disease trends
Monitoring behaviour during a simulated epidemic could help prevent the disease spreading

Trajectory of Encountering Sick People

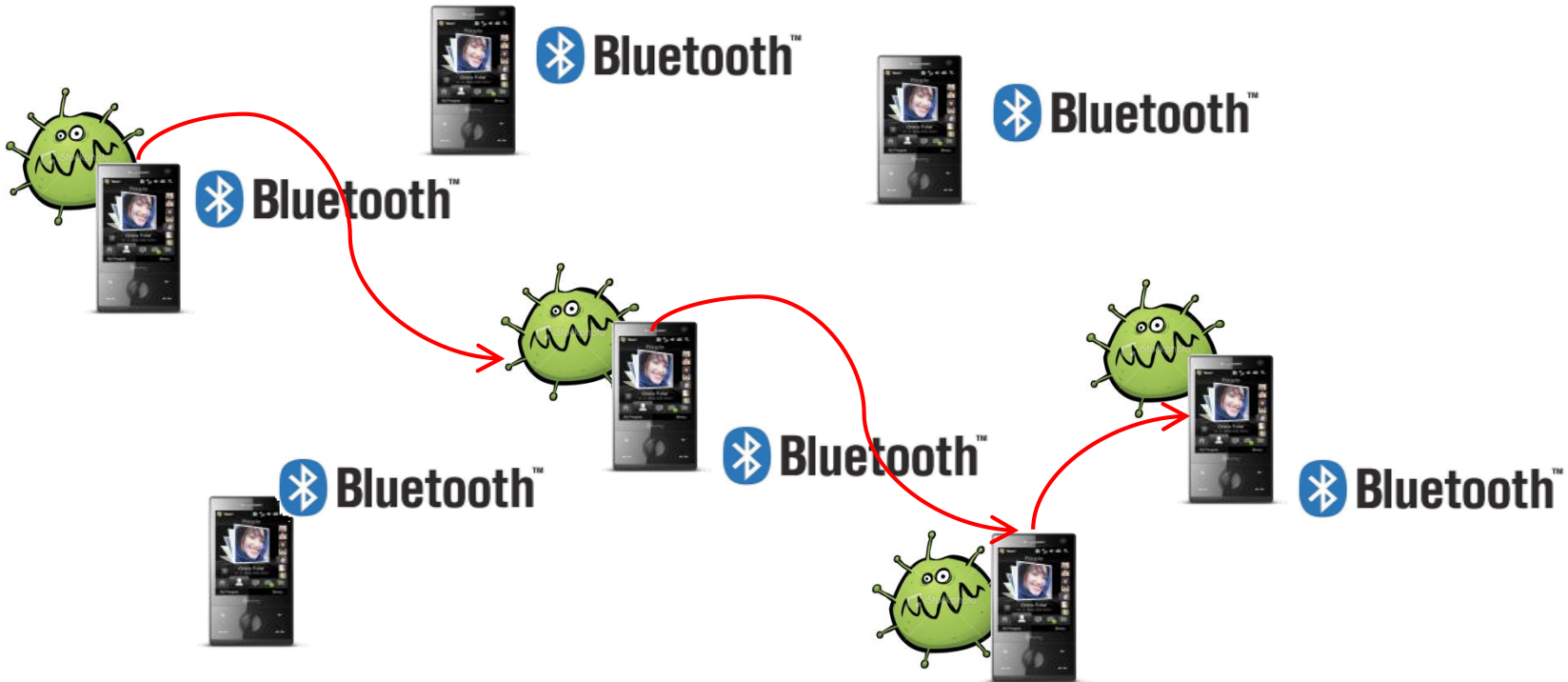
- Integration with GPS equipped Smartphones



Virtual Disease

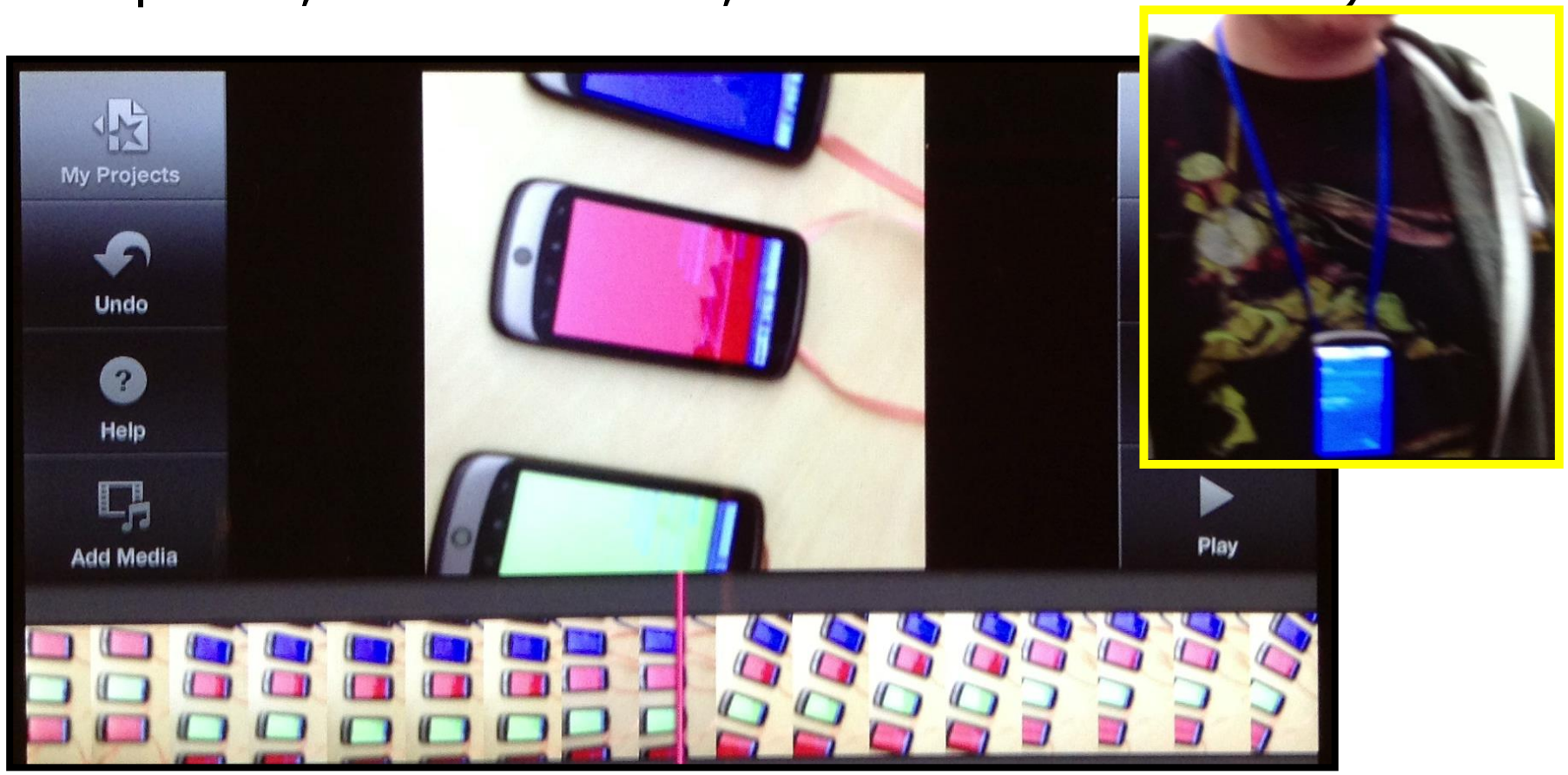
- Spread virtual disease via Bluetooth communication

Disease Name	Exposed Duration	Infectious Duration	Infection Probability
Base line	0	31536000000	1.0
SARS	86400000 **.5H**	108000000 **1H**	0.8
Flu	172800000 **1H**	216000000 **2H**	0.4
Cold	259200000 **2H**	432000000 **3H**	0.2



Virtual Disease: Infection State Coloring

- Spread of Disease (Green: Susceptible, Yellow: Exposed, Red: Infected, and Blue: Recovered)



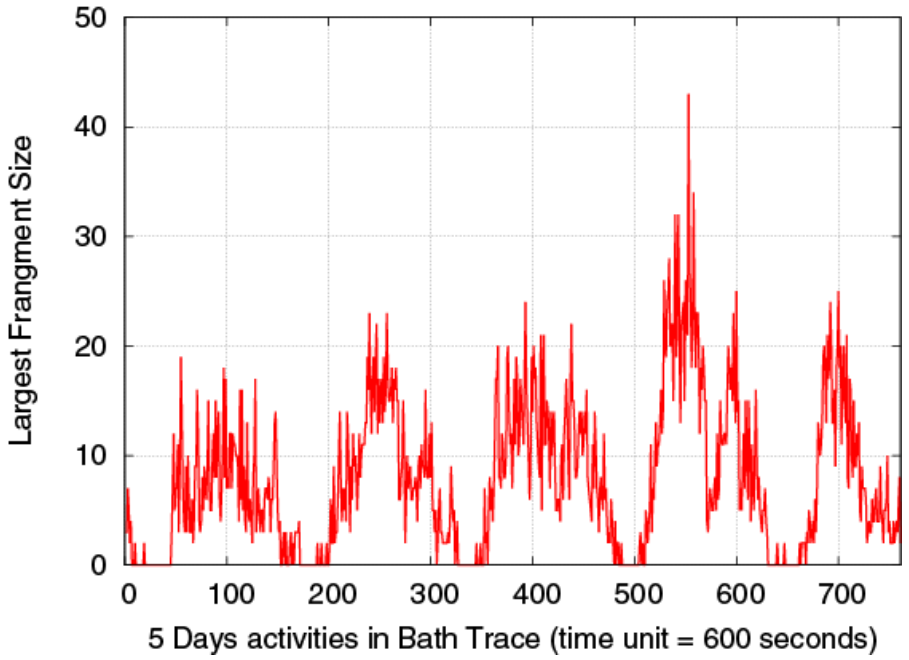
Trace Data of Contact Networks

- Analysis of dynamic network structure
  - How does community structure affect epidemic spread?
  - How do hubs influence temporal or spatial effects, and how does this affect the transmission characteristics of disease?

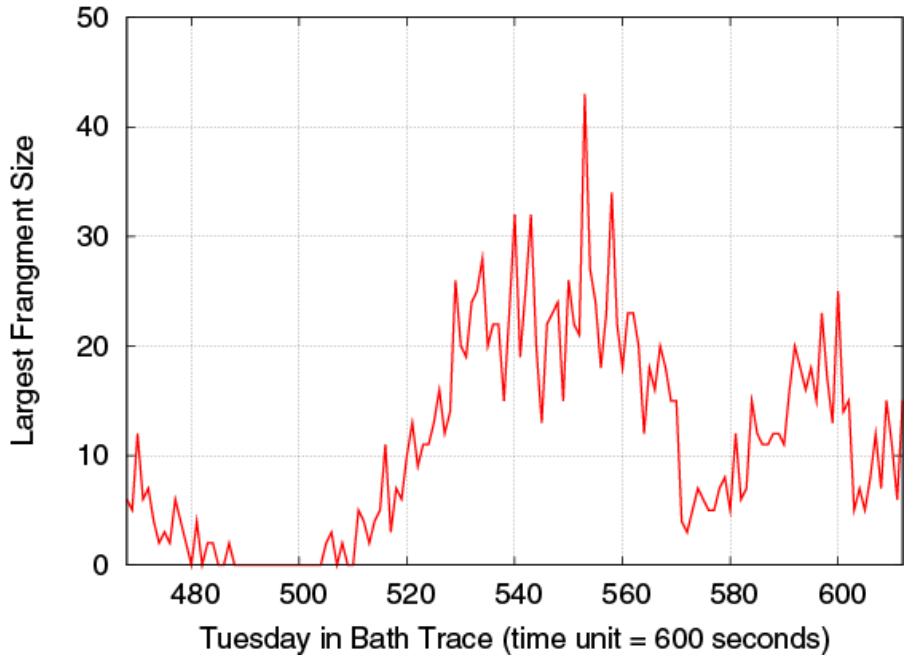
Experimental data set	Cambridge	Infocom06	MIT
Device	iMote	iMote	Phone
Network type	Bluetooth	Bluetooth	Bluetooth
Duration (days)	11	3	246
Granularity (seconds)	600	120	300
Number of Devices	36	78	97
Number of contacts	10,873	191,336	54,667
Average # Contacts/pair/day	0.345	6.7	0.024

