Influential Neighbours Selection for Information Diffusion

Eiko Yoneki

Systems Research Group
University of Cambridge Computer Laboratory

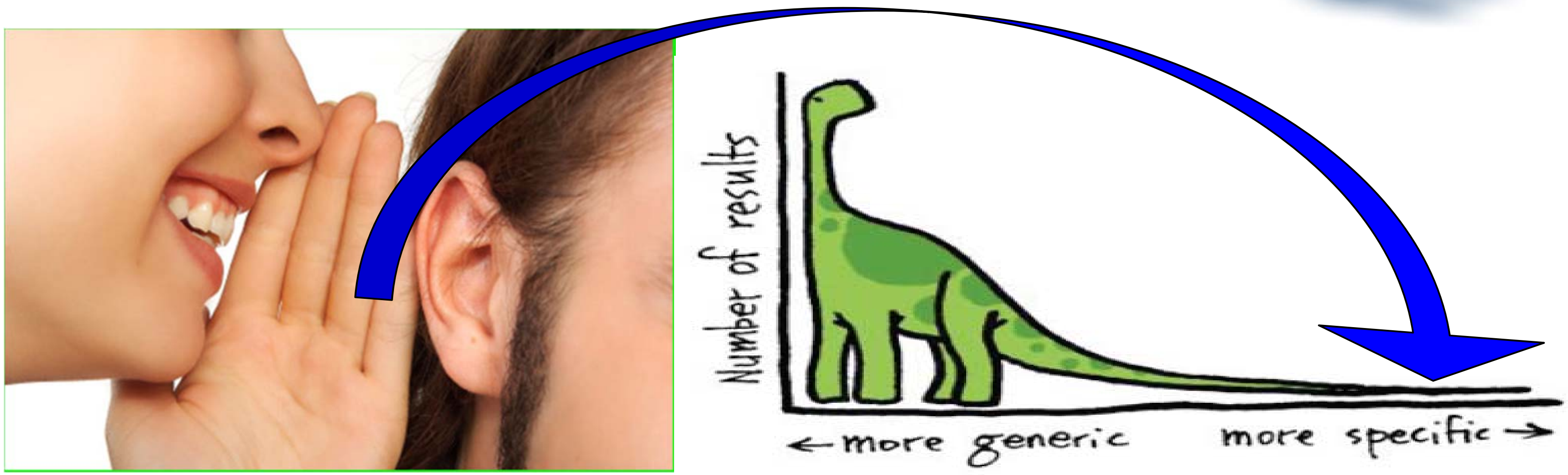
iSocial, Crete, May 21st, 2015

Outline

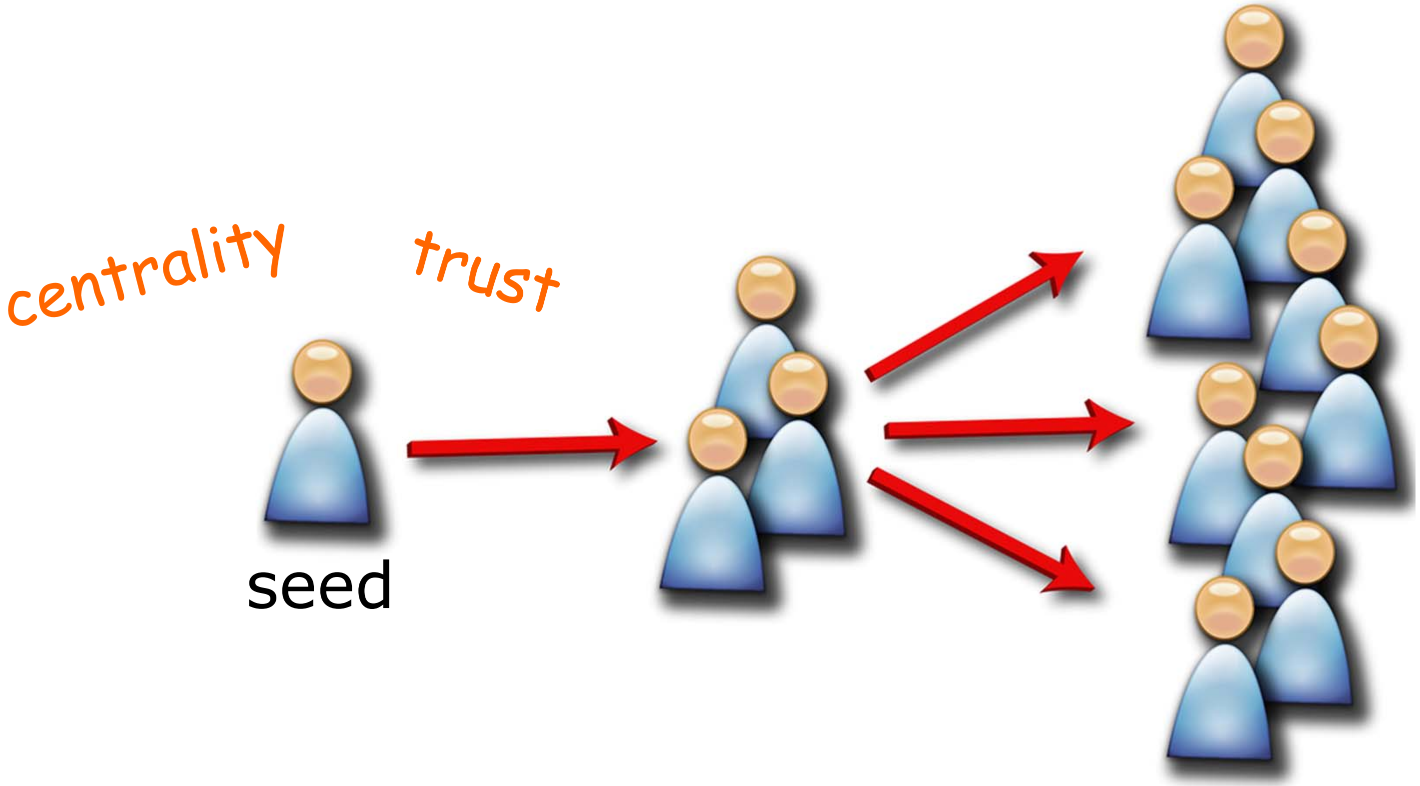
- Influence Maximisation Study: Selection of propagation nodes
- Decision at node: Social influence
  - EEG
  - Eye tracking
- Digital Epidemiology

OSNs for Spread of Influence

- Online Social Network (OSN) plays a fundamental role as a medium for the spread of influence among its members
- 90% of consumers trust peer recommendations while 14% trust advertisement



Importance of information seeds

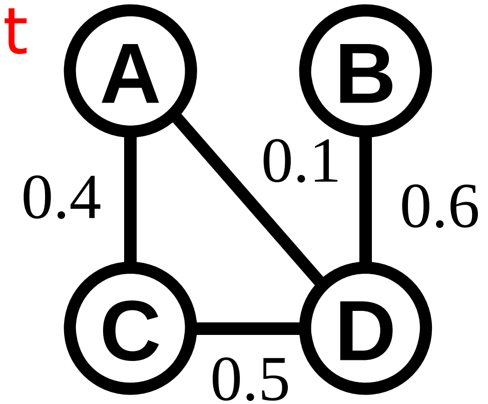


Influence might be changed with information seeds

Influence Maximisation

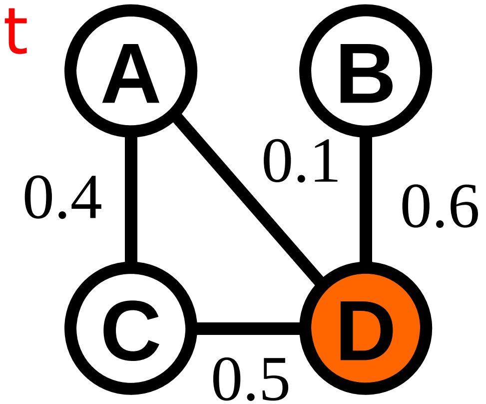
- Problem: (Domingos et al., 2001; Kempe et al., 2003)
  - Given a social graph $G = (V, E)$ with influence probabilities on edges, select k individuals such that by activating them, the expected spread of influence is maximised

Input



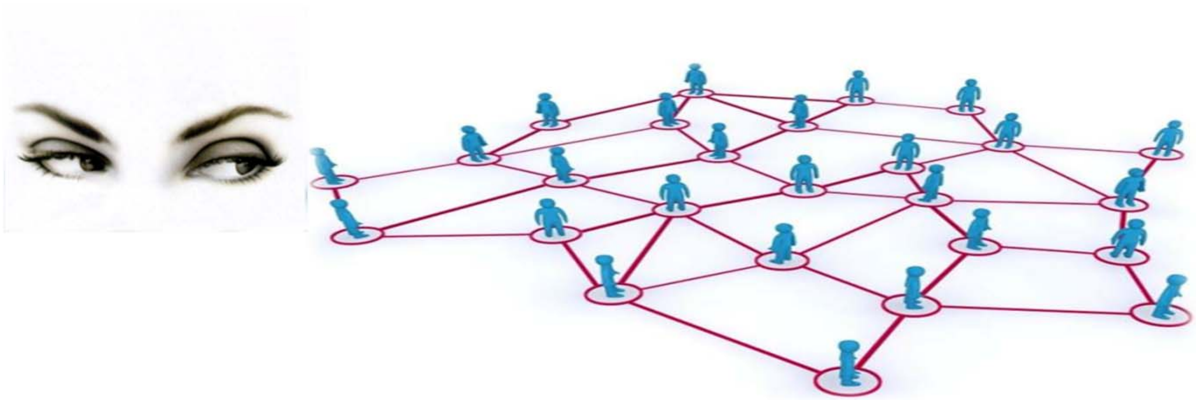
Social graph with influence probabilities of edges

Output



$k = 1$

Limitations of Influence Maximisation



This model requires a bird's eye view of an entire social graph. In real world, who knows the whole network topology?

In practice, a node can initially share the information with only some of its neighbours rather than a set of any arbitrary nodes.

Research Question

How can the neighbours be effectively chosen for information diffusion in OSNs?

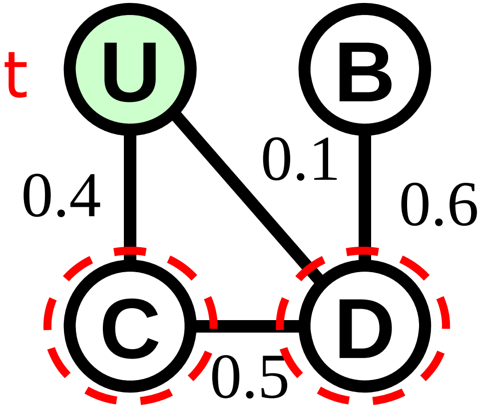
For example, when $k=1$, we may choose the most powerful(?) neighbour as the activated node

Influential Neighbour Selection

- Influential Neighbours Selection (INS) problem:

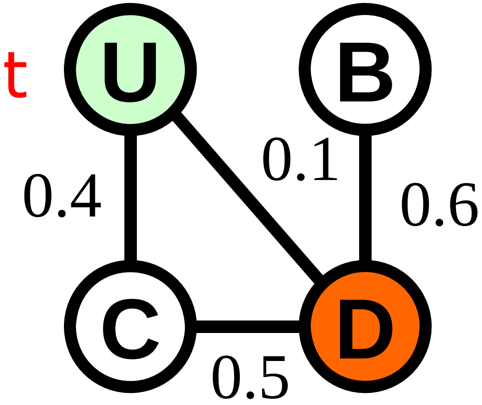
Given a social graph G = (V, E) with influence probabilities on edges and a node u, select u's min(k, degree(u)) neighbours such that by activating them, the expected spread of influence is maximised

Input



Social graph with influence probabilities of edges and *U*

Output



$k = 1$

Our Assumptions

1. Each node only communicates with its immediate neighbours

2. Each node has no knowledge about the global network topology

3. Each message size is bounded to O(log |V|) bits

4. For simplification, we use a constant influence probability for all edges

Neighbours Selection Strategies

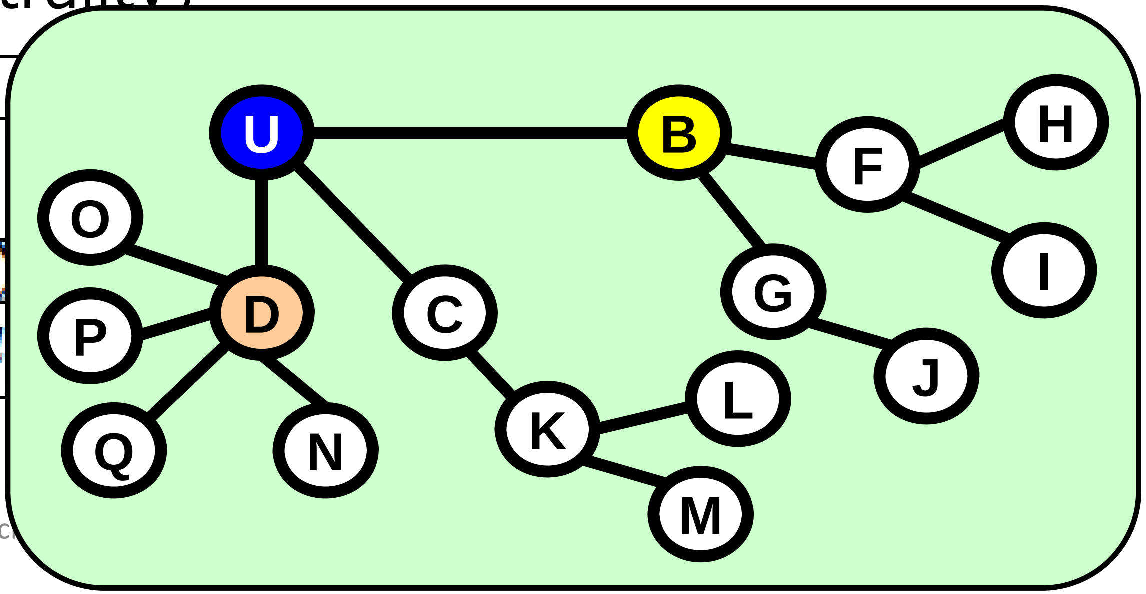
- Set 4 selection strategies based on local connectivity pattern such as degree and clustering coefficient
  1. Random selection
  2. High degree selection
  3. High volume selection (by Wehmuth and Ziviani) selection
  4. High weighted-volume selection (a good approximation of closeness centrality)

