UNIVERSITY OF CAMBRIDGE
Computer Laboratory

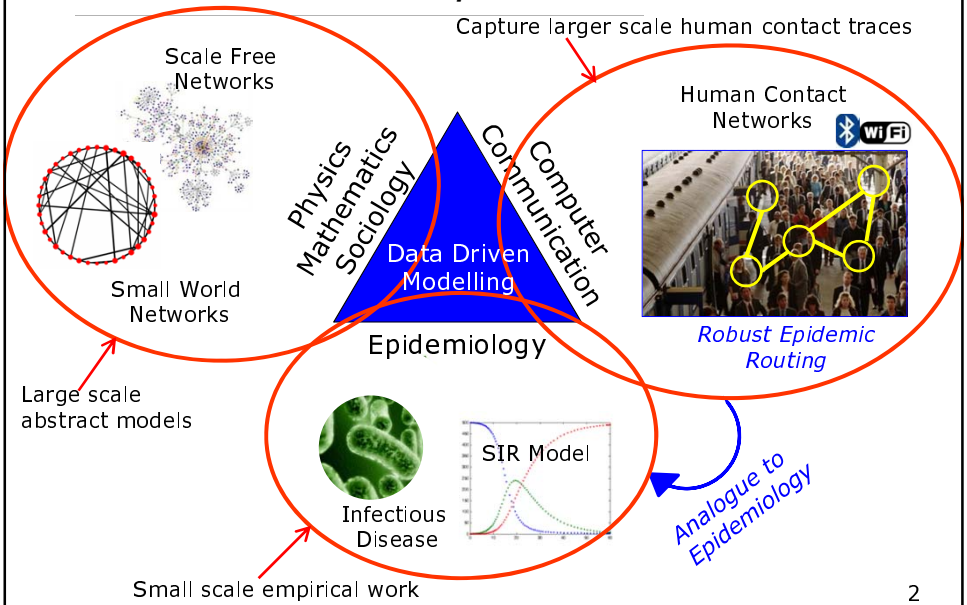
Data Collection using Short-Range Radio for Modelling Dynamic Human Contact Networks

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

Systems Research Group
University of Cambridge Computer Laboratory

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Data Driven Complex Network Research

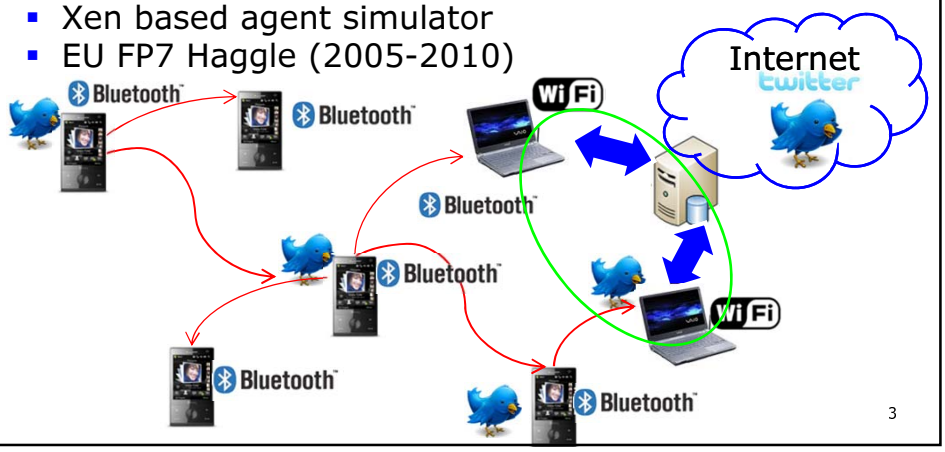


2

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Opportunistic Networks

- Opportunistic networks are Human in nature
  - Devices carried by people, thus 'do what users do'
- Exploit radio communication in proximity range
- Human social network structure to optimise protocol
- Xen based agent simulator
- EU FP7 Haggle (2005-2010)



3

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Empirical Approach

- Robust data collection from real world
- Post-facto analysis and modelling yield insight into human interactions
- Data is useful from building communication protocol to understanding disease spread

Modelling Contact Networks: Empirical Approach

4

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Outline

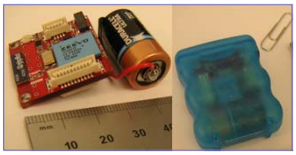
- New Communication Paradigm
  - → Opportunistic Networks
  - Empirical Approach to understand Network Structure
- Data Collection of Human Contact Networks
  - Bluetooth, GPS, RFID Tag…
  - SCaMPiE Day Experiments
- Network Characteristics
  - Inter-Contact Time
  - Social Communities based on Contacts

5

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Electronic Data for Contact Networks

- Sensors
  - Bluetooth Intel iMote
  - 802.15.4



- RFID Tags
  - UHF Tag Alien ALN-9640 - "Squiggle®" Inlay

- Mobile Phones
  - Virtual Disease Application Android Nexus One
  - FluPhone Application Nokia 6730
  - AroundYou Application Nokia 5200
  - GPS, Google latitude

- GPS Logger

- Online Social Networks
  - Foursquare: Checkin any location

6

Sensor Board or Phone or RFID Tag..