Regularity of Network Activity

- Size of largest connected nodes shows network dynamics



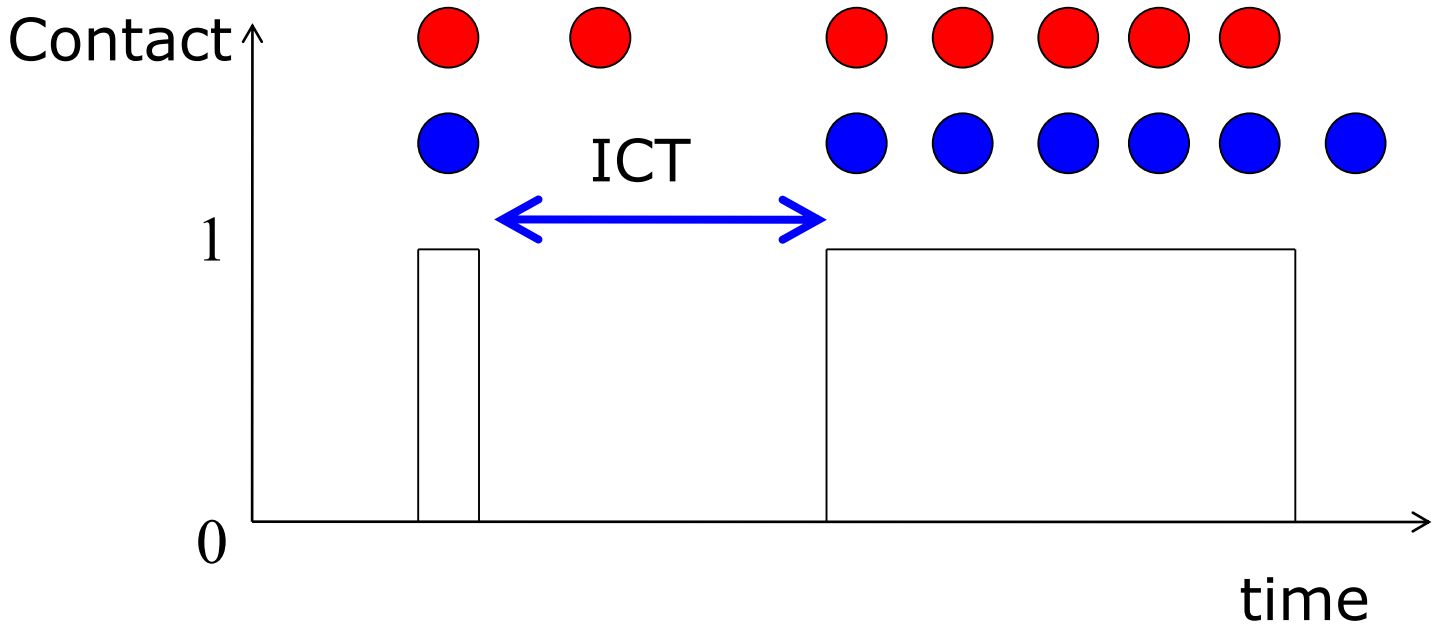
5 Days



Tuesday

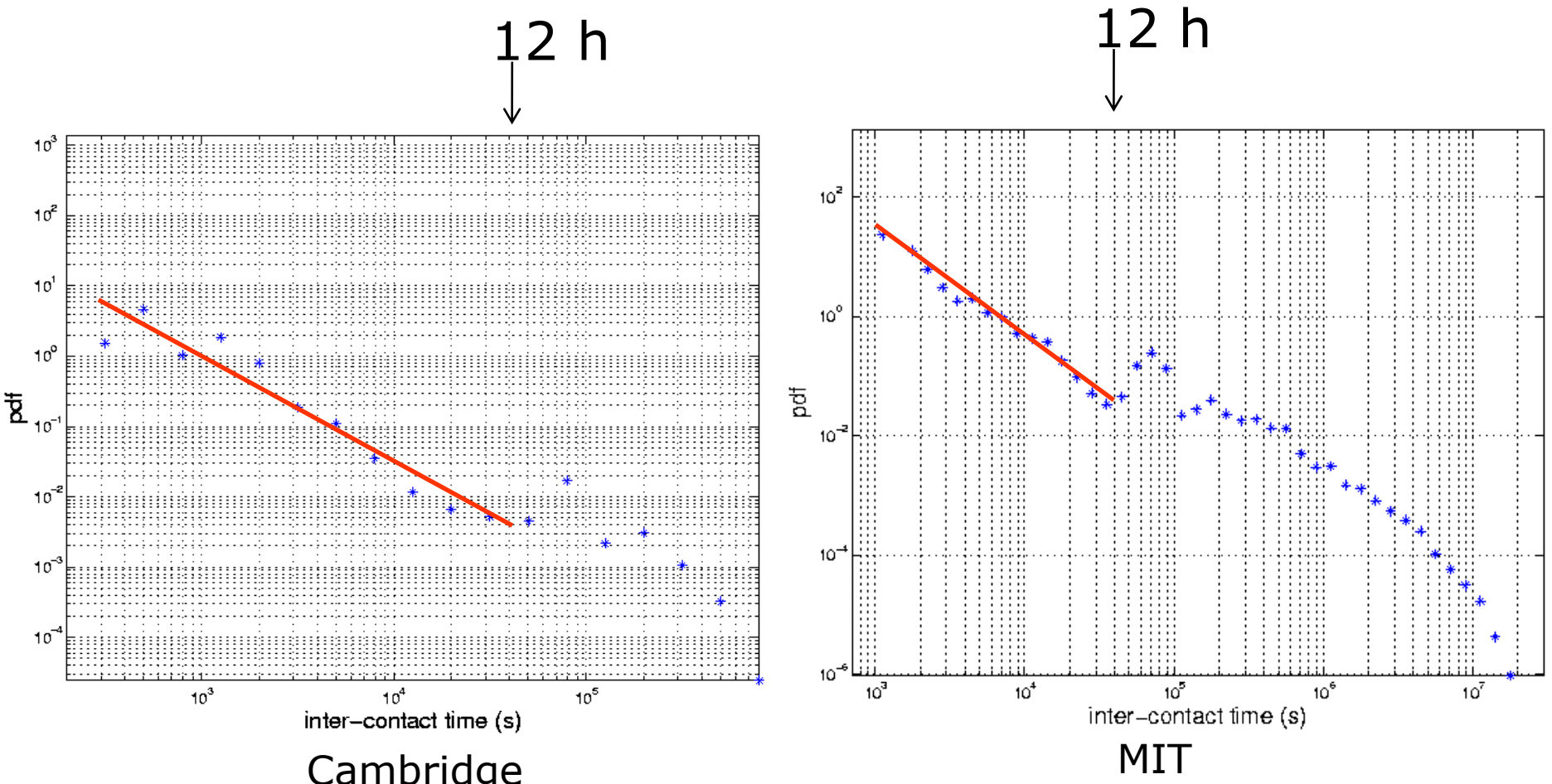
Inter-Contact Time (ICT)

- Calculated all possible inter-contact times between any two nodes, where ICT is defined as time between end of contact among two nodes and start of next contact among same two nodes



ICT: Random and Scale-free

- Sufficiently short time scales (<12 hours): ICT dist is approximated by power law

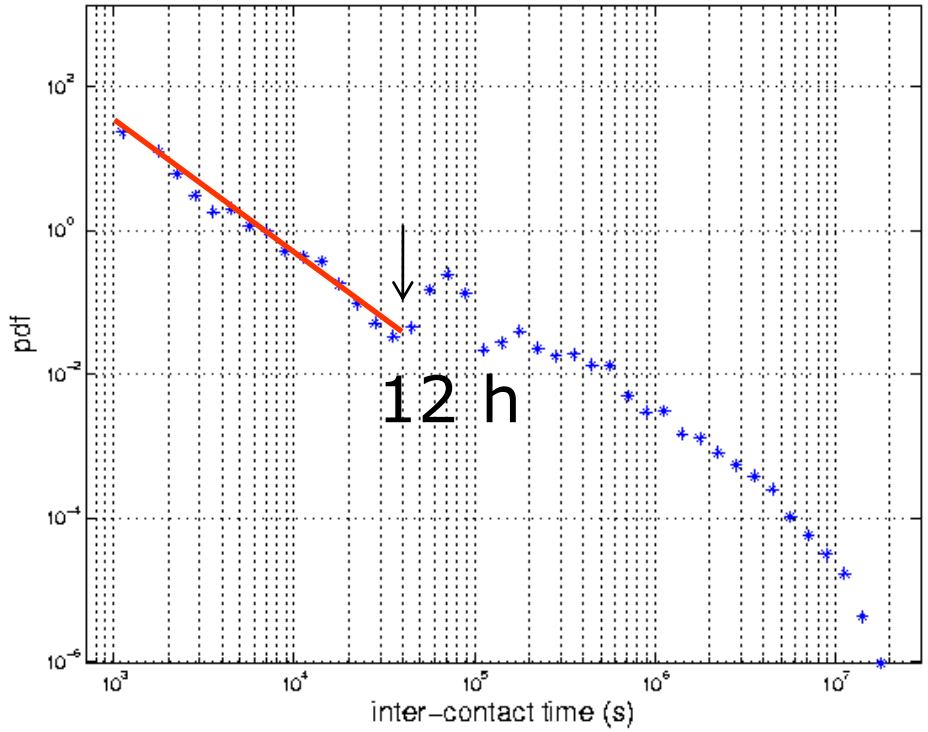


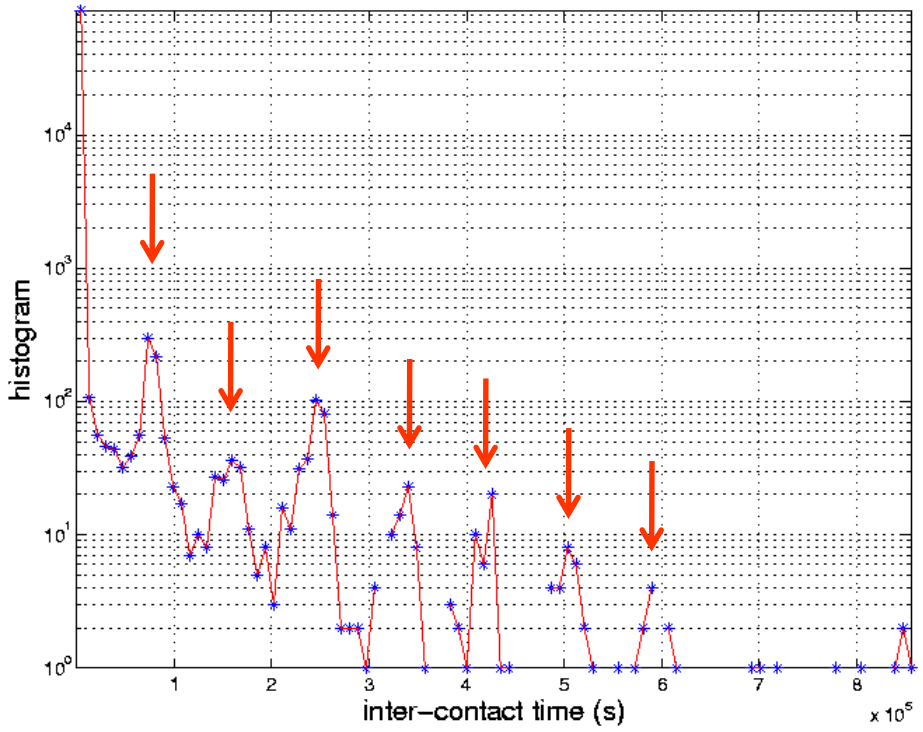
Cambridge

MIT

Inter Contact time: Random/Scale-free and Periodic

- Sufficiently short time scales (<12 hours): ICT distribution is approximated by power law
- Environmental, biological and social constraints show rhythms (ICT separated by 24 hours)