Function	Influence of v	Cost
Ran.	1	$O(1)$
Deg.	$d(v)$	$O(\kappa)$
Vol.	$\sum_{w \in N_h(v)} d(w)$	$O(\kappa^{(h+1)})$
Wei.	$\sum_{w \in N_h(v)} d(w) \cdot (1 - c(w)) \cdot (1/2^{\delta(v,w)})$	$O(\kappa^{(h+1)})$

Neighbours Selection Strategies

- Set 4 selection strategies based on local connectivity pattern such as degree and clustering coefficient
  - Random selection
  - High degree selection
  - High volume selection (by Wehmuth and Ziviani) selection
  - High weighted-volume selection (a good approximation of closeness centrality)

Function	
Ran.	
Deg.	
Vol.	$\sum$
Wei.	$\sum_{w\in N_h(v)} d$



Datasets for Simulation

- We test the four real-world network datasets:

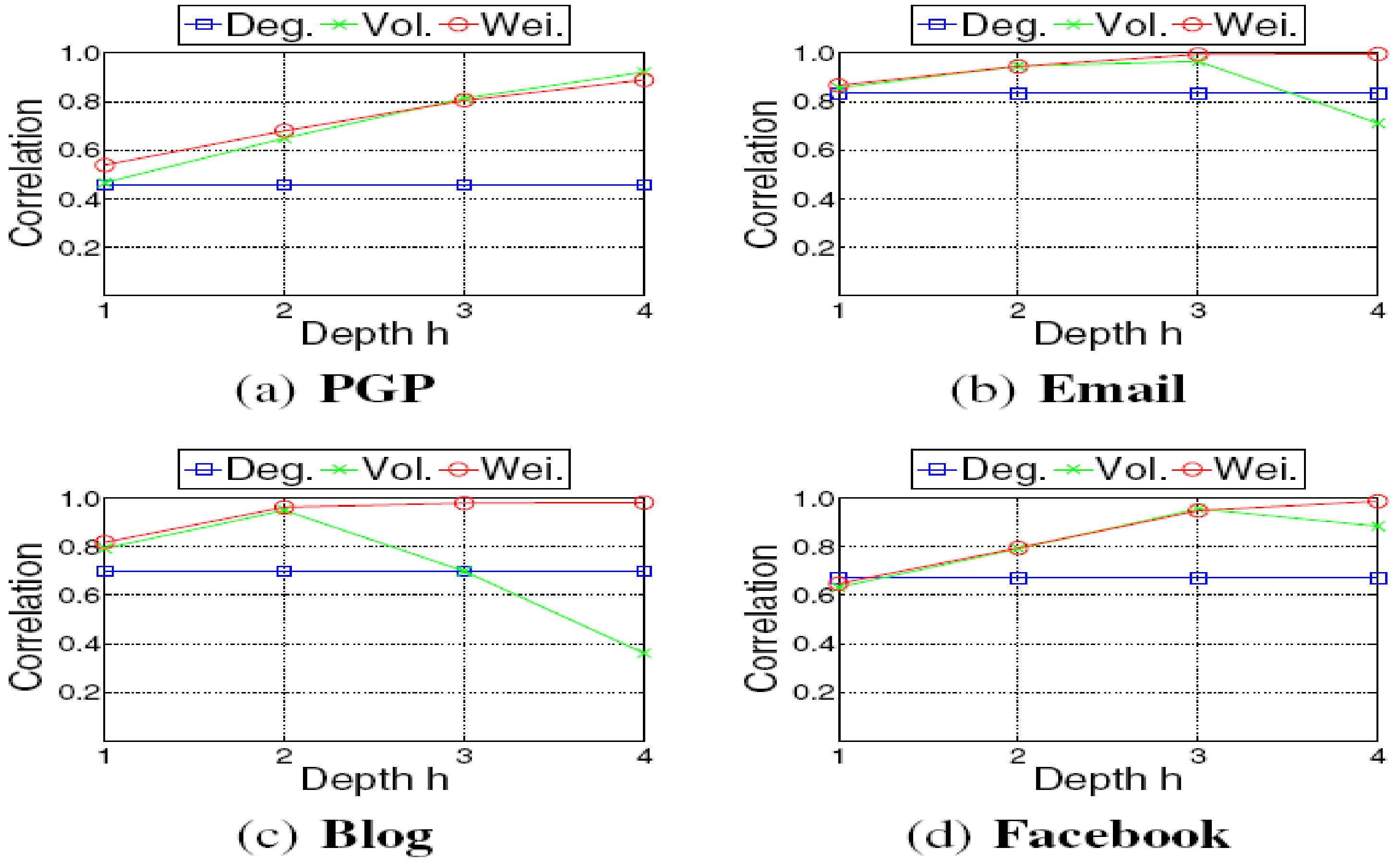
Network	$\|V\|$	$\|E\|$	κ	$\mathcal{C}$	$\mathcal{D}$
PGP [6]	10,680	24,316	4.55	1	24
Email [7]	1,134	5,453	9.62	1	8
Blog [8]	1,224	16,718	27.32	2	inf
Facebook	26,701	251,249	18.82	1	15

- Simulation: random selection of origin x 1000
- Used constant probability *λ at node*

k: average degree
C: number of connected components
D: network diameter

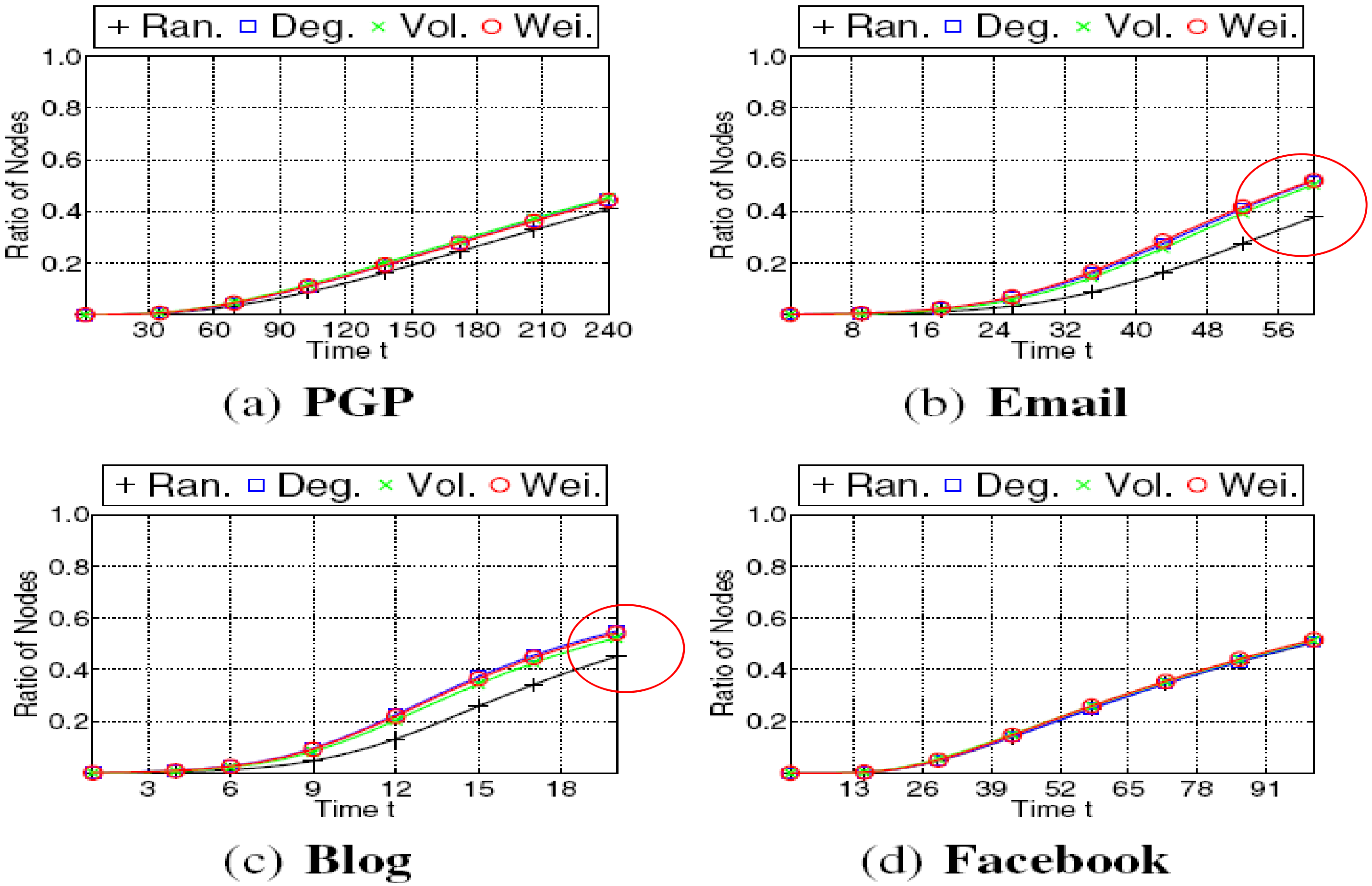
Correlation to Closeness Centrality

- Pearson correlation coefficients between node property and closeness centrality



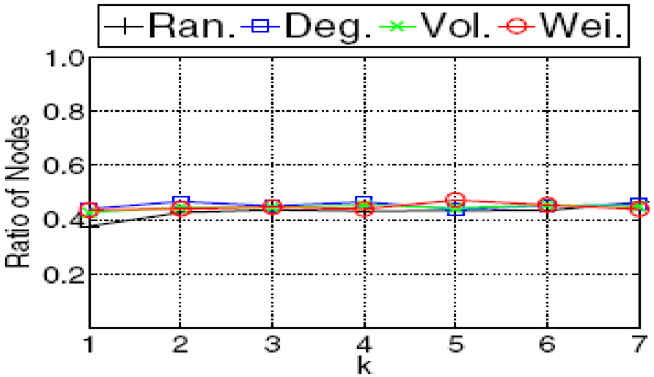
Simulation (use IC model)

Ratio of average # of activated nodes to total # of nodes over time *t* – *weight h=3*

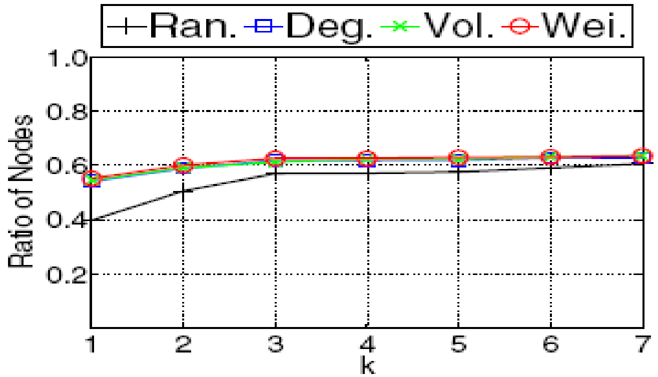


Effect of Size of K – Long Term

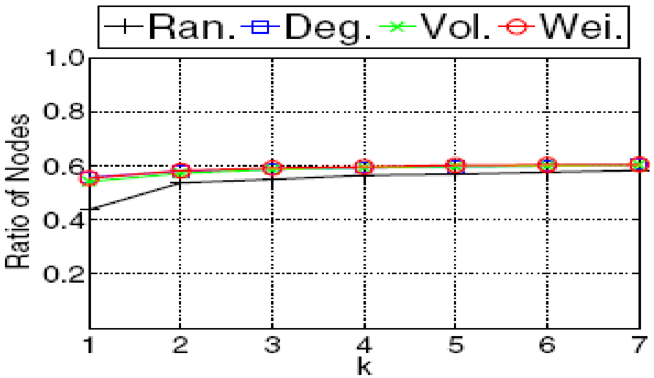
Ratio of average # of activated nodes to total # of nodes with # of initial activated neighbours *k*