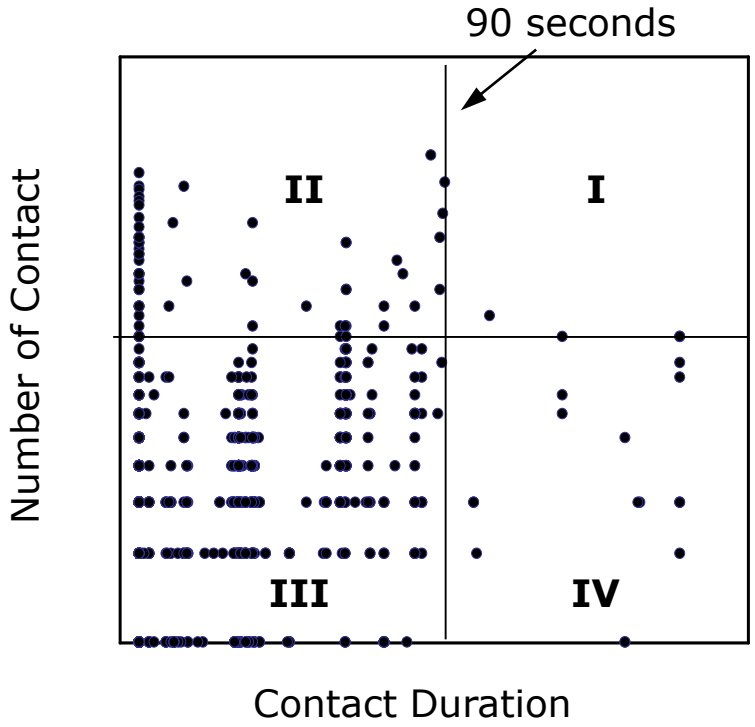




10 Days

Edge Weight

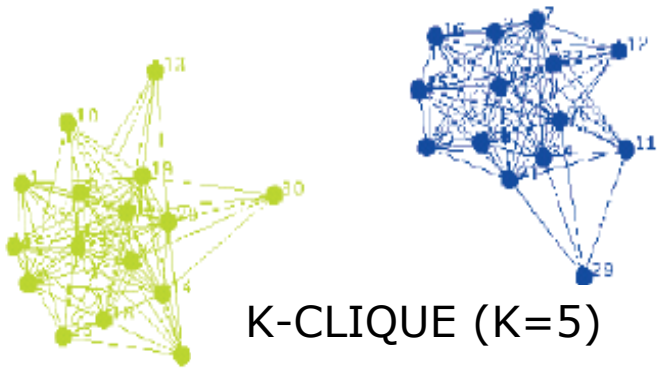
I. High Contact Nº - Long Duration: Community
II. High Contact Nº - Short Duration: Familiar Stranger
III. Low Contact Nº - Short Duration: Stranger
IV. Low Contact Nº - Long Duration: Friend



Uncovering Community

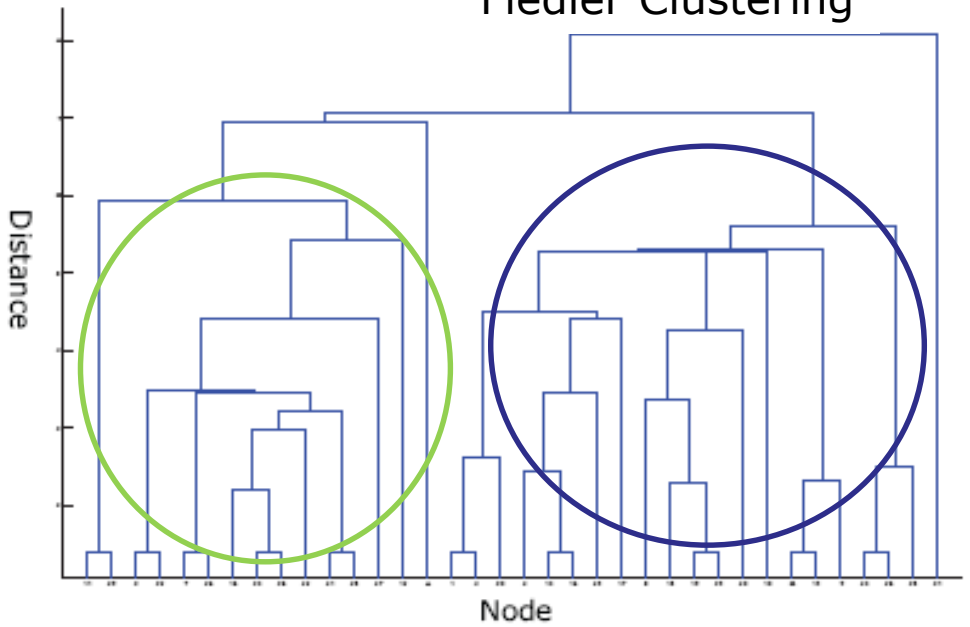
- Contact trace in form of weighted (multi) graphs
  - Contact Frequency and Duration
- Use community detection algorithms from complex network studies
  - K-clique [Palla04], Weighted network analysis [Newman05], Betweenness [Newman04], Modularity [Newman06], Fiedler Clustering etc.

Cambridge Trace

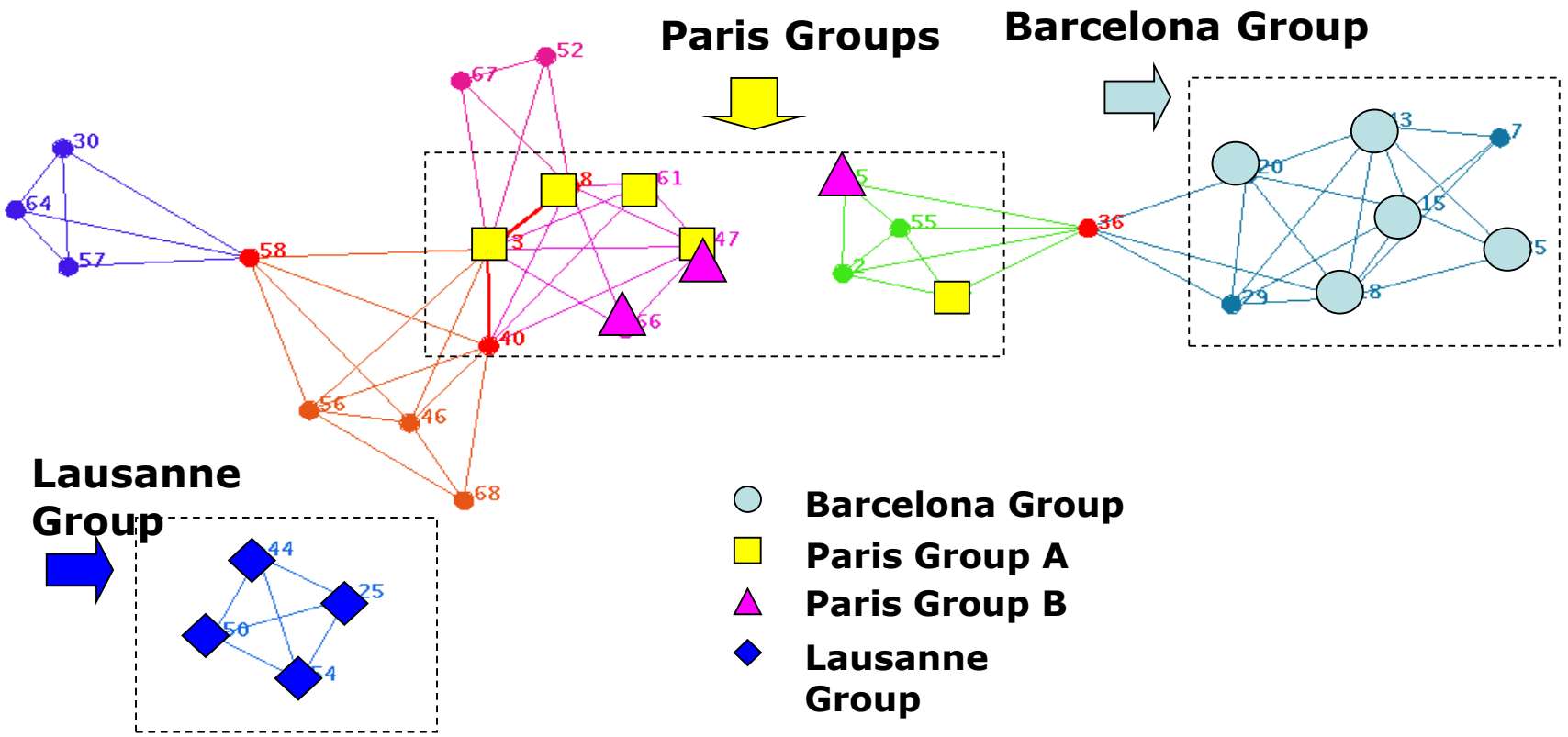


K-CLIQUE (K=5)

Fiedler Clustering

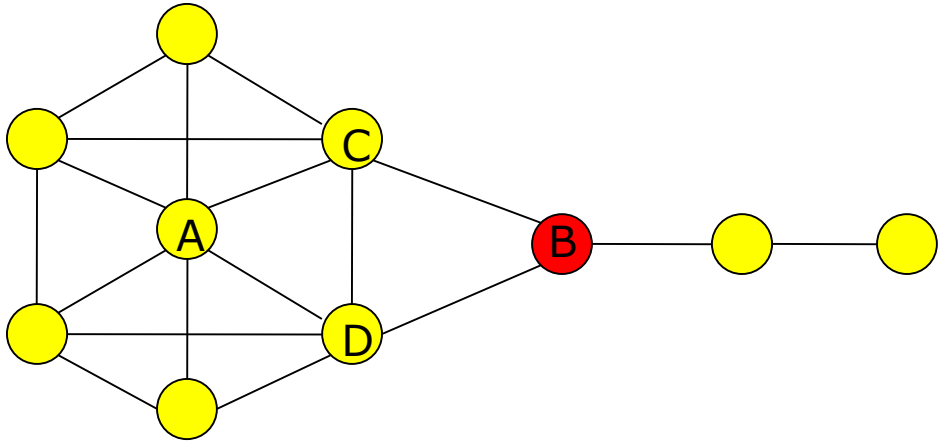


K-CLIQUES Communities with INFOCOM Conference



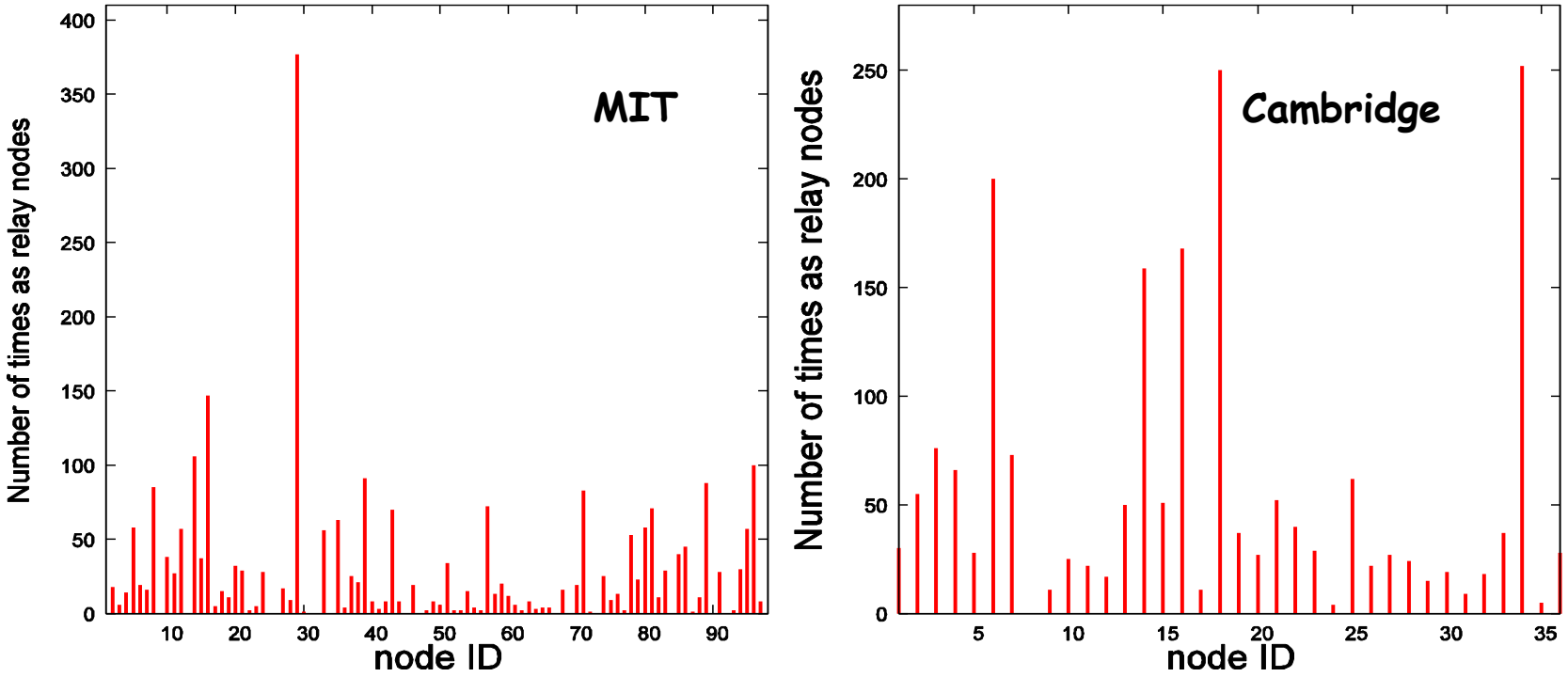
Centrality

- Betweenness Centrality: Control over information flowing between others
  - High betweenness node is important as a relay node
  - Large number of unlimited flooding, number of times on shortest delay deliveries → Analogue to Freeman centrality