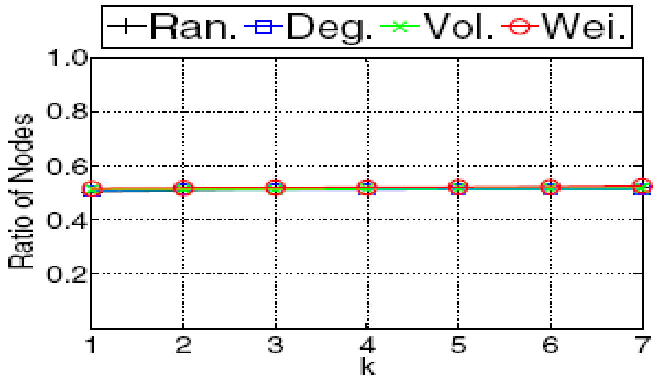
(a) **PGP**



(b) **Email**



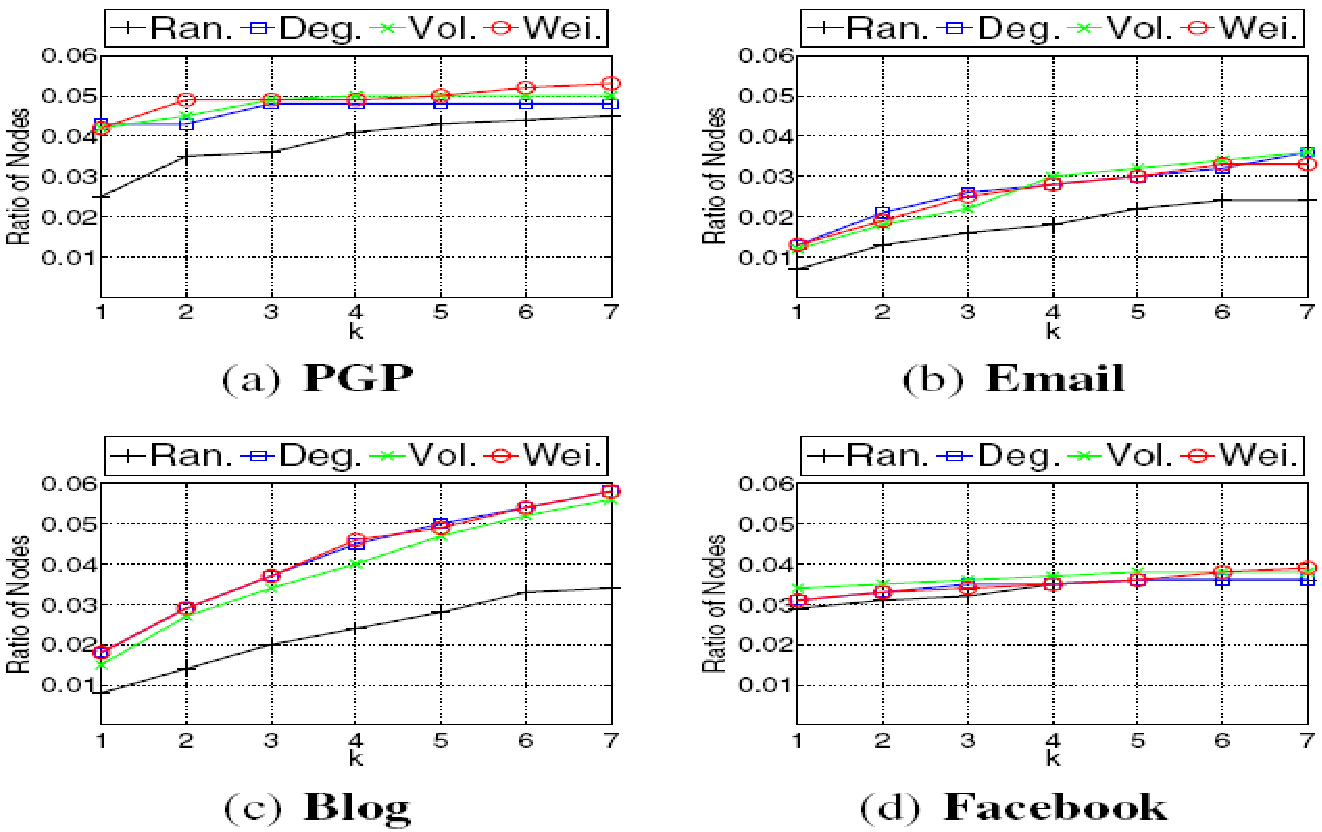
(c) **Blog**



(d) **Facebook**

Impact of Size of K – Short Term

Changes in ratio of average # of activated nodes to total # of nodes with # of initial activated neighbours k (1/4 of full timeline)



(a) **PGP**

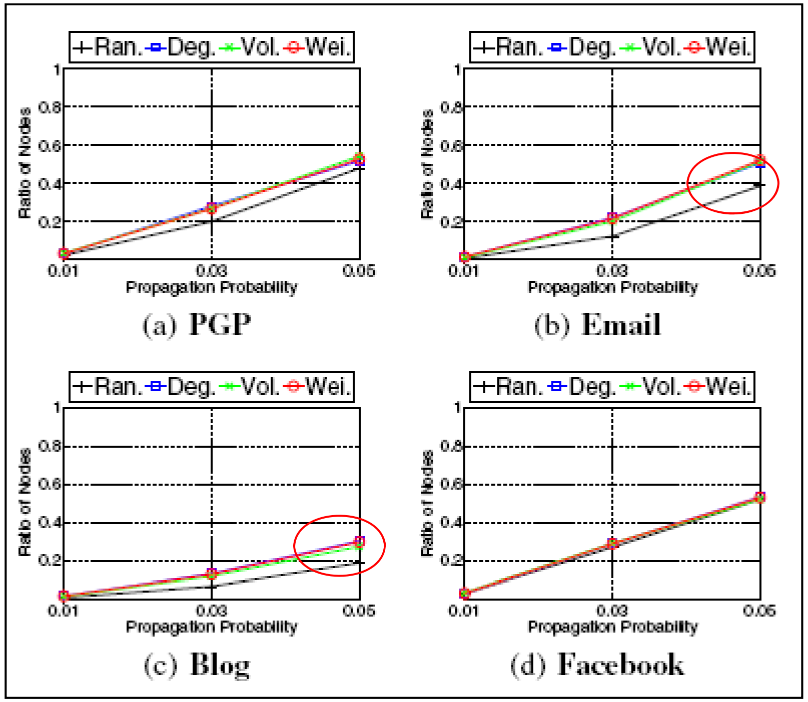
(b) **Email**

(c) **Blog**

(d) **Facebook**

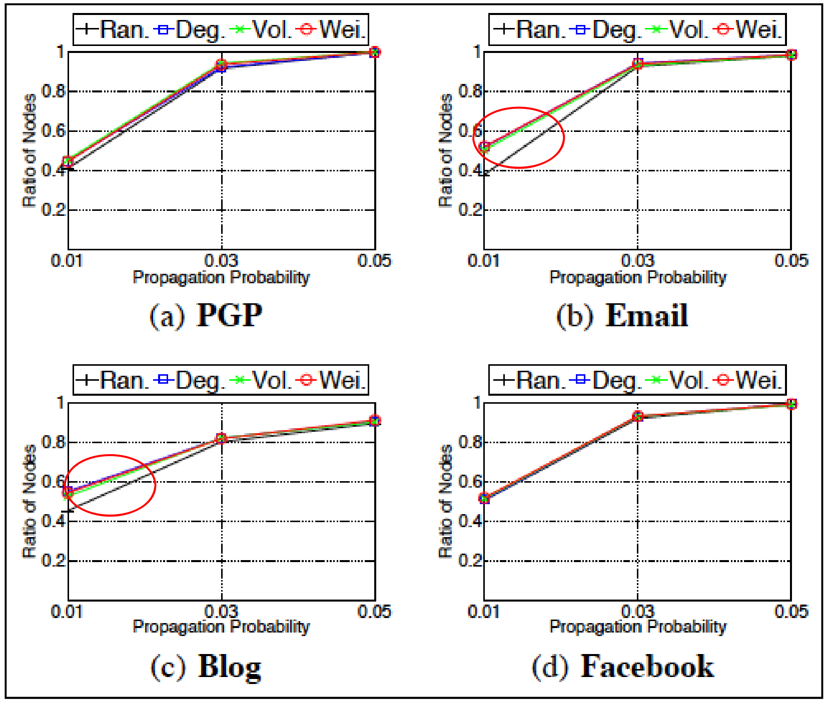
Propagation Probability

The ratio of average # of activated nodes to total # of nodes with influence probability λ (k=1)



Short Term

increase gap



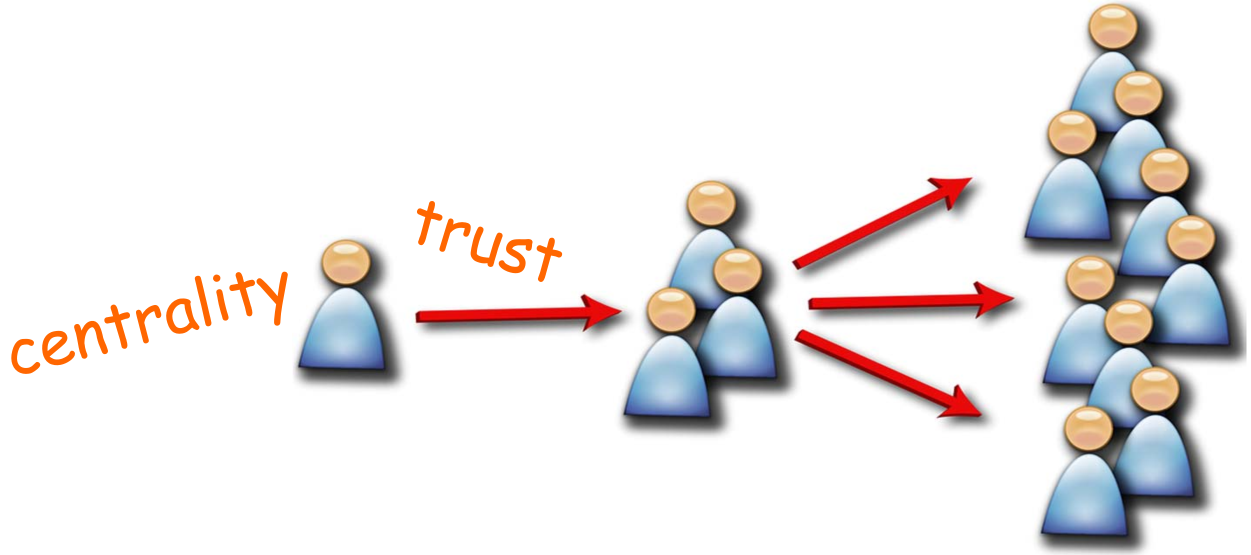
Long Term

decrease gap

INS: Outlook

- Introduced INS problem: Select a node's neighbours to efficiently diffuse its information
- Empirically tested 4 selection strategies through intensive simulation with 4 real-world network topologies
  - Degree selection strategy for short-term propagation
    Random selection strategy for long-term propagation
  - Volume and Weighted produce similar results to those obtained by degree - we recommend using degree, which less costs
  - Speed of information diffusion is dramatically improved with higher probability λ

Propagation Decision at Node



- Individual influence probability λ

  → Need to model decision making mechanism at each node

  → Towards psychological behaviour embedded model

Outline

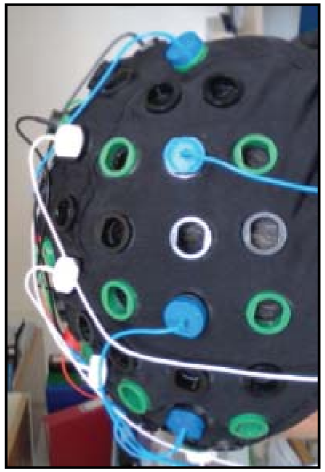
- Influence Maximisation Study: Selection of propagation nodes

- Decision at node: Social influence
  - EEG
  - Eye tracking

- Digital Epidemiology

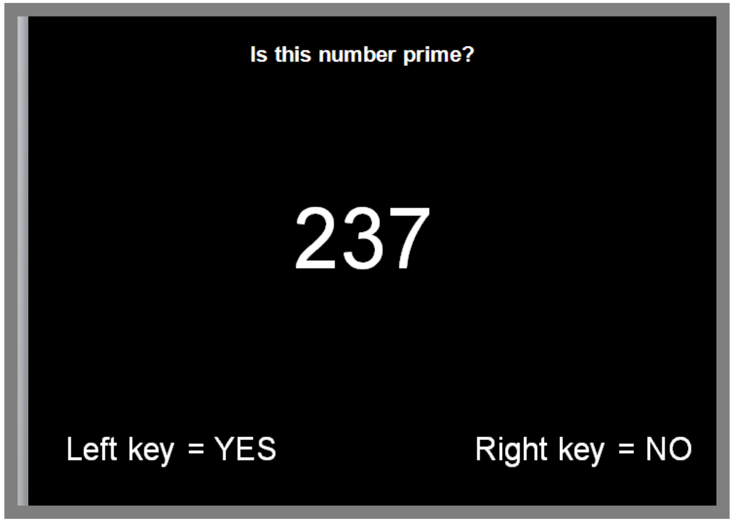
EEG System

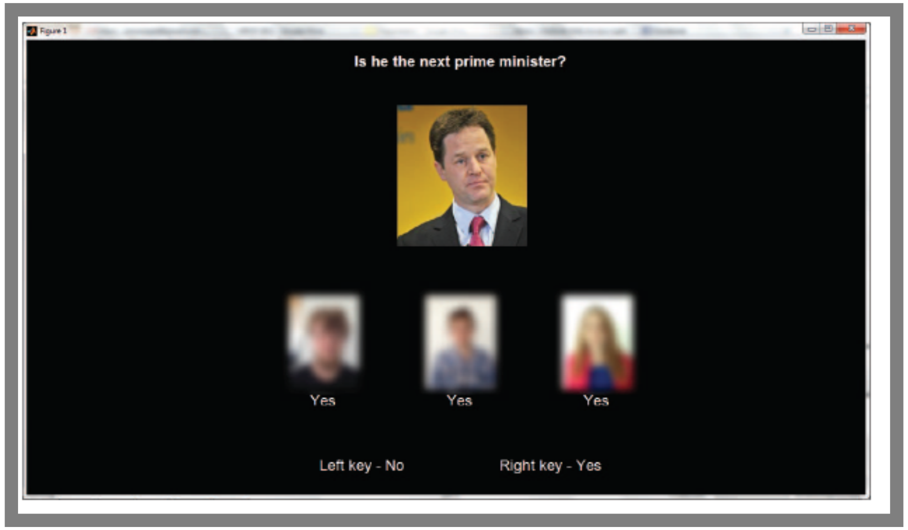
- Employ Social Neuroscience
- Electroencephalography (EEG): measures electrical activity from firing of neuron populations
- Signal propagation patterns among channels for understanding decision making mechanism



Experiment Setup

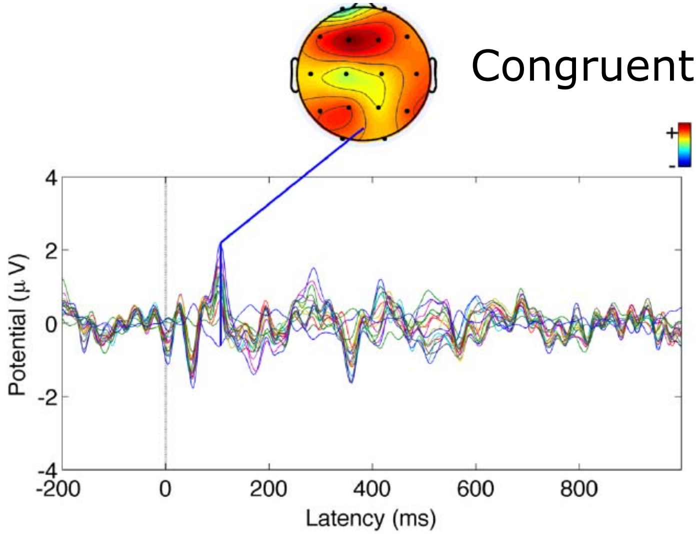
- 500 Questions
  - Arithmetic, quizzes, and recognition/preference of images/photos
  - Blind stage and Friends stage (plus manipulation stage)
- 20 Participants
- Together with Pre-Experiment Survey and OSN info



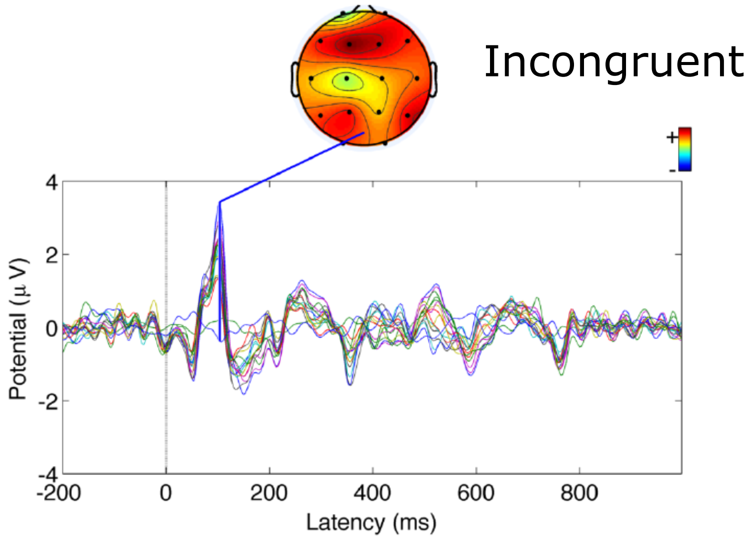


EEG Response Preliminary Analysis

- Voltage topography of the EEG response over scalp, sampled at the 16 locations



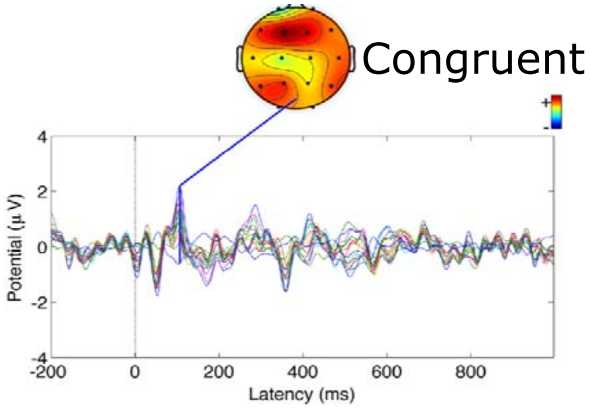
Average EEG response: Friend Stage answer was Congruent (Agreement)

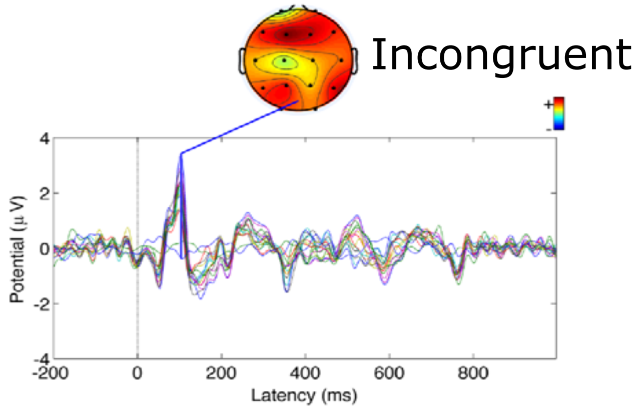


Average EEG response: Friend Stage answer was Incongruent (Disagreed)

- Millisecond-scale temporal dynamics of the brain response following this instant track cognitive processing of congruency of friend's answer

EEG Response Preliminary Analysis





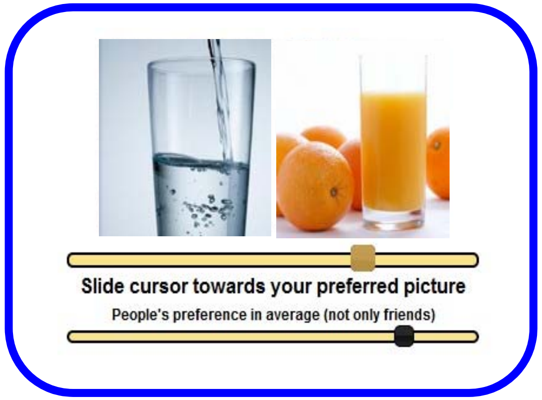
- Preliminary but key difference appears to be in the amplitude of brain response in congruent vs. incongruent scenarios
- Participants engaged in increased amounts of cognitive processing on receiving negative feedback
- Brains generate heightened responses indicating that this Social/Interpersonal Conflict is processed with increased valence and cognitive/emotional import

Eye Tracking

- Photo rating study by understanding unconscious behaviour
- Tobi: automatically records web page coordination

- Is social desirability bias higher when you know rating of a specific person, average rating, or rating is known to friends?

Blind Phase

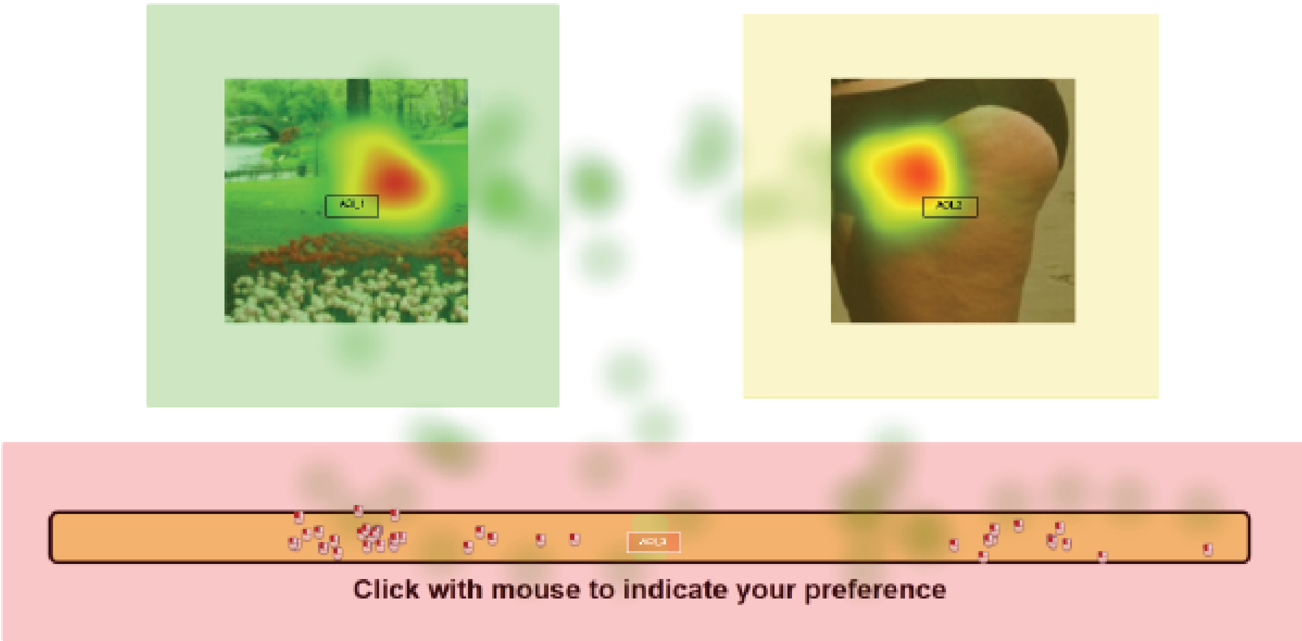


Average of Others Phase

Friends Phase

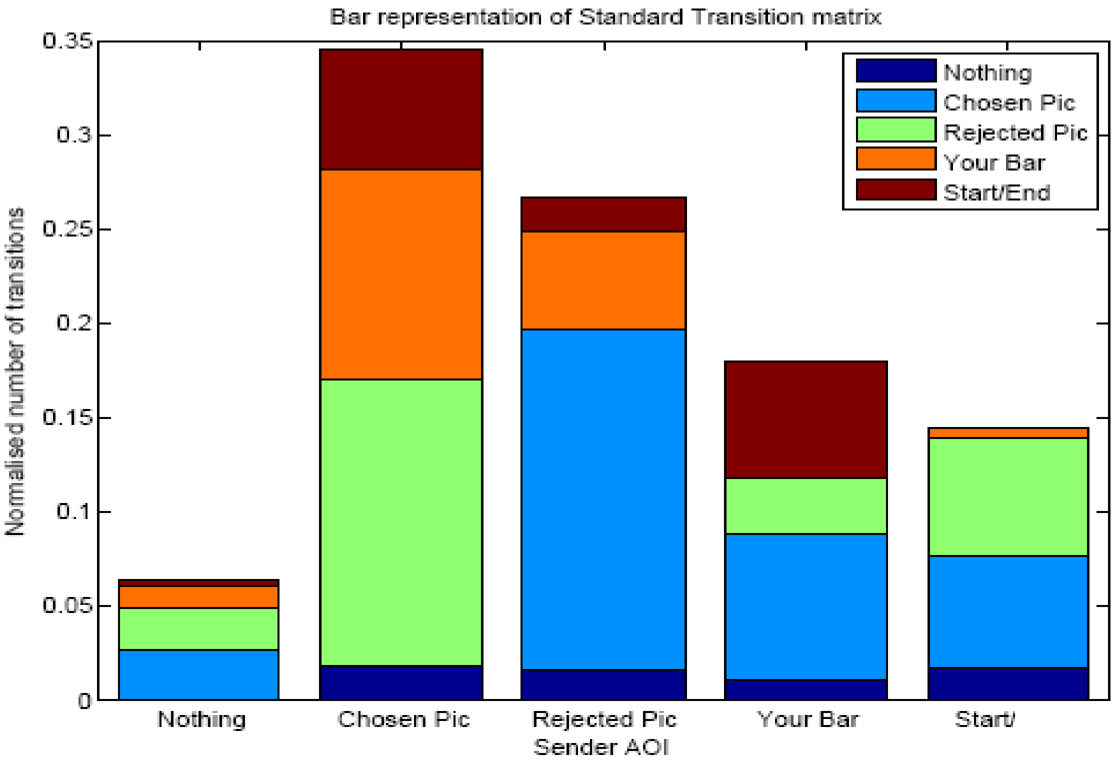
Area of Interest Heat Map

- Spatial Distribution of Eye Gaze over 2 Photos



Area of Interest Transition

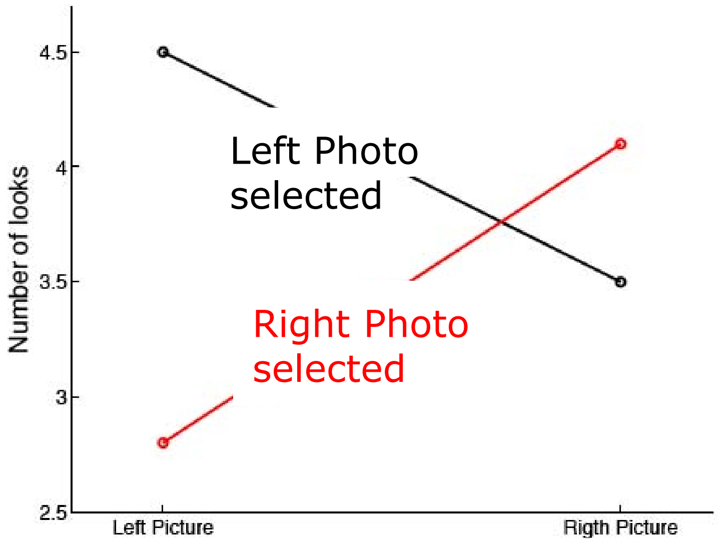
- Chosen picture is popular AOI
- Decision to finish rating is based on the picture you prefer and not the rejected picture