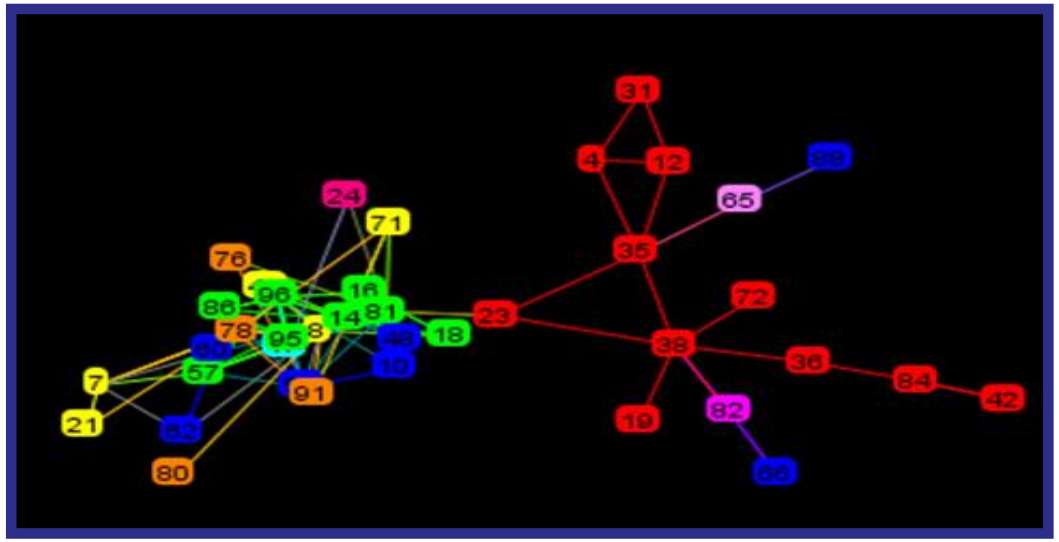
Betweenness Centrality

- Frequency of a node that falls on the shortest path between two other nodes
- High ranking nodes ~= Popular nodes



Contact Networks

- Contact networks: time dependent contacts
  - A → B → C → A
- Think about spread of messages, infectious disease, and gossip
  - Physical contacts, virtual online contacts
- Understand flow and control: key to uncover centrality nodes

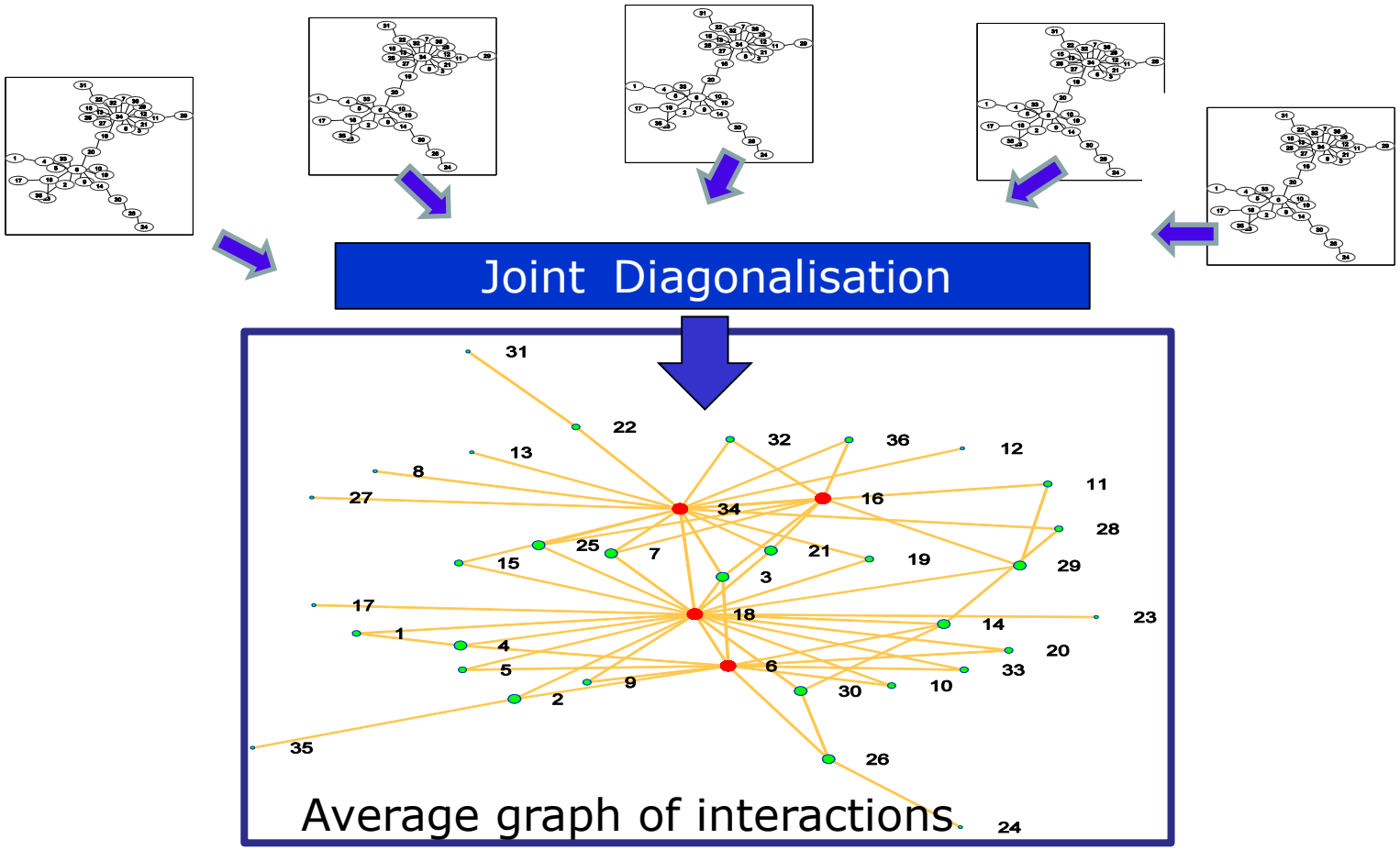


Multiple Spread Modes

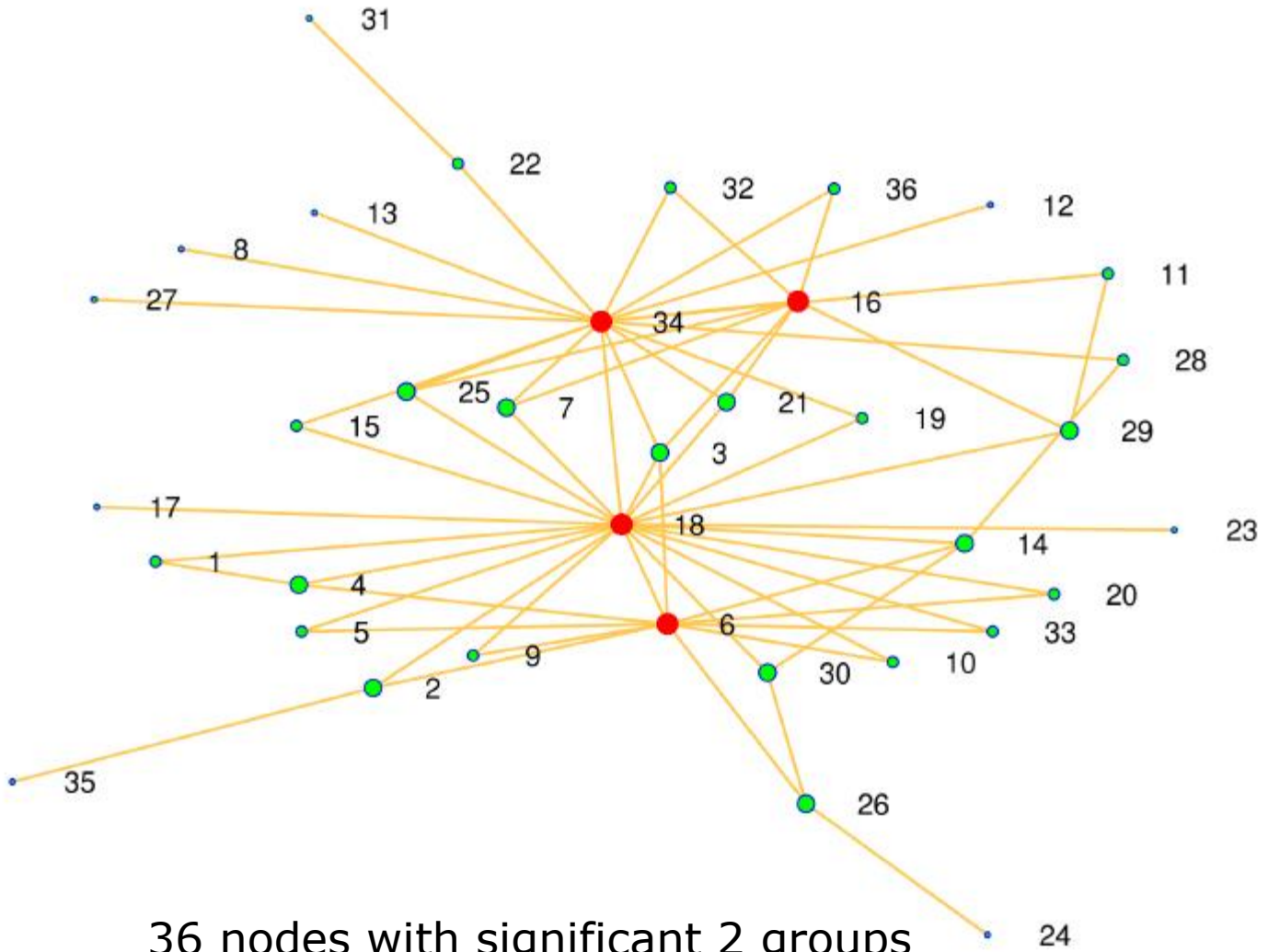
- Typical approach: Cluster nodes to build single network or multiple networks within the sliced time windows
  - Aggregate the number of contacts
  - Ignores time
  - Ignores correlation between links
- Solution: Use spanning tree based samples of a network
  - Akin to spreading a disease in the population and recording the order of infection
  - Define an eigen-space average across these trees
  - Distribution of deviations gives the required groups

Joint Diagonalisation

- Build by combining many of spanning tree based samples of a network using Joint Diagonlisation → Average Interaction Network