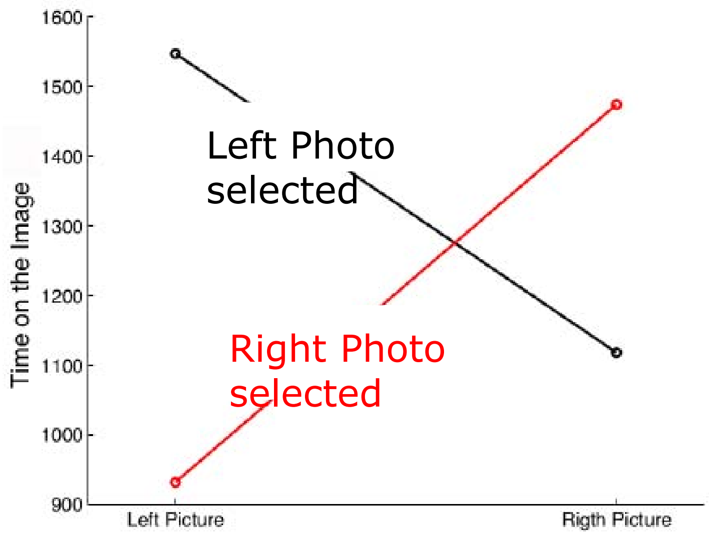
Bar shows NEXT AOI after X-axis labeled AOI

Indicator of Liking

- Choice of photo can be predicted by how much time observer looks at a particular picture (and number of saccades)



Number of Looks



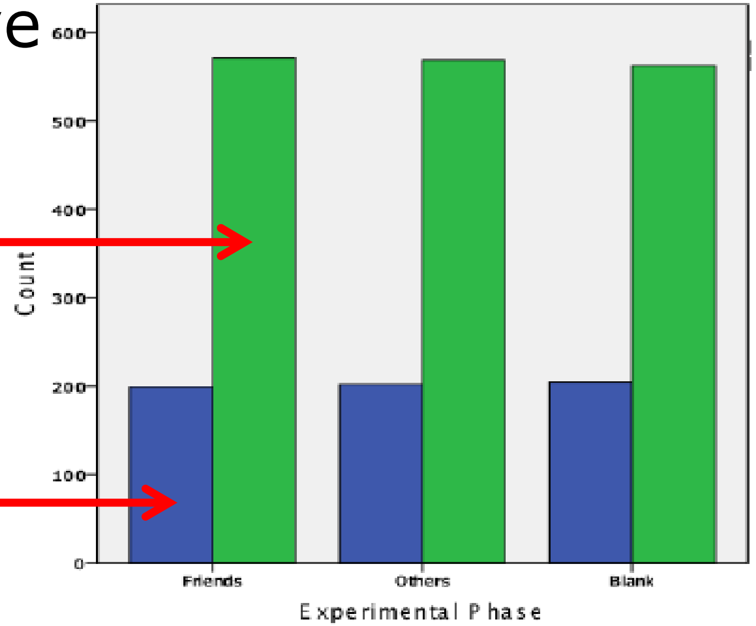
Duration of Looks

Consistent Choices in Different Phases

- Consistent Choices: independent from experimental conditions (i.e. Blind, Average, and Friend phases)
- In-Out Group Bias (based on psychological theory) shows personal preference is preserved no matter the conditions are
- However, underlying cognitive processing is different

Same Choice w/ others within the Phase

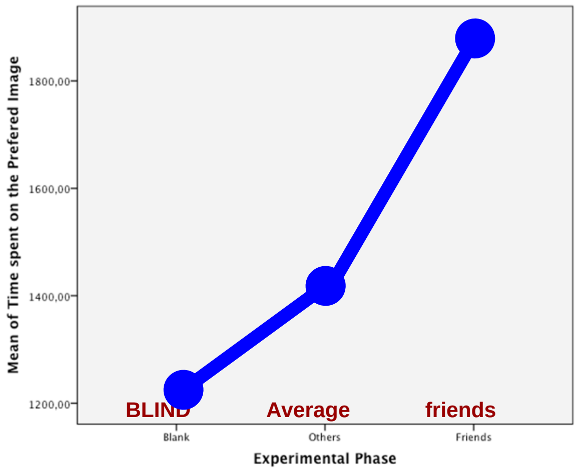
Opposite Choice from others within the Phase



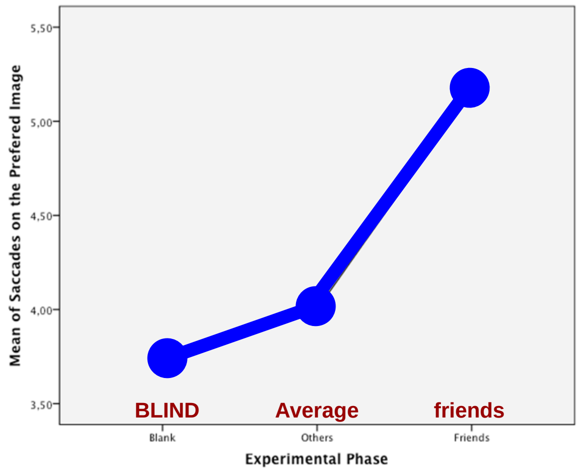
Consistent Choices in Different Phases

- When more information is given, participants spend more time to make final decision by looking at preferred photo, but choices are consistent

Time Spent on Preferred Photo



Saccades on Preferred Photo



EEG and Eye Tracking: Outlook

- Employ Neuropsychology (Eye Tracking) and Social Neuroscience (EEG) for understanding decision making process
- Helps to understand different cognitive processes (mental schemes)
- Eye Tracking/EEG can be used to equip and to train an artificial systems

Outline

- Influence Maximisation Study: Selection of propagation nodes
- Decision at node: Social influence
    - EEG
    - Eye tracking
- Digital Epidemiology

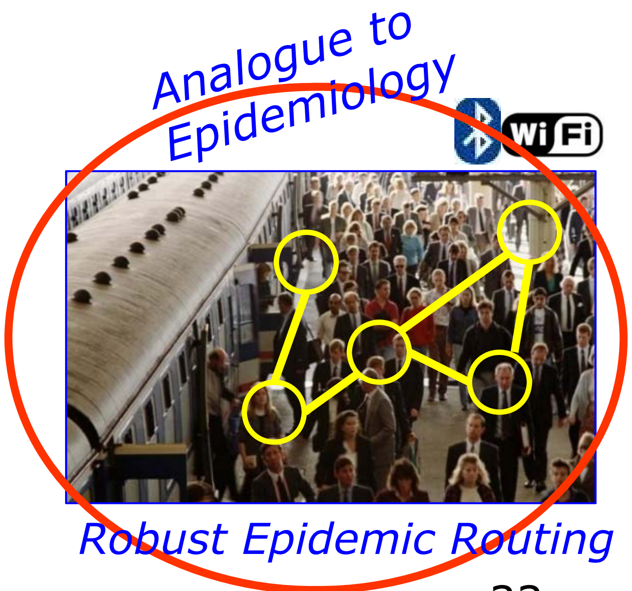
Spread of Infectious Diseases

- Thread to public health: e.g., 

SARS, AIDS, Ebola
- 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



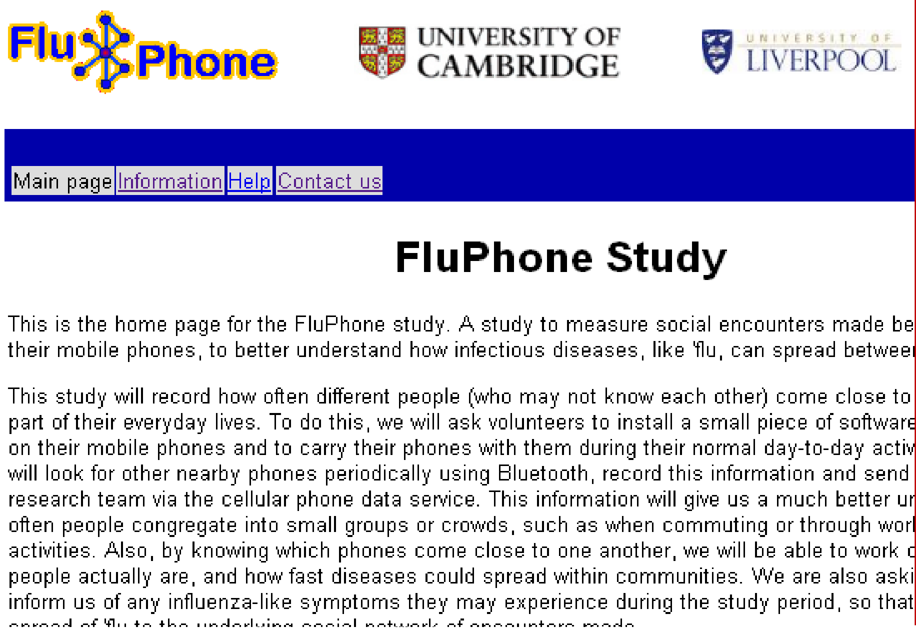
FluPhone Project

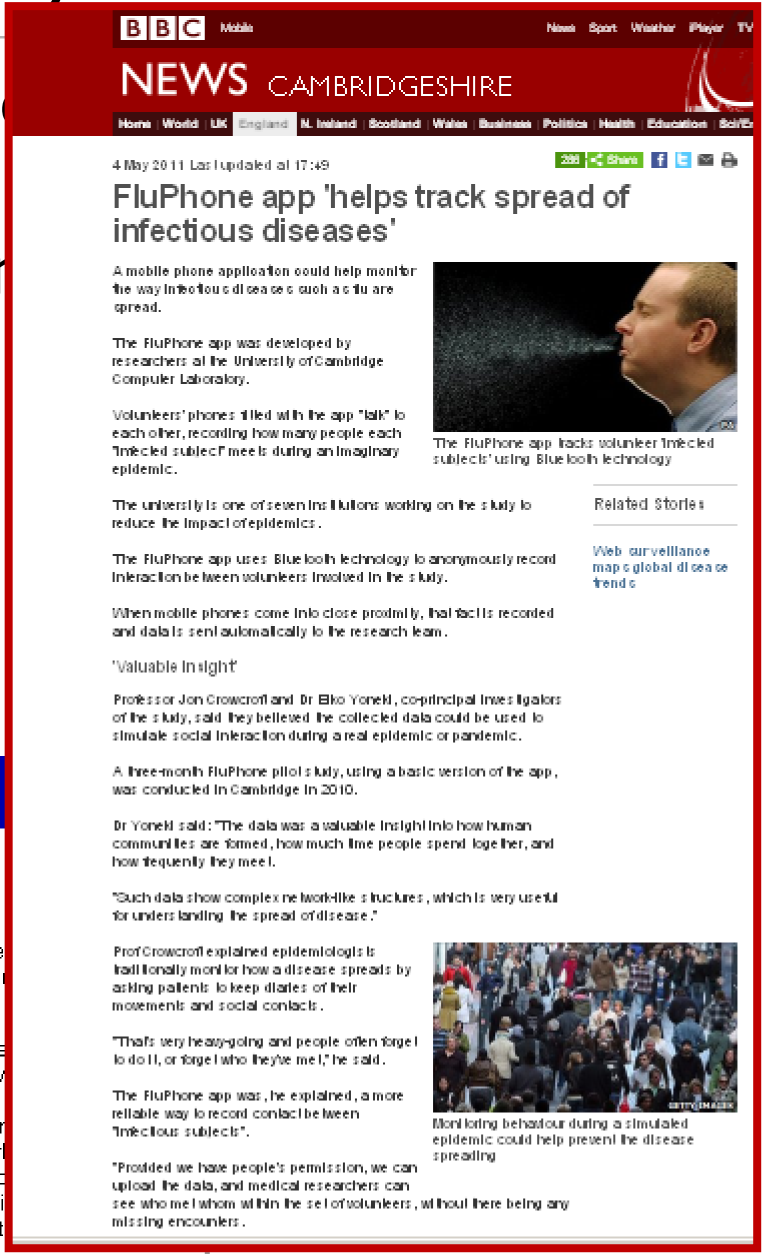
- Scan Bluetooth devices every 2 minutes
- Ask symptoms

FluPhone Project

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

https://www.fluphone





BBC Mobile News Sport Weather iPlayer TV

NEWS CAMBRIDGESHIRE

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/En

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 fitted 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.

Related Stories

Web surveillance maps global disease trends

'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.

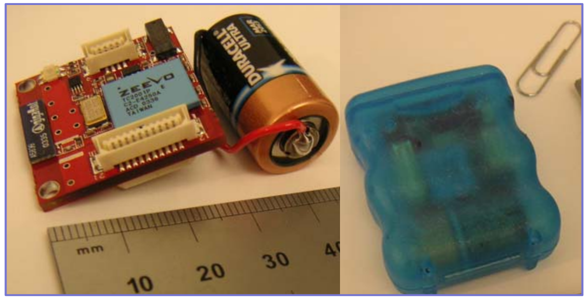
"That's very heavy-going and people often forget to do it, or forget who they've met," he said.

The FluPhone app was, he explained, a more reliable way to record contact between "infectious subjects".

Monitoring behaviour during a simulated epidemic could help prevent the disease spreading

"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.

Electronic Proximity Sensing



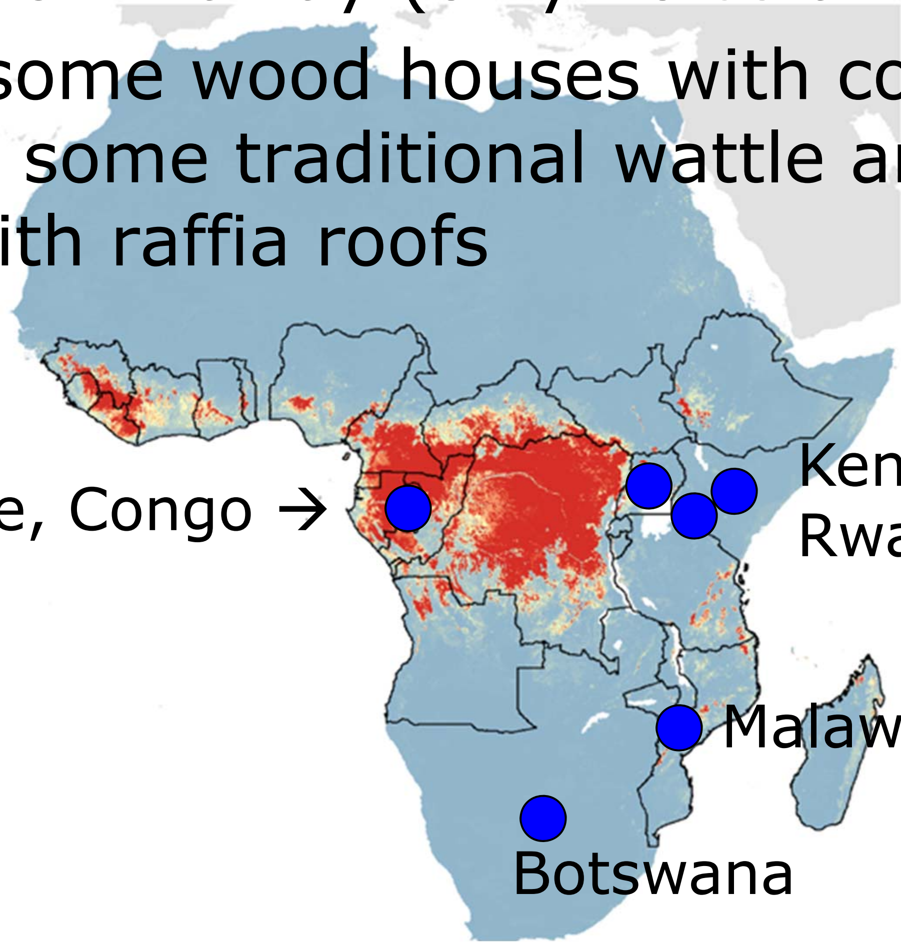
- Sensors
  - Bluetooth Intel iMote
  - Zigbee
- RFID Tags

  - UHF Tag Alien ALN-9640 - "Squiggle®" Inlay
  - OpenBeacon active RFID Tag

- Mobile Phones
  - FluPhone Application
  - GPS, Google latitude
- GPS Logger
- Online Social Networks
  - Twitter, Facebook, Foursquare…

Lengue Lengue, Congo

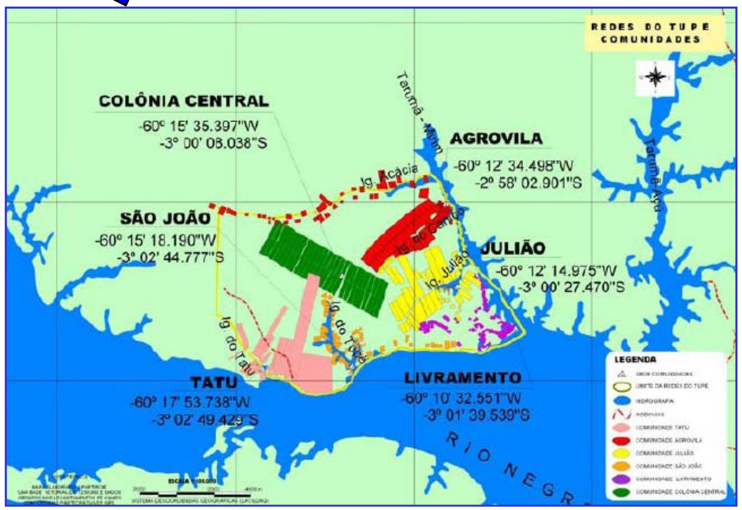
- Village of about 100 people with no power
- Access to cell phone network from town of ~2000 200km away (only voice and text)
- Houses: some wood houses with corrugated roofs and some traditional wattle and daub houses with raffia roofs



South America – Manaus in Brazil

- Several disconnected communities along river
- Deploy delay tolerant communication

Tupé Sustainable Development Reserve



UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Sensing Platform in Remote Region

- Build a platform for sensing and collecting data in developing countries
  - e.g. OpenBeacon Active RFID tags based contact network data collection
  - Build a standalone network for data collection and communication using Raspberry Pi → RasPiNET
  - Inexpensive network setting
  - Support streaming model