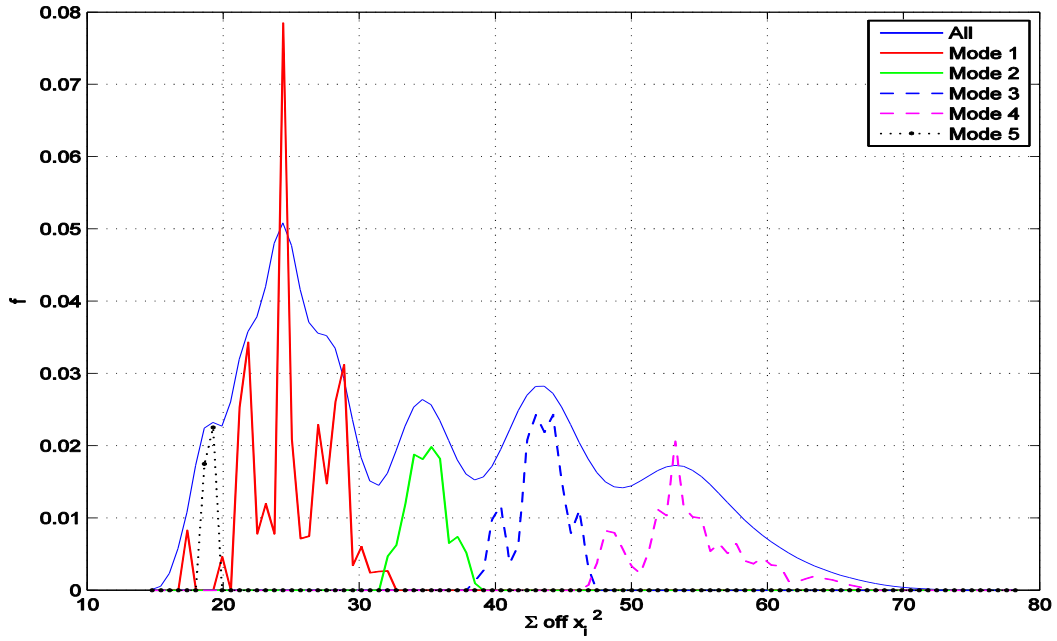
Average Graph of Interactions



36 nodes with significant 2 groups

Multiple Network Modes

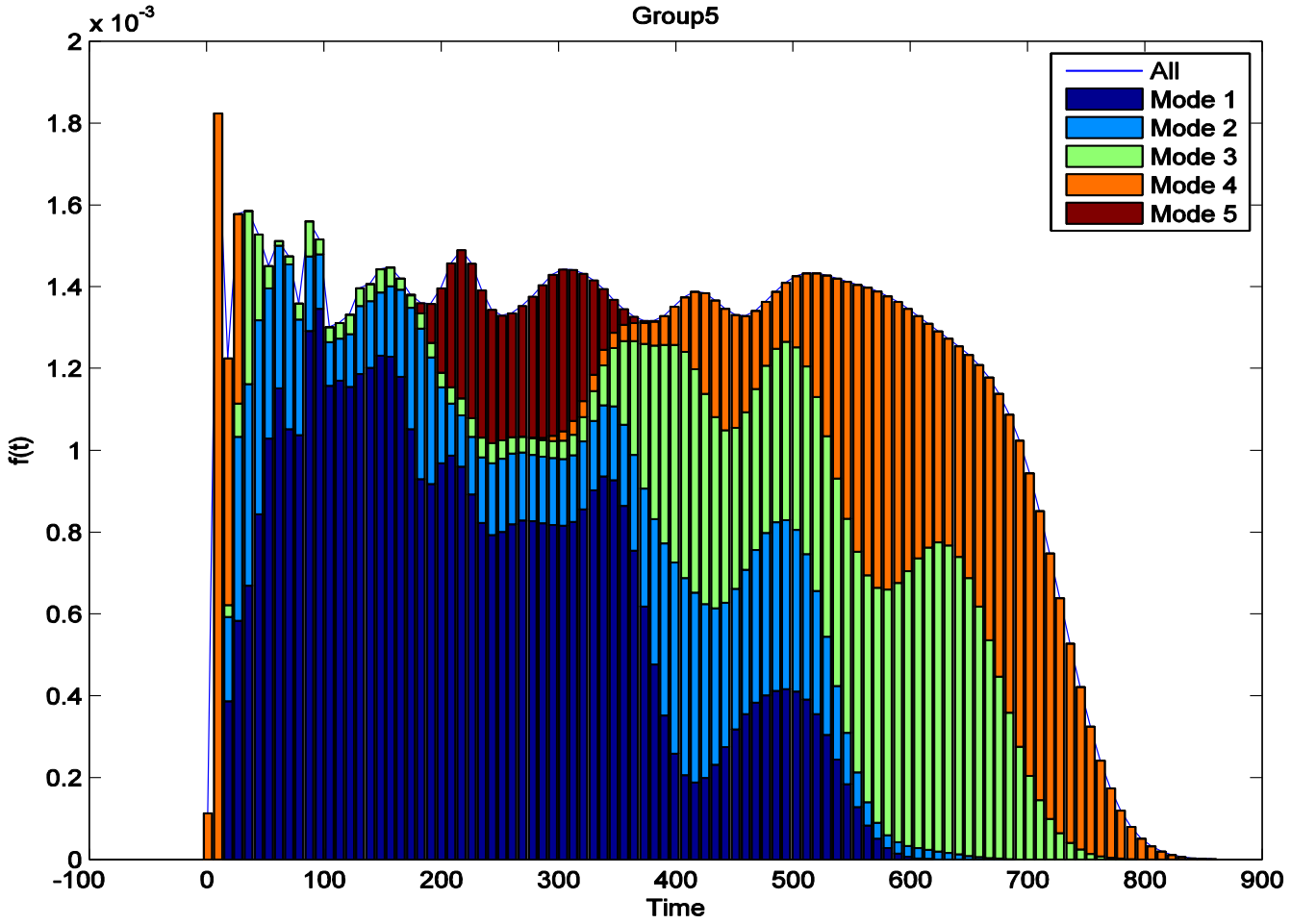
- Build average network with Joint Diagonalisation method
- Define deviation from mean eigen-space as sum of off-diagonal elements
- Use Gaussian mixture model for mode determination
- Distribution of deviation → different behaviour of network



Distribution of deviation: random network shows one mode

Extract Spread Modes

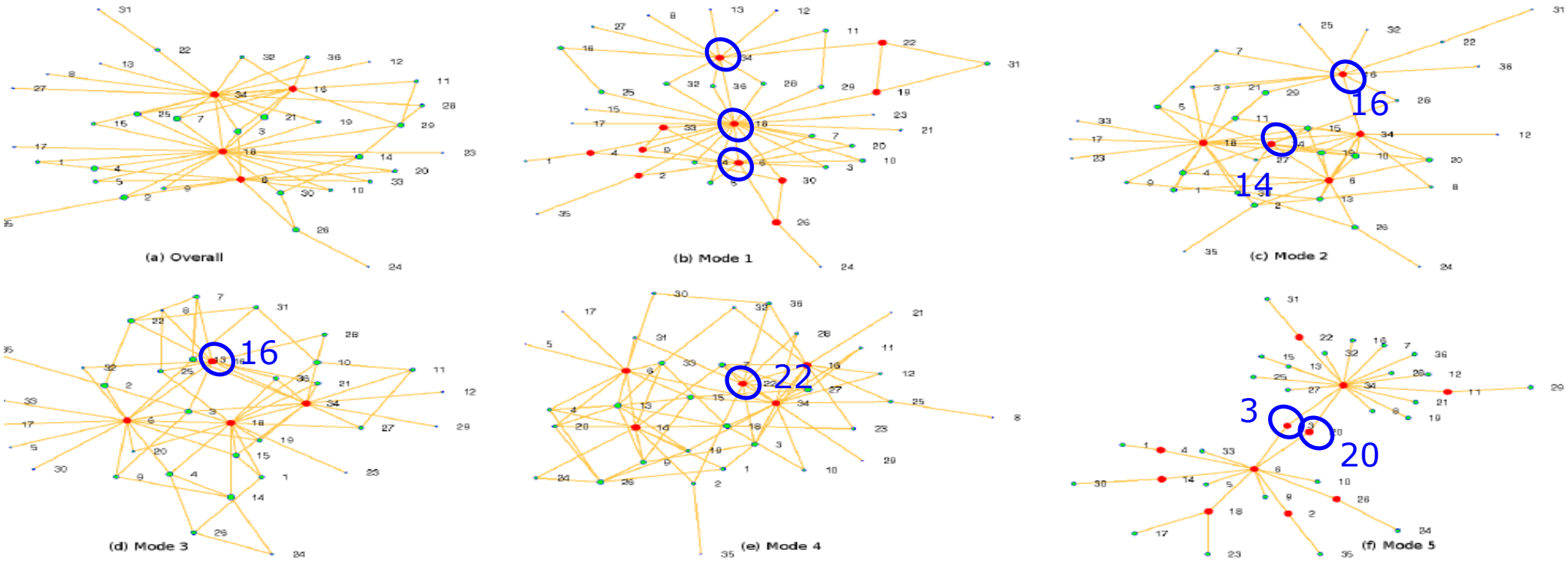
- Change of mode corresponds with state transition



Distribution of times by mode

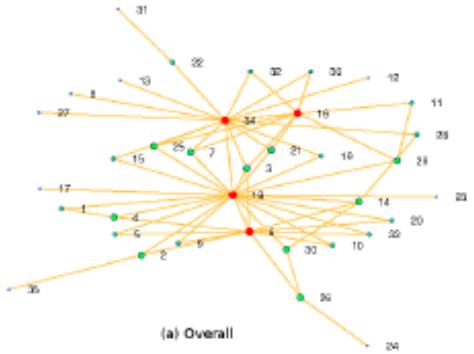
Network Structure of Each Mode

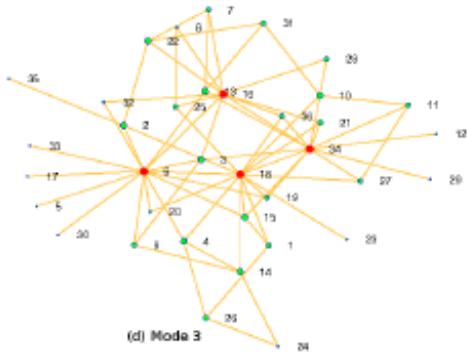
- Mode 1 shows highly structured network corresponding to day when groups are well defined by group dependent activity
- Mode 5 is particularly interesting as there is obvious bridge formed by nodes 3 and 20

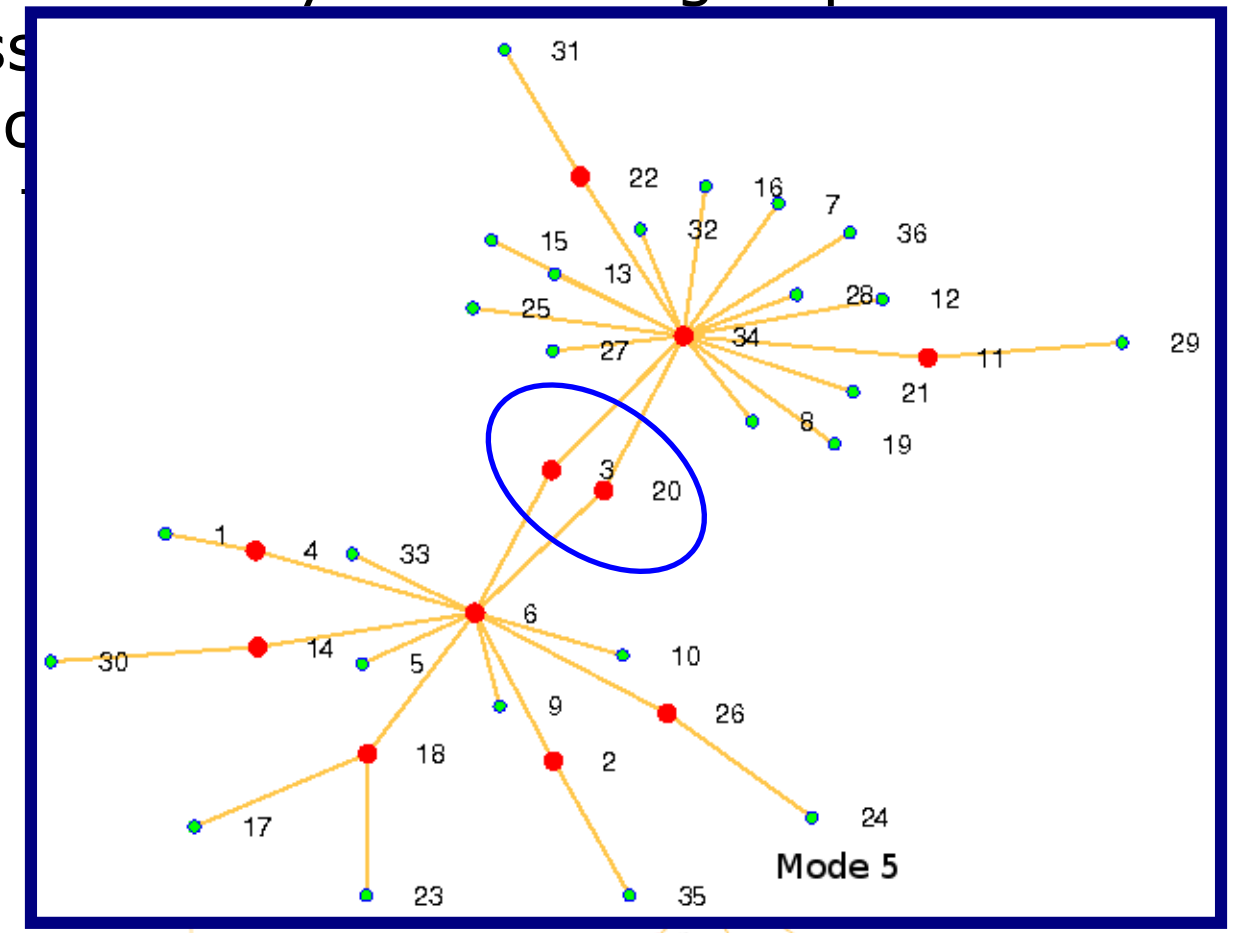


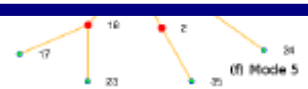
Network Structure of Different Modes

- Mode 1 shows a highly structured network corresponding to the day when the groups are well defined by class
- Mode 5 is partic
obvious bridge