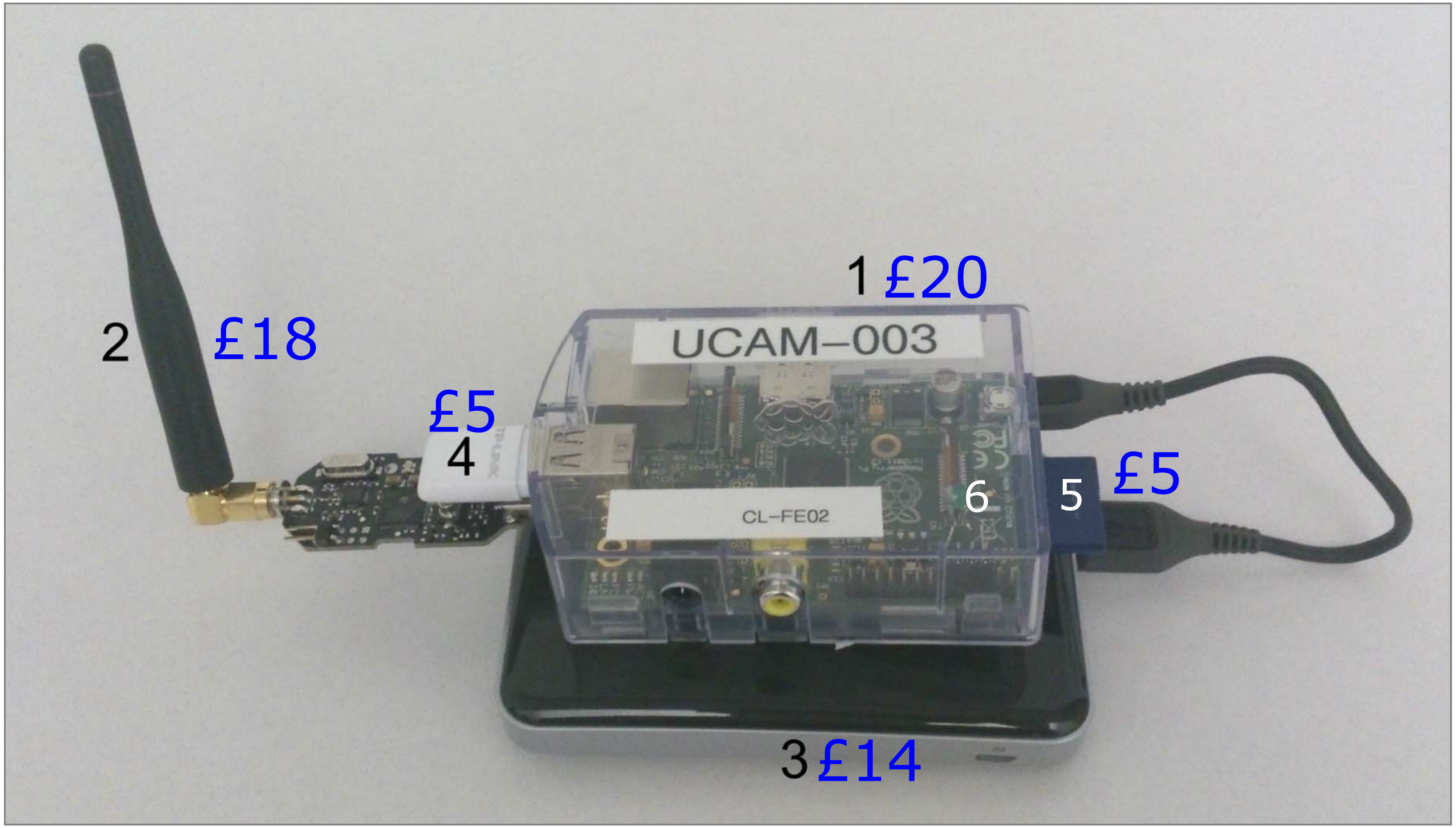
OpenBeacon RFID Tags

- 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

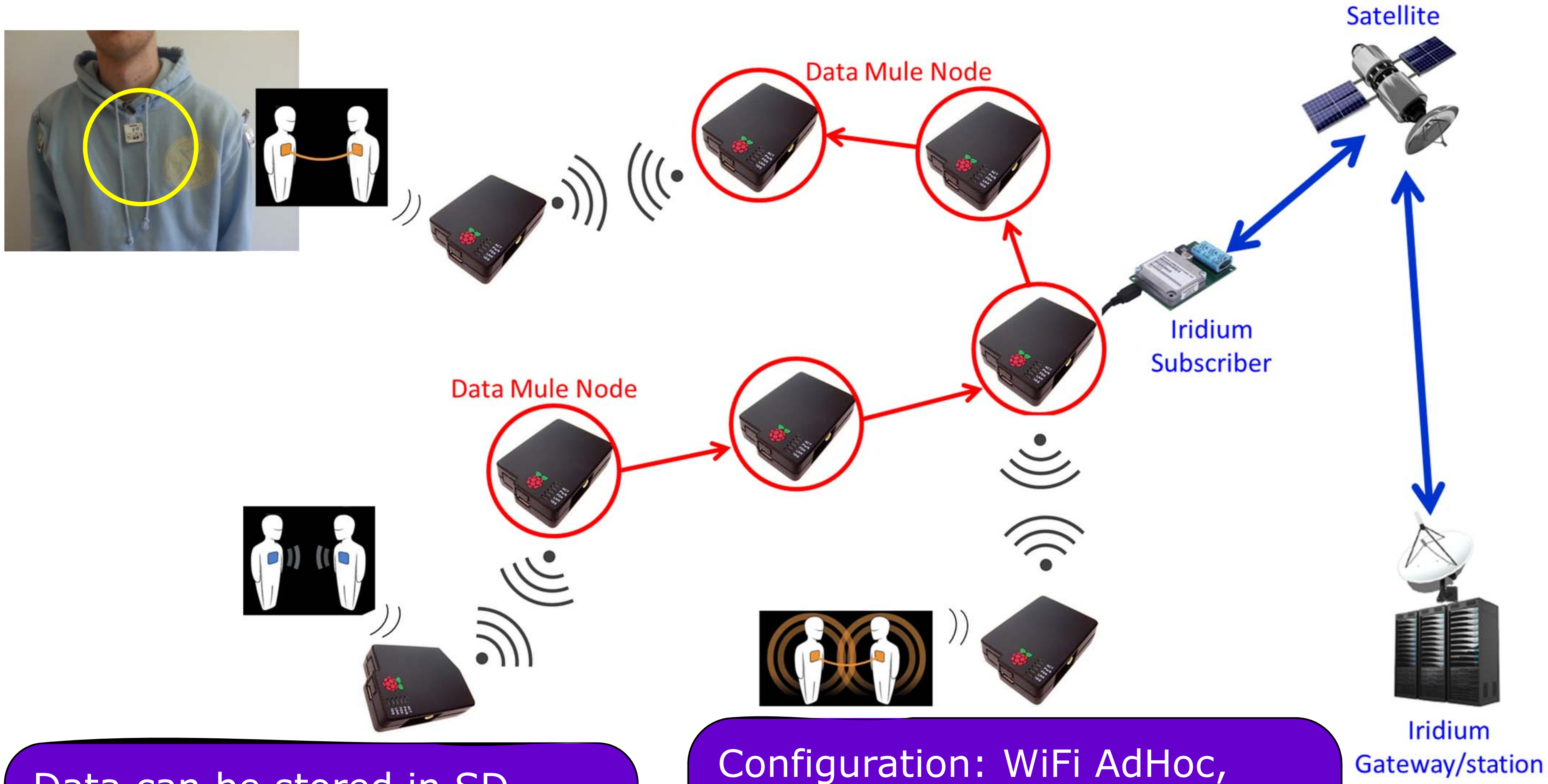
OpenBeacon Ethernet EasyReader

Raspberry Pi OpenBeacon Reader



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

Raspberry Pi based Sensing Platform



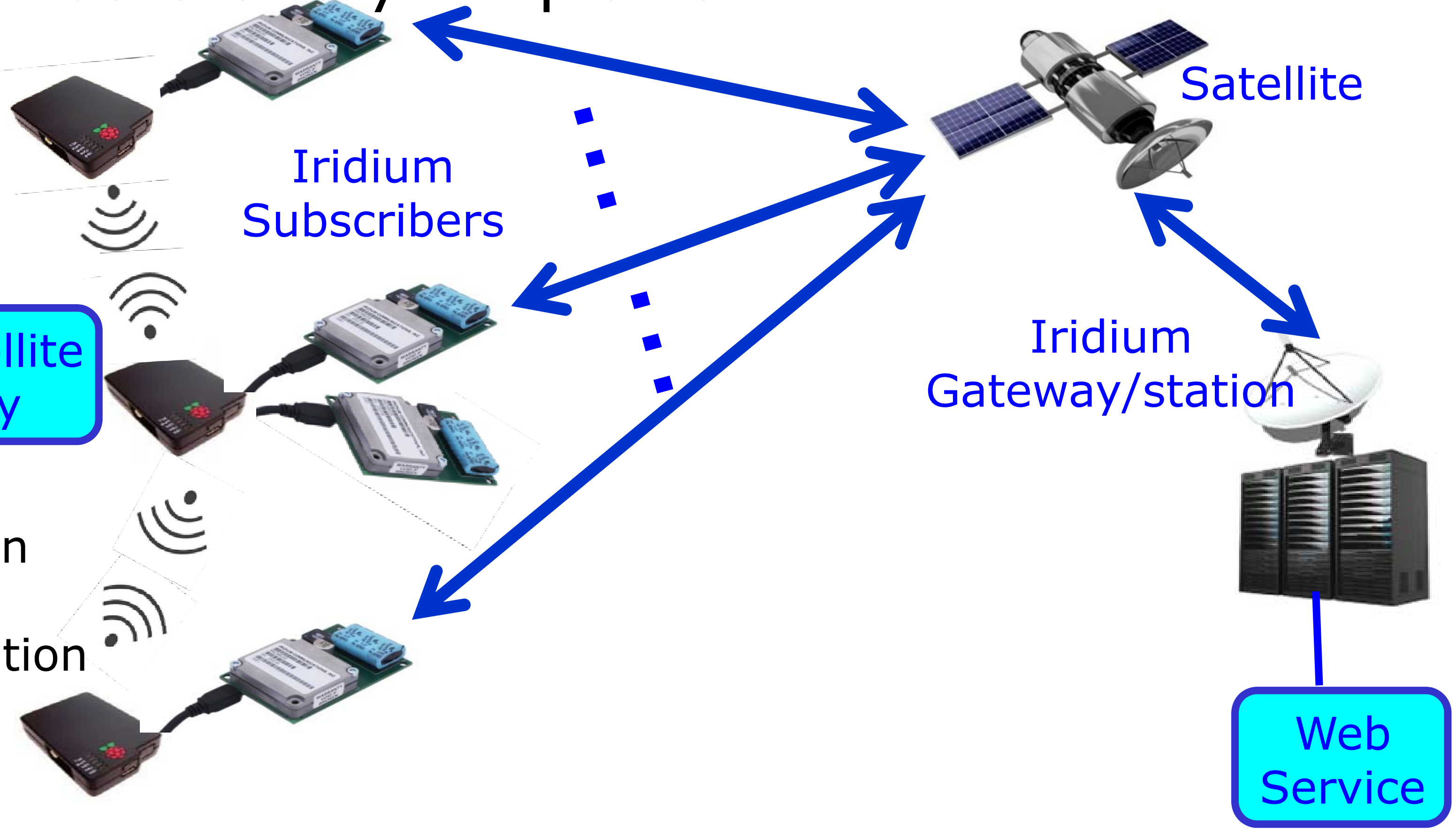
Data can be stored in SD card, transmitted to Data Mule node, or use of WiFi AdHoc mode transmission to Gateway

Configuration: WiFi AdHoc, Software Access Point, WiFi Direct, DTN Data mule

Single/Multiple Satellite Gateway nodes

RasPi Satellite Gateway

- Build stream processing paradigm
- RasPi data analytics platform



RasPi Satellite Gateway

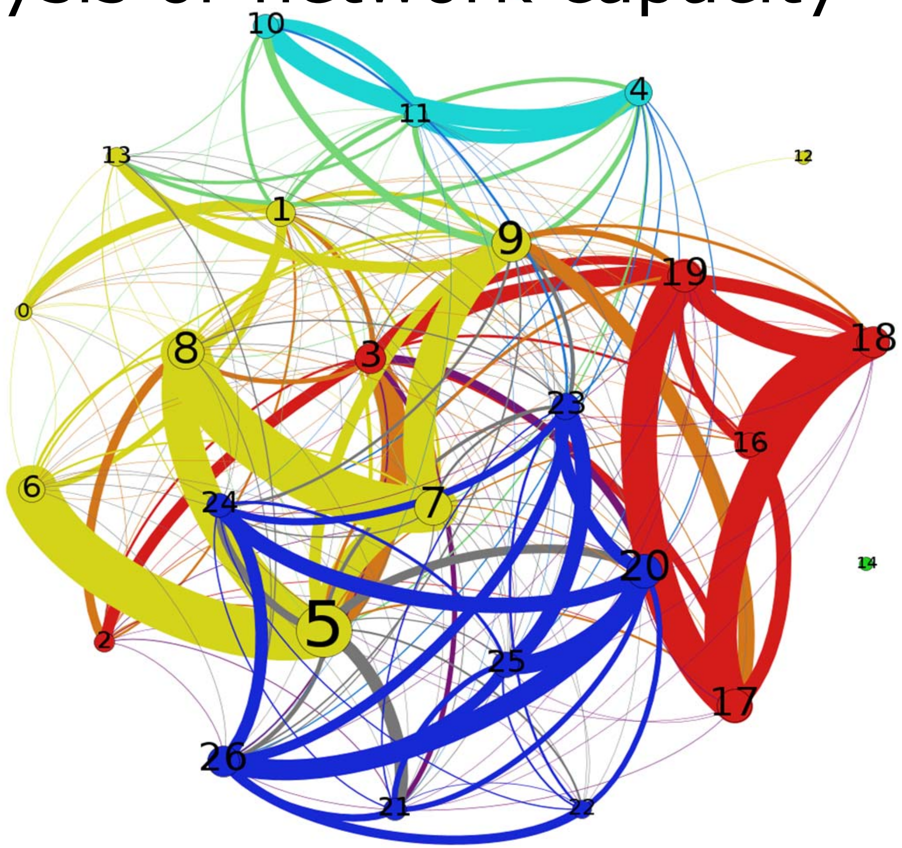
- Data filter, aggregation
- Analysis
- Fragmentation
- Multicast support

Web Service:

- Combine split data

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)

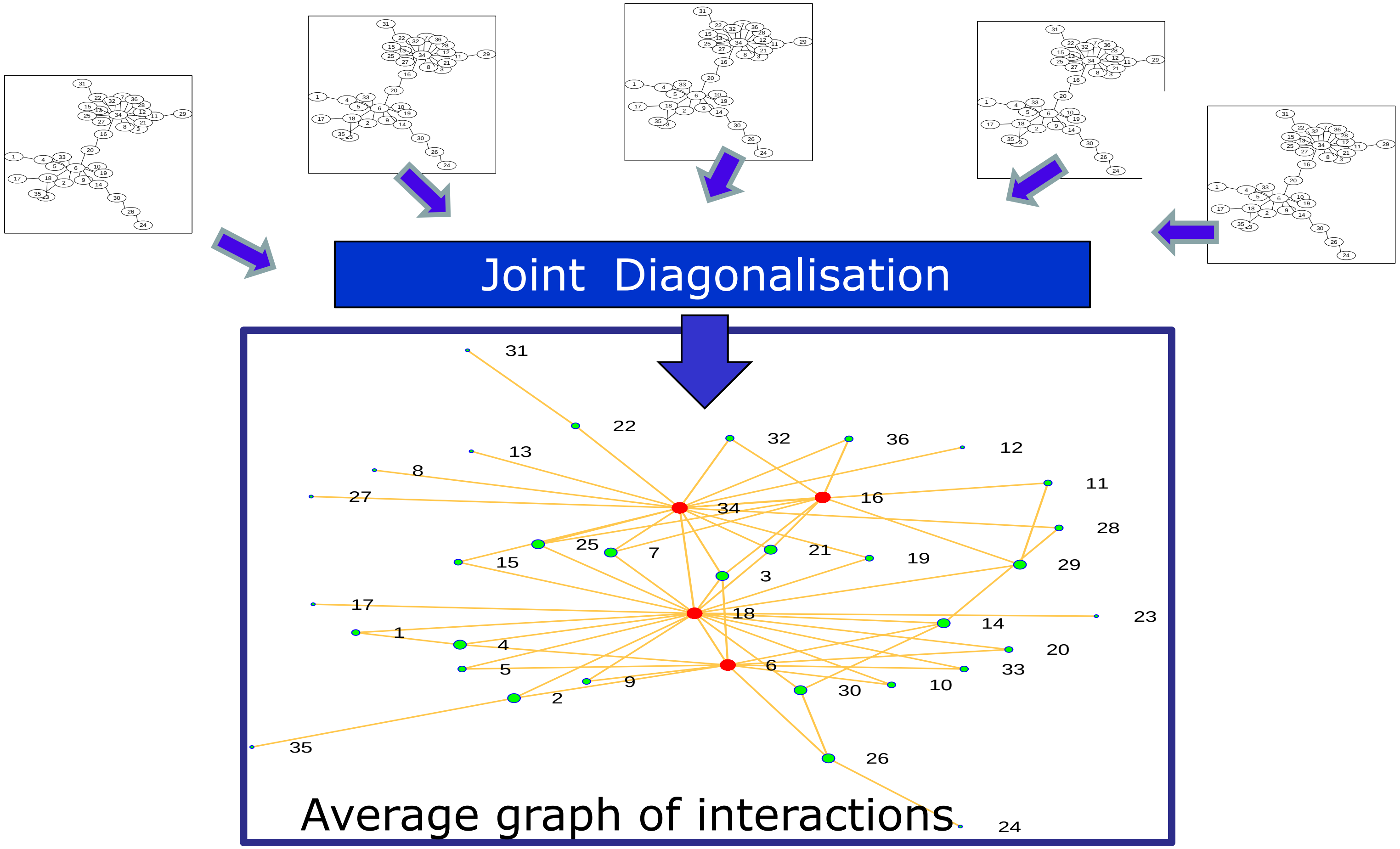


Extract Different Modes of Spread

- Typical approach: cluster nodes to build single network or multiple networks within the sliced time windows
  - 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

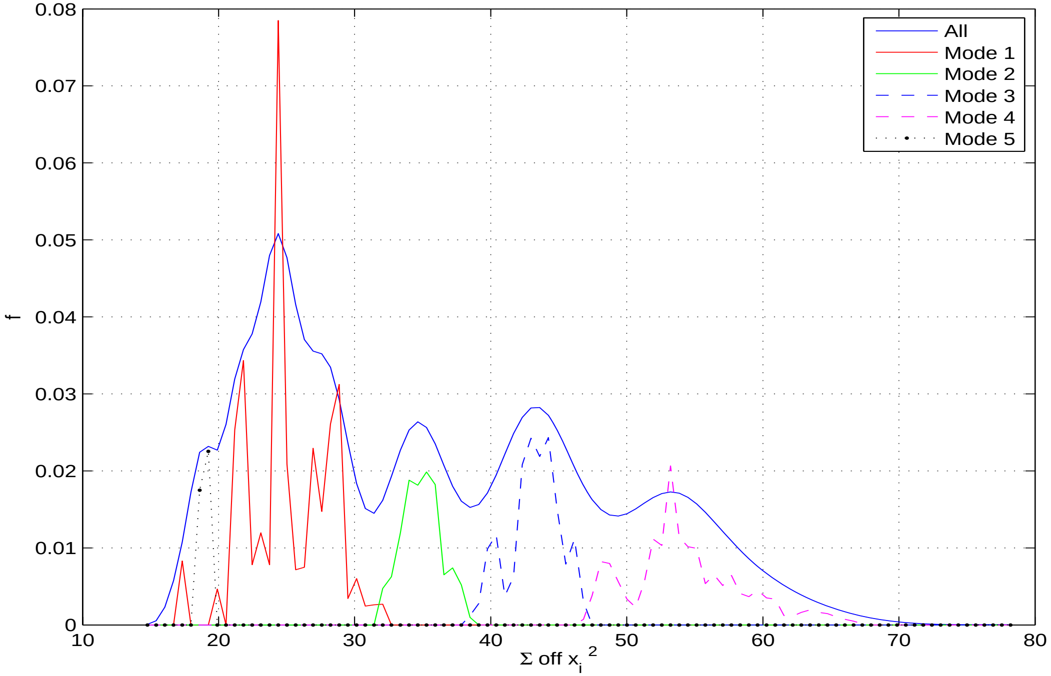
Joint Diagonalisation

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



Multiple Network Modes

- Define deviation from the average eigen-space as the sum of off-diagonal elements
- Use Gaussian mixture model for mode determination
- Distribution of deviation from average graph is multi-modal ➡ different behaviour of network

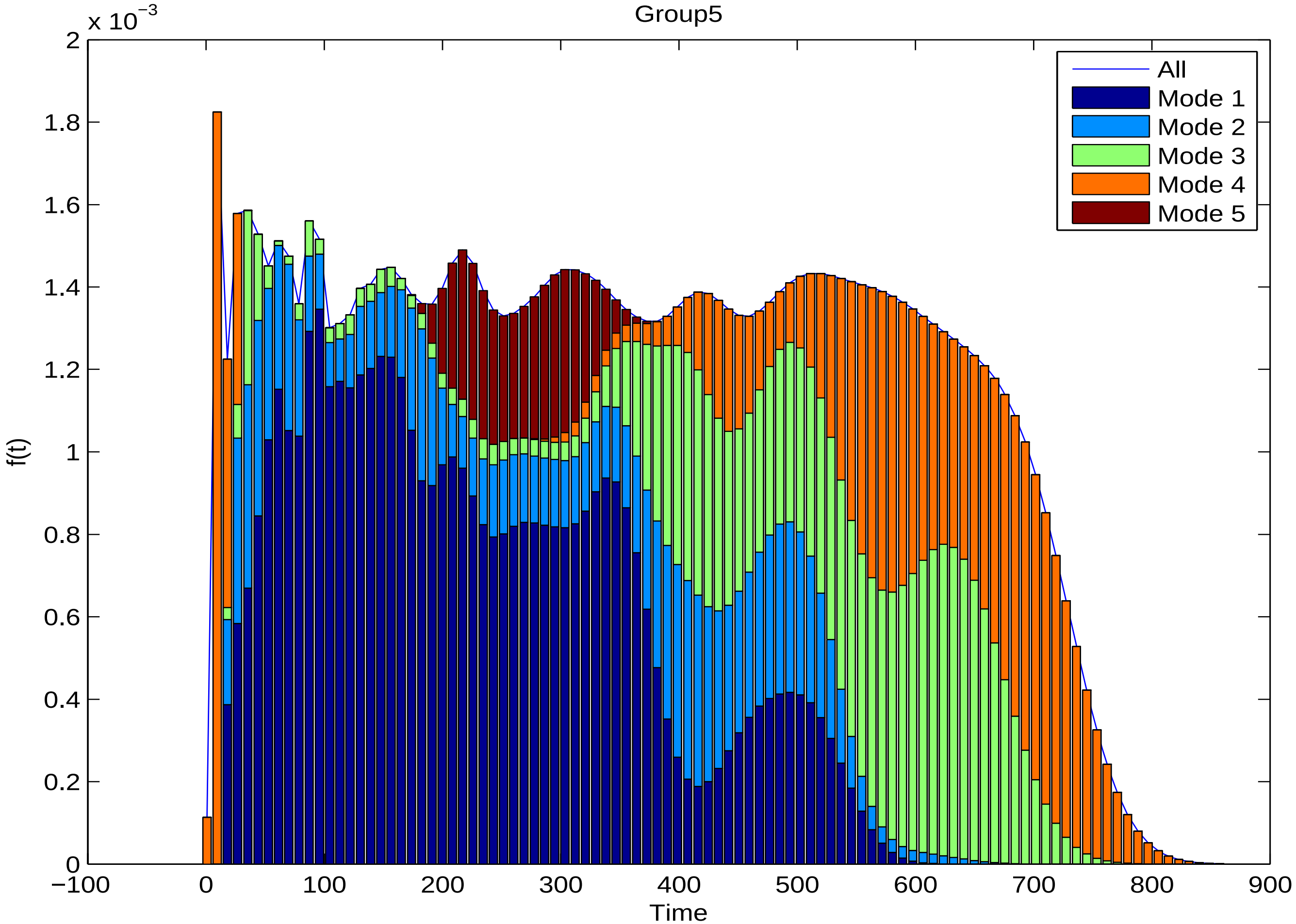


Distribution of deviation (Cambridge data)

Note: A random network shows only one mode

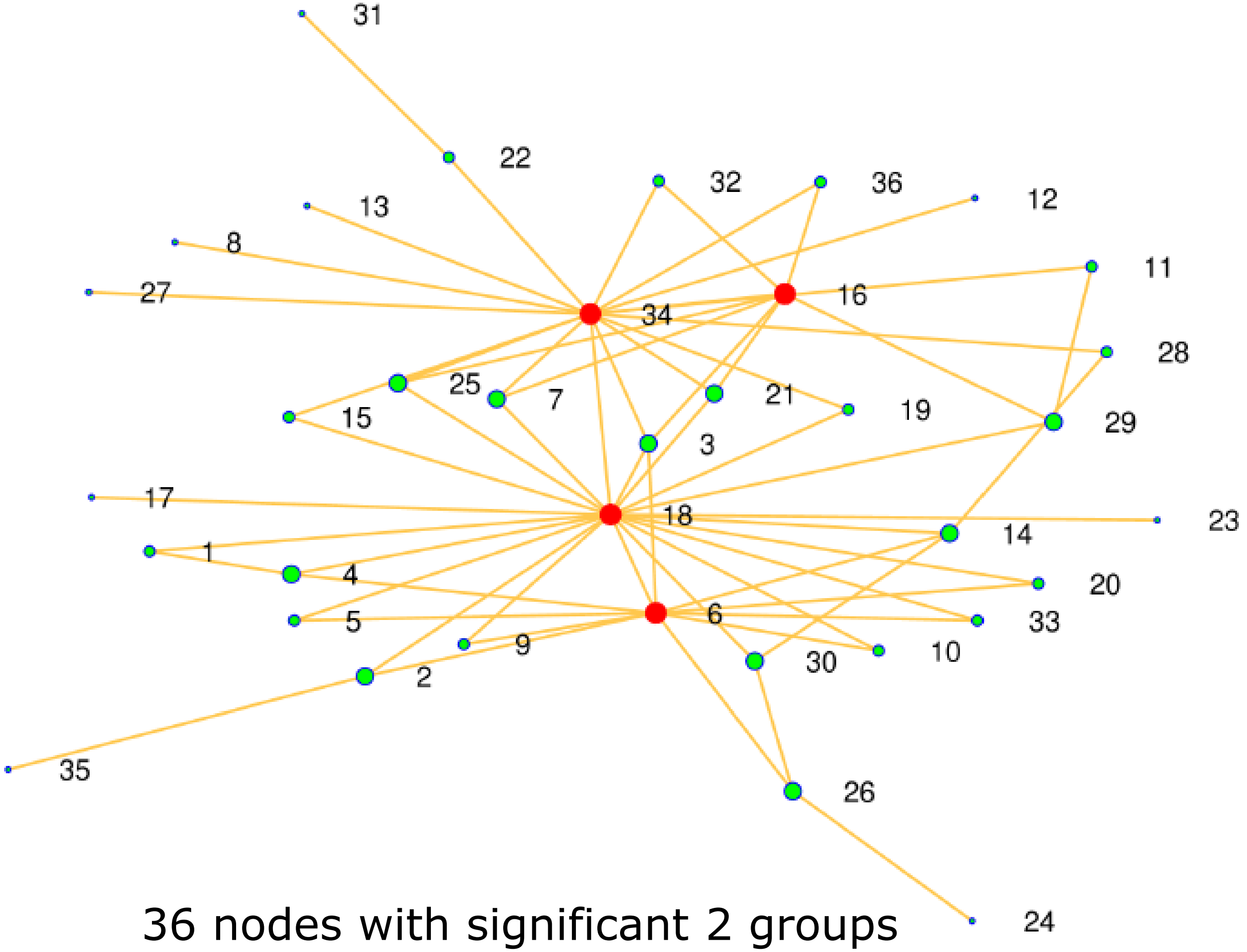
Extract Spread Modes

- Change of mode corresponds with state transition



Distribution of times by mode

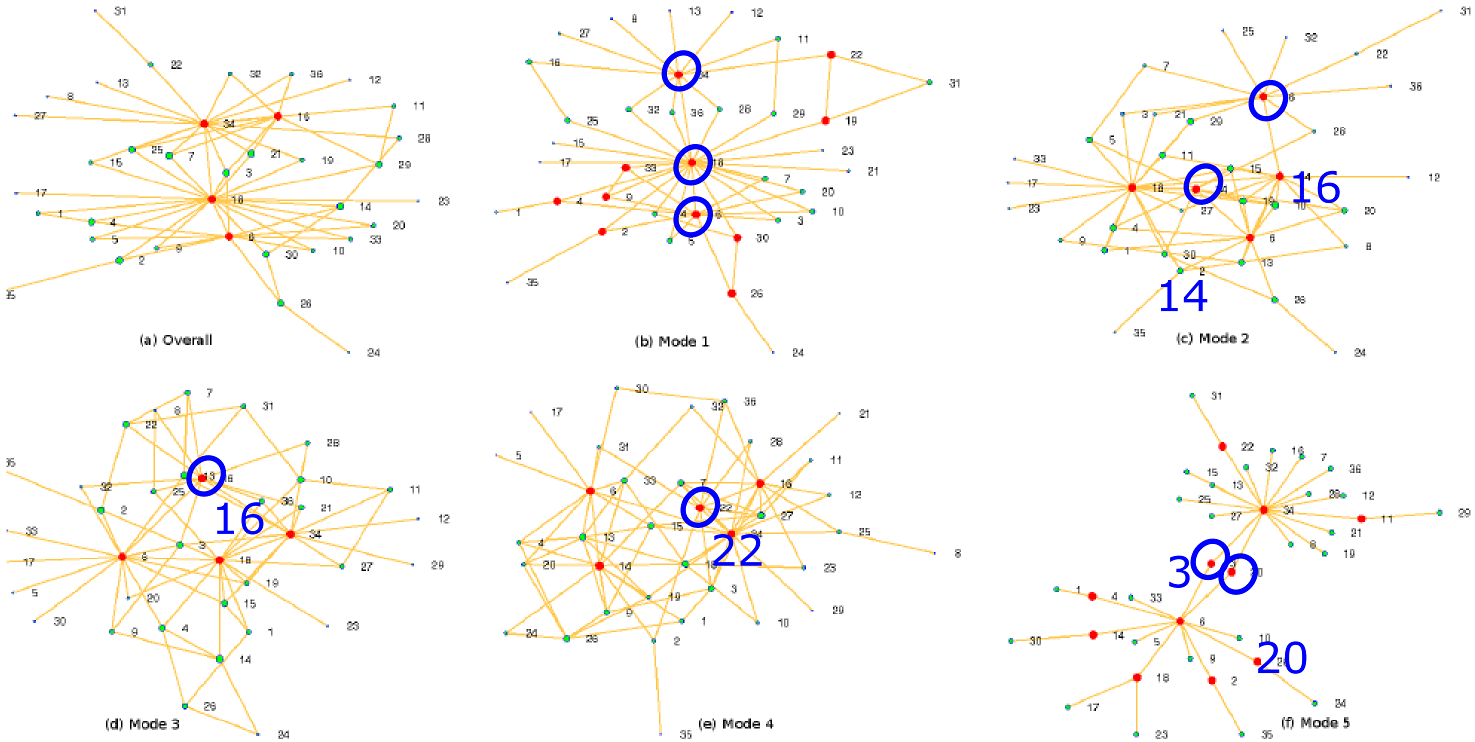
Average Graph of Interactions



36 nodes with significant 2 groups

Network Structure of Each Mode

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



(a) Overall (b) Mode 1 (c) Mode 2 (d) Mode 3 (e) Mode 4 (f) Mode 5

Questions?

References:

- Influential Neighbours Selection for Information Diffusion in Online Social Networks, ICCCN, 2012.
- A Study on the influential neighbors to maximize information diffusion in online social networks, Springer Computational Social networks, 2015.
- Cognitive dissonance and social influence effects on preference judgements: An eye tracking based system for their automatic assessment, Int. J. Human-Computer Studies, 2015.
- Centrality and Mode Detection in Dynamic Contact Graphs; a Joint Diagonalisation Approach, IEEE/ACM ASONAM, 2012.

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