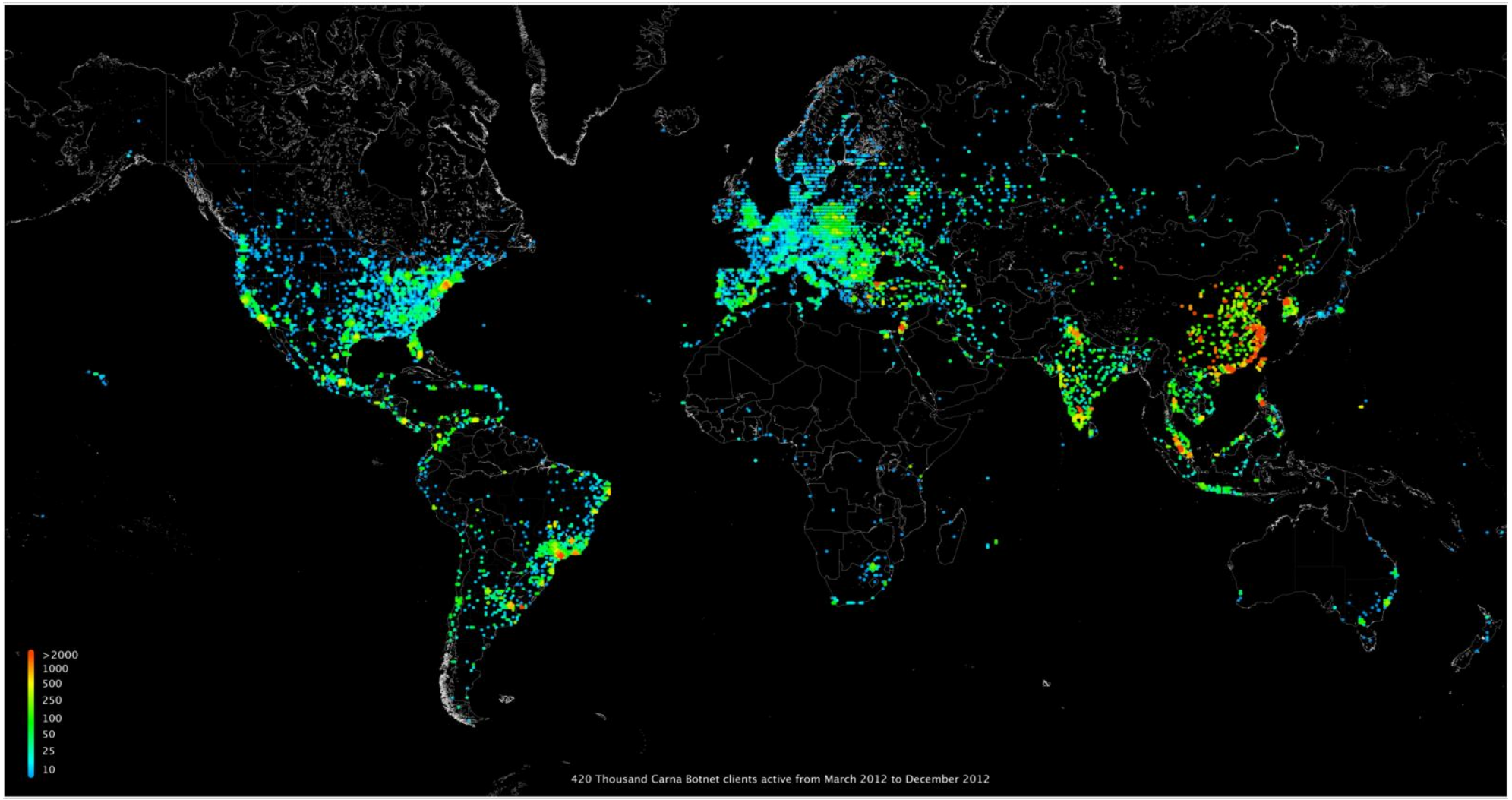
Data Collection in Developing Countries

- Studies in Africa and South America are desired
- Target diseases: Measles, tuberculosis, meningococcal, respiratory syncytial virus and influenza

Sensing Platform in Remote Region

- Build a platform for sensing and collecting data in developing countries
  - Build a standalone network for data collection and communication using Raspberry Pi → RasPiNET
  - Inexpensive network setting
  - Support streaming model
  - In-network partial data processing

Internet Coverage



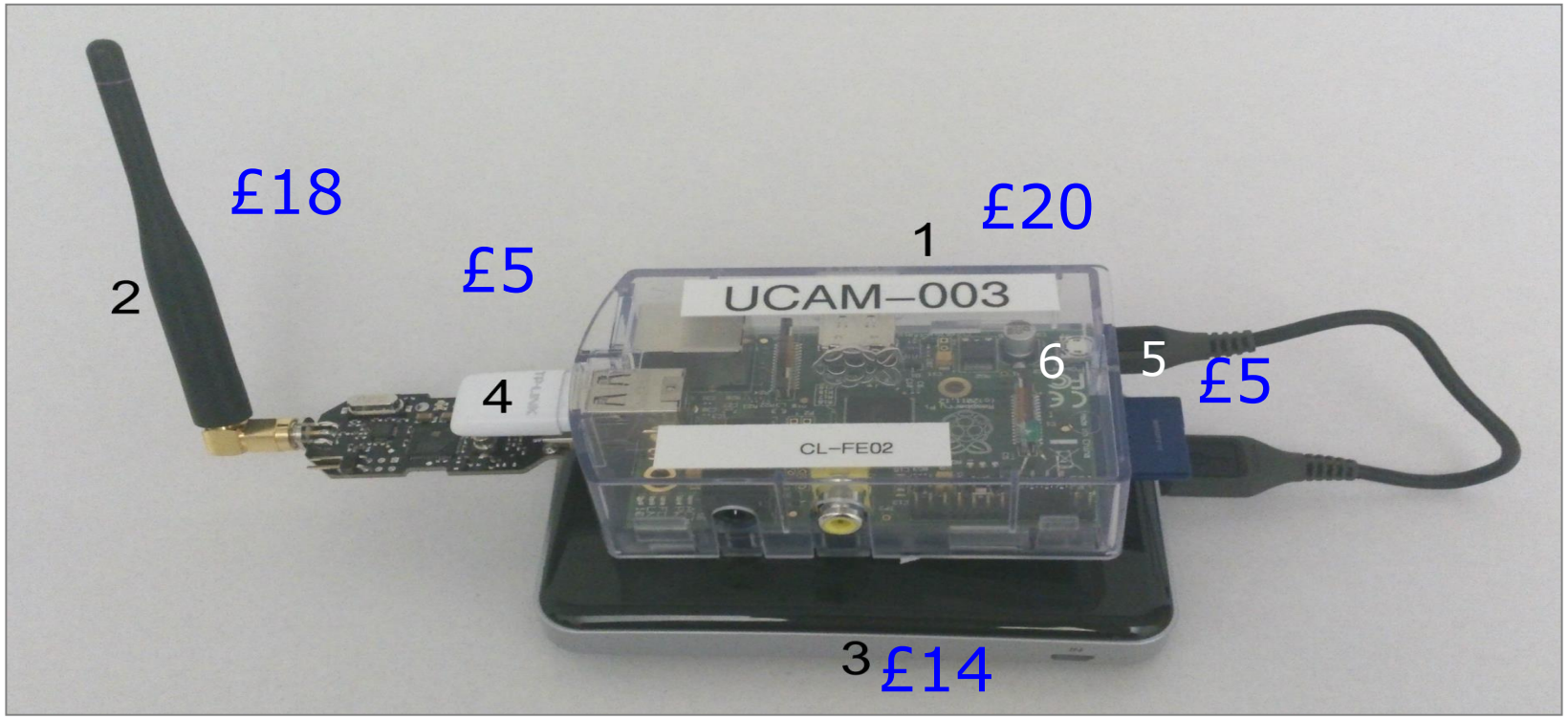
OpenBeacon RFID Tags

- OpenBeacon Active RFID Tags
- Bluetooth has an omnidirectional range of ~10m
- OpenBeacon active RFID tags: Range ~1.5m and only detect other tags are in front of them
- Low Cost ~=10GBP
- Face-to-Face detection
- Temporal resolution 5-20 seconds
- On-board storage (up to ~4 logs)
- Battery life ~2-3 weeks

An OpenBeacon RFID tag

OpenBeacon Ethernet EasyReader

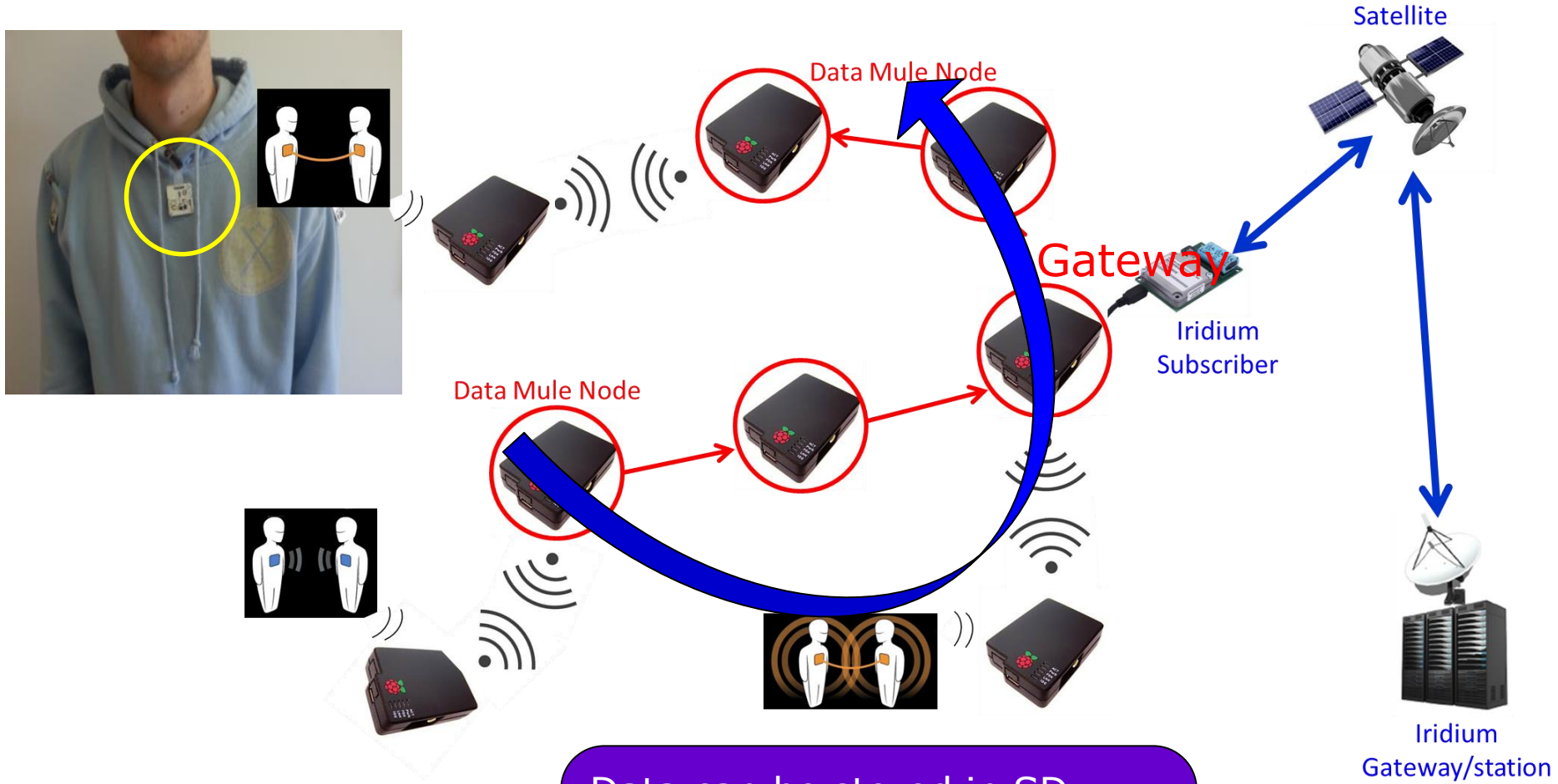
Raspberry Pi Open Beacon RFID Tag/Reader



Open Beacon RFID Tag
Range ~1.5m
Low cost 10GBP

1. Raspberry Pi
2. OpenBeacon USB reader
3. Battery Pack (7000mAh)
4. WiFi dongle 5. SD Card 6. LED

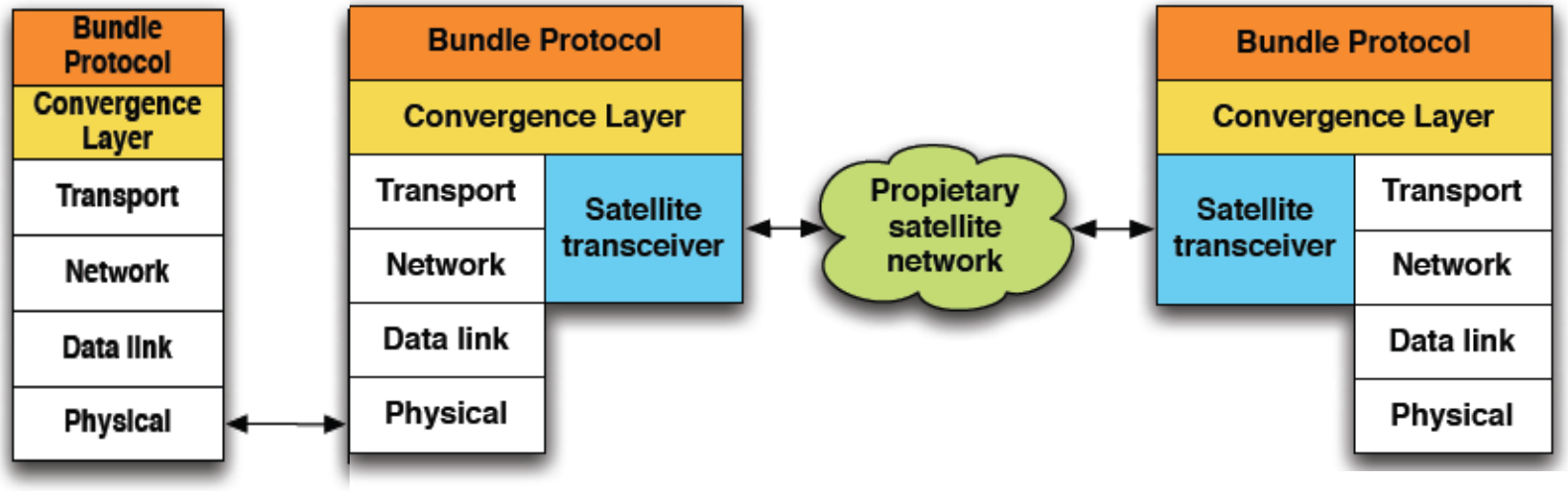
Raspberry Pi based Sensing Platform



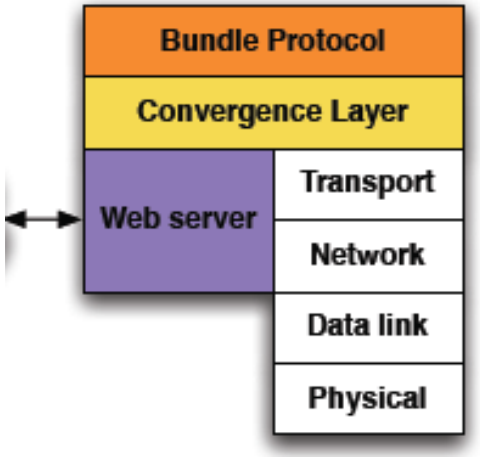
Data can be stored in SD card, transmitted to Data Mule node, or use of WiFi AdHoc mode transmission to Gateway

Communication Protocol

- Protocol for communication between devices with satellite transceiver



- Rockblock provides Web Service Interface
- also Email Interface



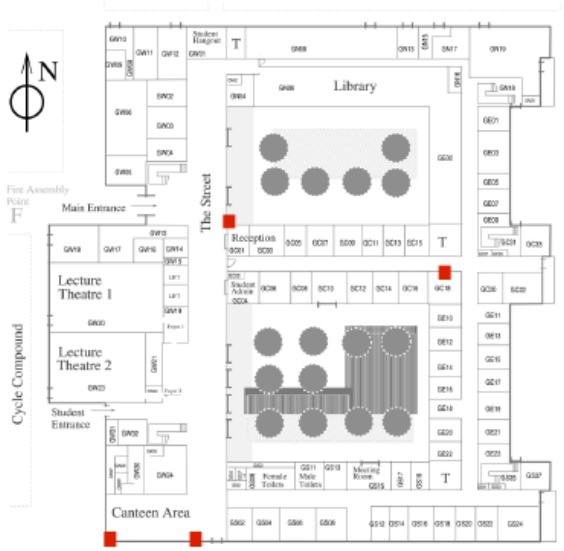
Pilot Study in Computer Laboratory

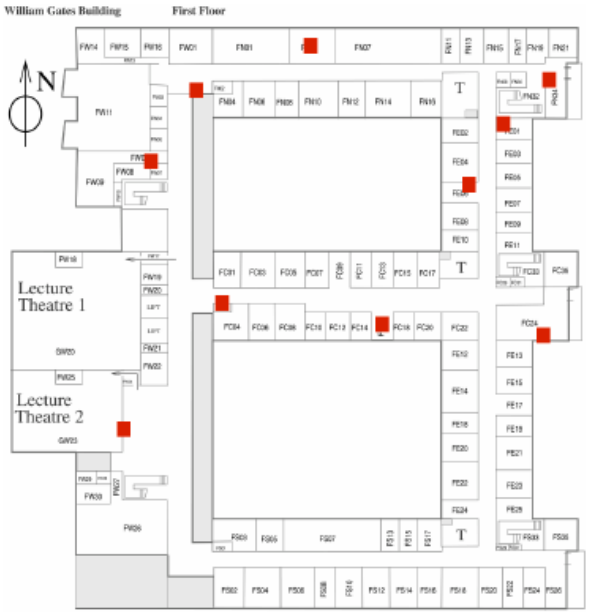
- 15 RasPi OpenBeacon Readers around Computer Laboratory
- 30 participants (4 groups)
- 3 days of data collection

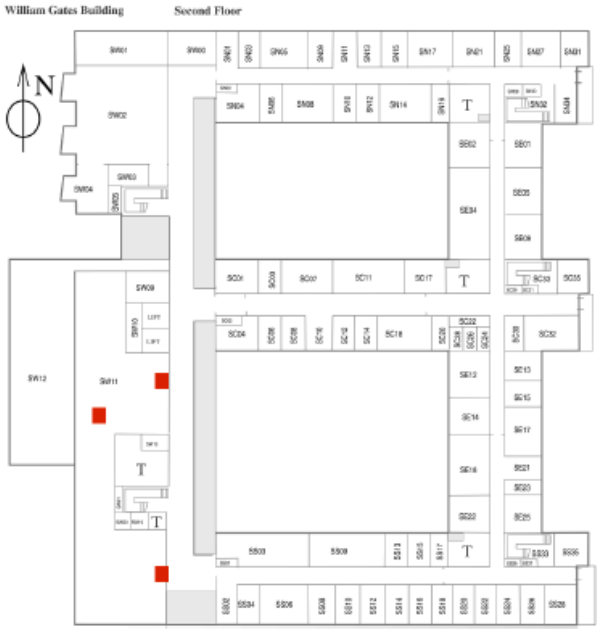
A participant wearing three RFID tags

Setting RasPiNET on 3 Floors

- Use of Data Mule approach for Data Collection
- Satellite Communication for sending statistics and changing sensing rate

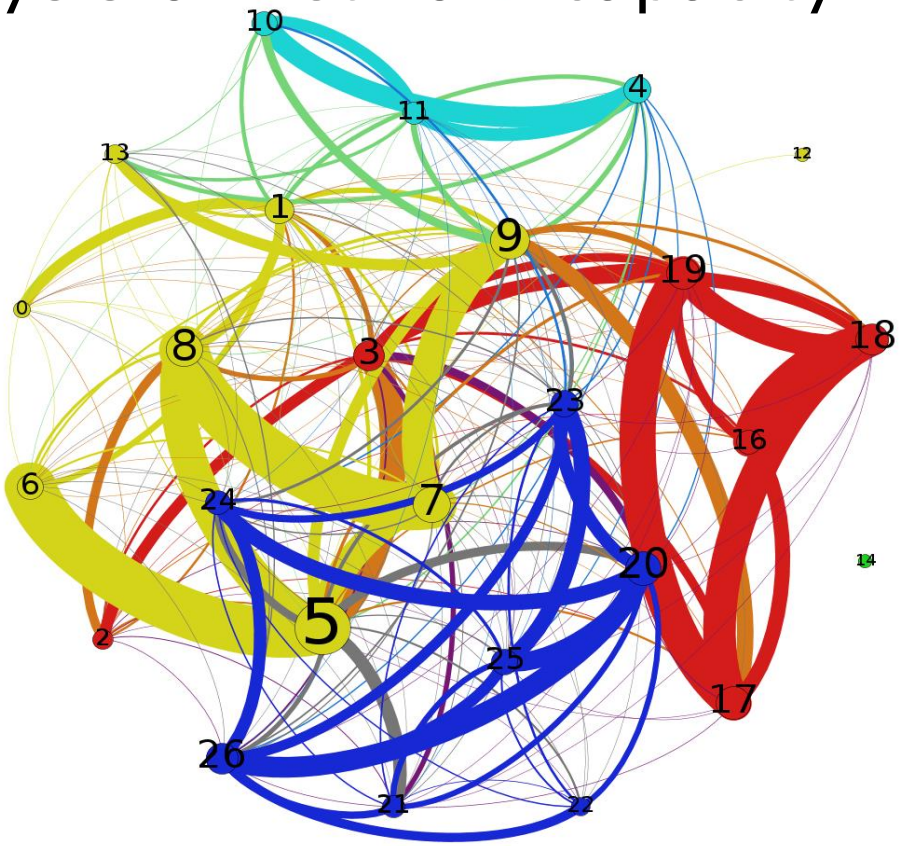






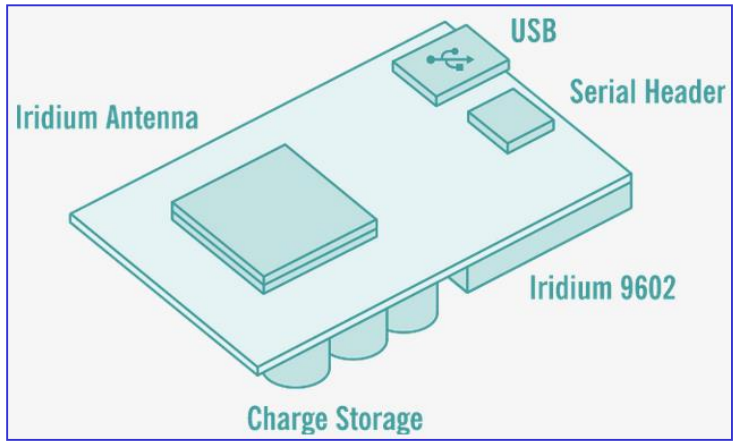
Post Data Analysis on Pilot Study

- Community Detection (4 groups and bridging nodes can be identified)
- No in-depth traffic analysis or network capacity evaluation yet
- One simulator based Simulator (w and w/o satellite connectivity)



Satellite Communication

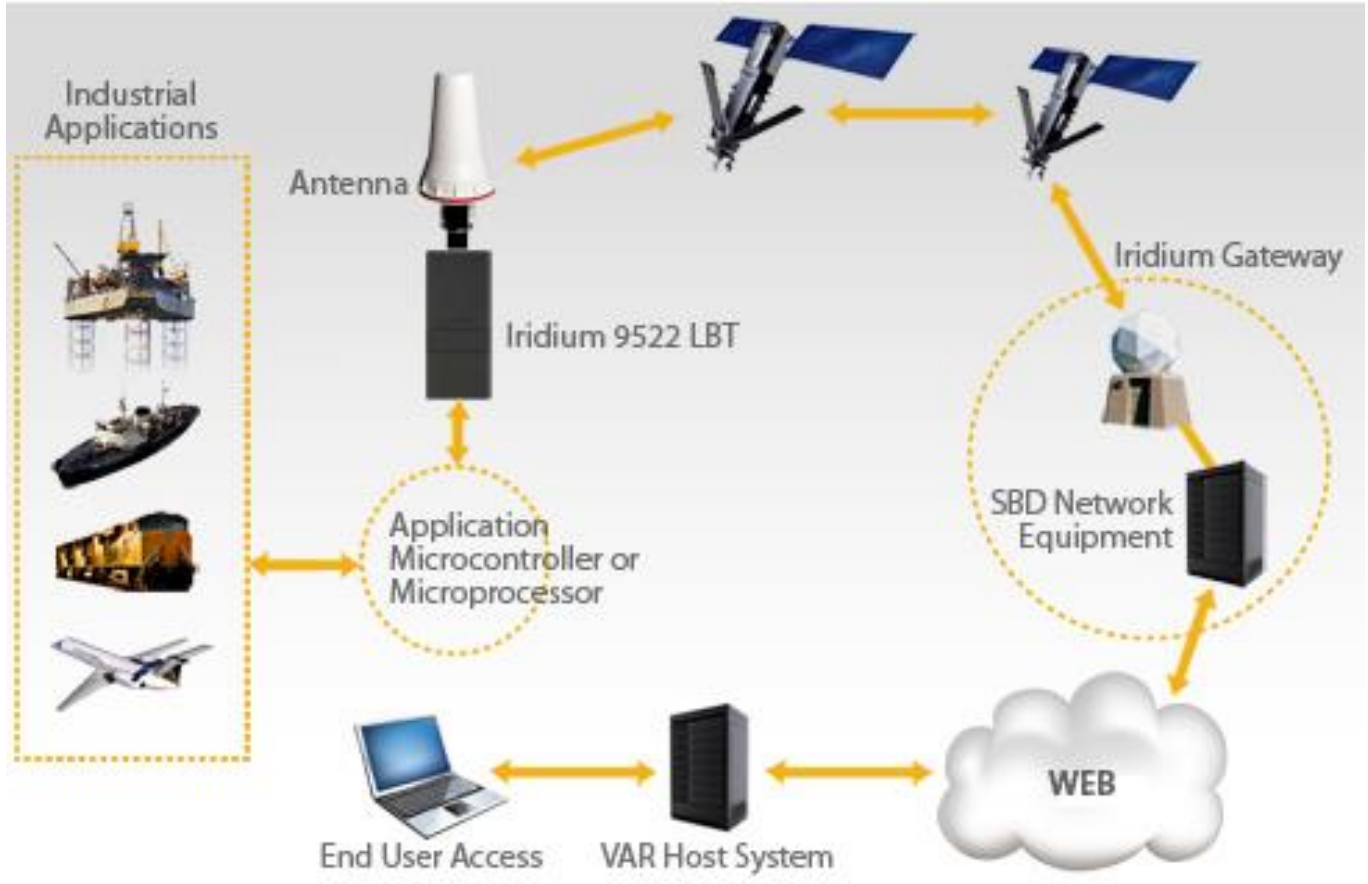
- Satellite module integration in Raspberry Pi
  - RockBLOCK Satellite Module (~=£120)
  - Uses Iridium Satellite Network: Short Burst Data(SBD)
  - Iridium SBD session roughly every 10 seconds
  - To email address, or own web service (i.e. HTTP POST)
  - pay-as-you-go – 34 bytes per message (Hex encoded)
    - 50 credits - 12p/message
    - 20000 Credits – 4p/message



RockBLOCK satellite module

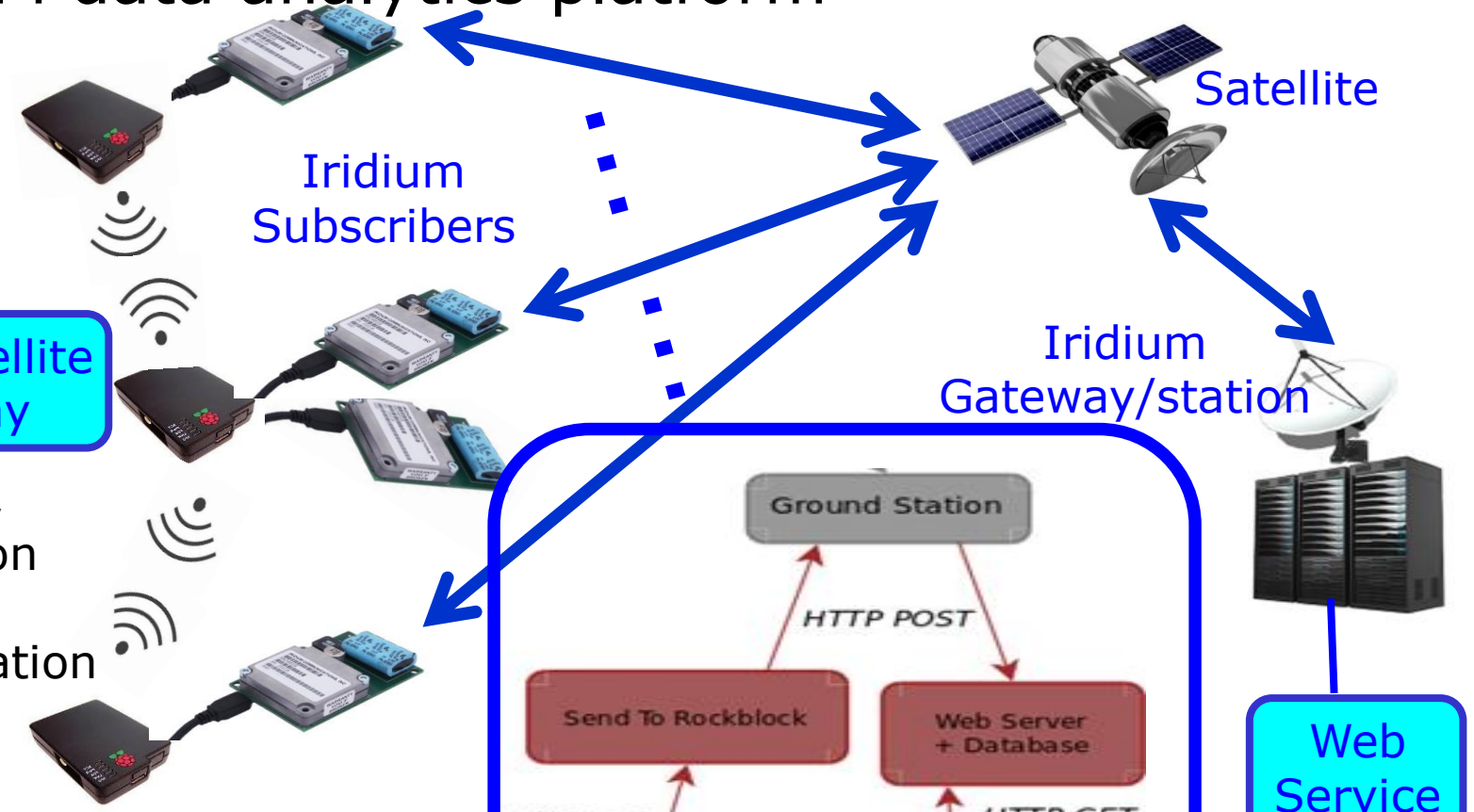
RasPiNET with Satellite Communication

- Satellite module is integrated
- Useful in developing country



RasPi Satellite Gateway

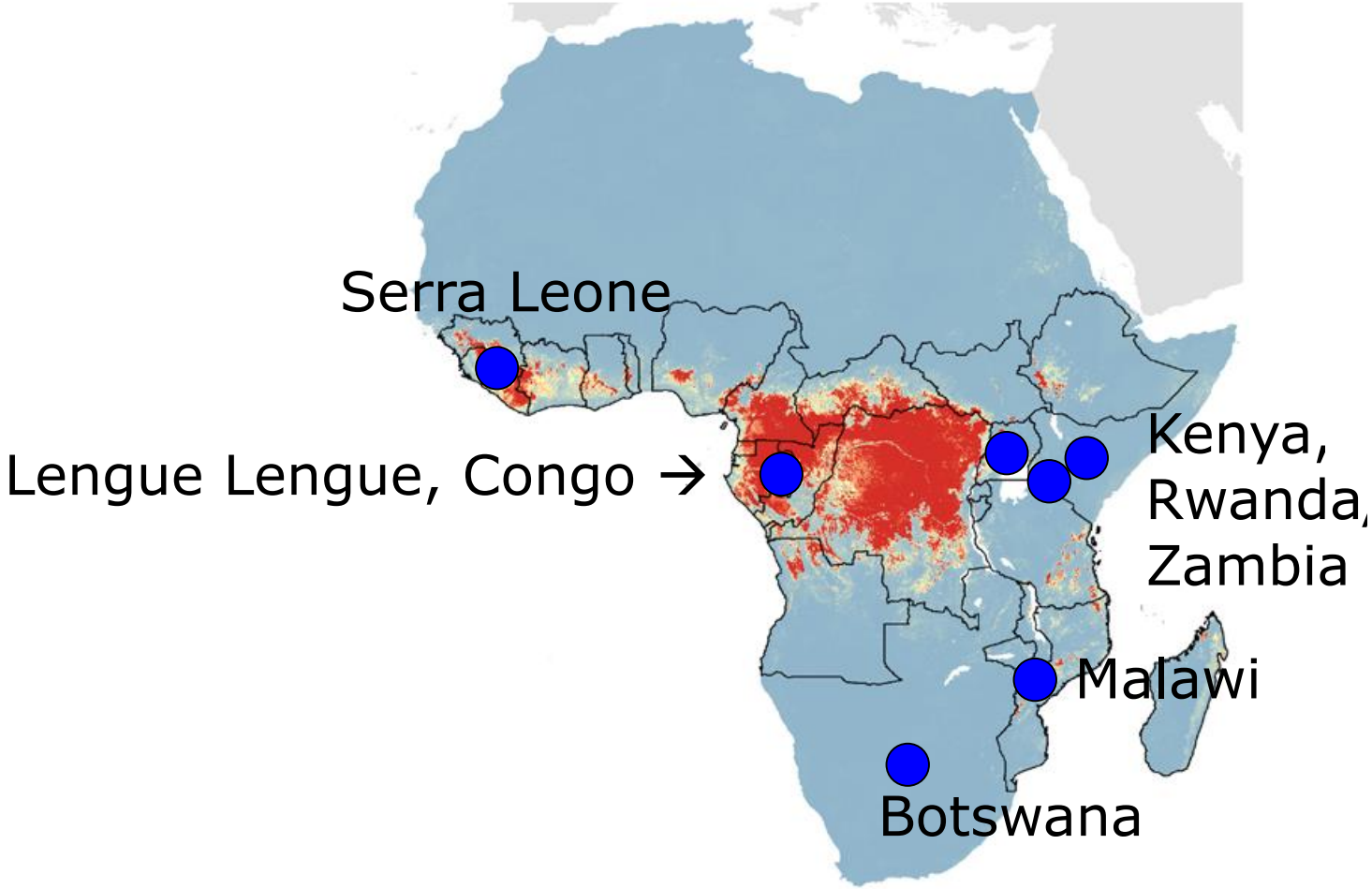
- Build stream processing paradigm
- RasPi data analytics platform



- Data filter, aggregation
- Analysis
- Fragmentation
- Multicast support
- Combine split data

In Africa, Several Projects Planned…

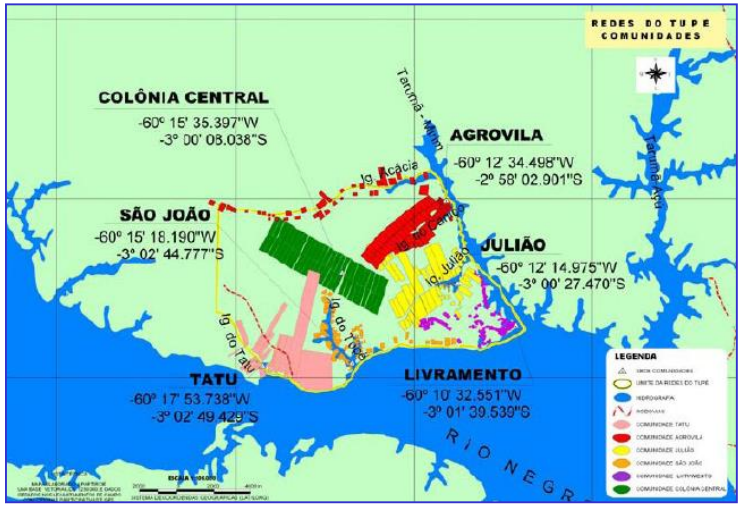
- Complicated ethical approval → killed the projects



2015 Brazilian Amazon

- Local help by Federal University of Amazonas
- Small scale data collection from few fragmented rural communities along Rio Negro (~90 People, ~70 Raspberry Pis)

Tupé Sustainable Development Reserve



Rio Negro Communities

Digital Epidemiology with RasPiNet

- Rhythm and Randomness in Human Contact:
  http://arxiv.org/abs/1009.3980, 2010
- On Joint Diagonalisation for Dynamic Network Analysis:
  http://arxiv.org/abs/1110.1198, 2011
- EpiMap: Towards Quantifying Contact Networks for Understanding Epidemiology in Developing Countries: Elsevier Ad Hoc Networks Journal: Special Issue on Wireless Technology for Humanitarian Relief, 2014

RasPiNET: Decentralised Network for Data Collection and Communication with Raspberry Pi

eiko.yoneki@cl.cam.ac.uk
http://www.cl.cam.ac.uk/~ey204