- iMote needs battery
  - Expensive
  - Third world experiment
  - New packaging (wrist band, medallion)
- Mobile phone
  - Rechargeable
  - Additional functions (messaging, tracing)
  - Smart phone: location assist applications
- Stationary or Mobile detection
- Provide device or software
- Combine with online information (e.g. Foursquare, Twitter)

7

Experiment Parameters vs Data Quality

- Battery life vs Granularity of detection interval
- Duration of experiments
  - Day, week, month, or year?
- Data Storage
  - e.g. FluPhone: Contact/GPS data < 50KB per device per day (in compressed format)
  - Server data storage for receiving data from devices
  - Extend storage by larger memory card
- Incentive for participating experiments
- Target populations

8

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Phone Price vs Functionality

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

9

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Location Data

- Location data necessary?
  - 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 2009

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.

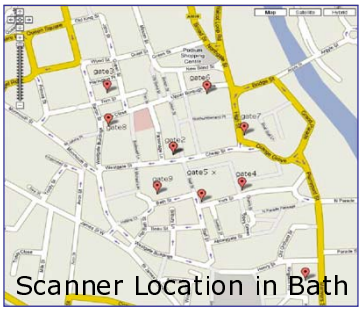
Scientists track people walking around cities, using the Bluetooth signals from their mobiles, laptops and handheld computers.

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

SCANNING YOUR EVERY STEP

1 Bluetooth mobile phones, laptops, handheld computers, earpieces and digital cameras broadcast a unique code and name.

BATH


Scanner Location in Bath

10

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Data Retrieval Methods

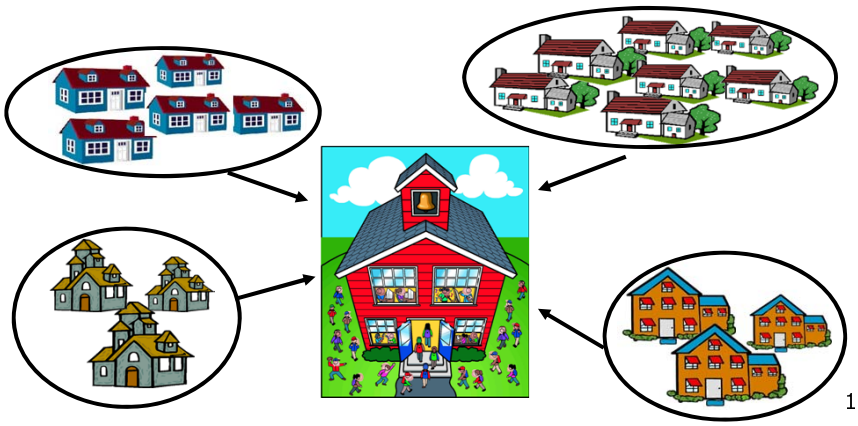
- Retrieving collected data:
  - Tracking station
  - Online (3G, SMS)
  - Uploading via Web
  - via memory card
- Incentive for participating experiments
- Collection cycle: real-time, day, or week?

11

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Target Population

- Provide devices to limited population or target general public
  - For epidemiology study ~=100% coverage may be required
- Or school as mixing centres



12

Data Transformation for Analysis

- Transform to discrete version of contact data
- Deal with noise and missing data
  - Ex. transitivity closure
- Data analysis requires high performance computer and storage
  - Low volume - raw data in compact format
  - Transformation of raw data for analysis increases data volume

13

Security and Privacy

- Current common method: basic anonymisation of identities (e.g. MAC address)
- Data packets encrypted over Internet
- Anonymising identities may not be enough?
  - Simple anonymisation does not prevent to be found the social graph
- Ethic approval is requied

14

UNIVERSITY OF CAMBRIDGE Computer Laboratory

Proximity Detection by Bluetooth

- Only ~=15% of devices Bluetooth on
- Scanning Interval
  - 2 mins iMote (one week battery life)
  - 5 mins phone (one day battery life)
  - or continuous scanning by station nodes
- Bluetooth inquiry (e.g. 5.12 seconds) gives >90% chance of finding device
- Complex discovery protocol
  - Two modes: discovery and being discovered
- 5~10m discover range
- Advantage: most phones have Bluetooth

Can it produce reliable data (negligible noise)?

15

UNIVERSITY OF CAMBRIDGE Computer Laboratory

RFID Tags

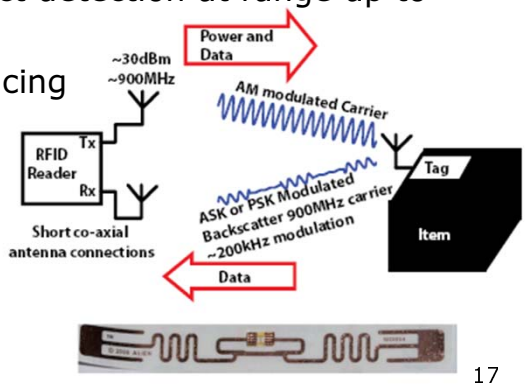
- Radio-frequency identification (RFID): Use of radio waves to exchange data between reader and electronic tag
- Either passive (no battery) or active (with an on-board battery that always broadcasts or beacons its signal) or battery assisted passive
- High-frequency RFID or HFID/HighFID tags
  - library book or bookstore tracking, Oyster card
- UHF, Ultra-HighFID or UHFID tags
  - Shipping container tracking
  - Ski lift ticket
- Fixed RFID and Mobile RFID
  - Fixed: Stationary position
  - Mobile: hand helds, carts and vehicle

16

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Passive UHF RF tag

- TINA Project: Intelligent Airport
  - Location based service
  - RFID to boarding pass
- Passive UHF RFID allows very lo cost tags (10p) to be used for object detection at range up to 10m (tuneable)
- Closer antenna spacing



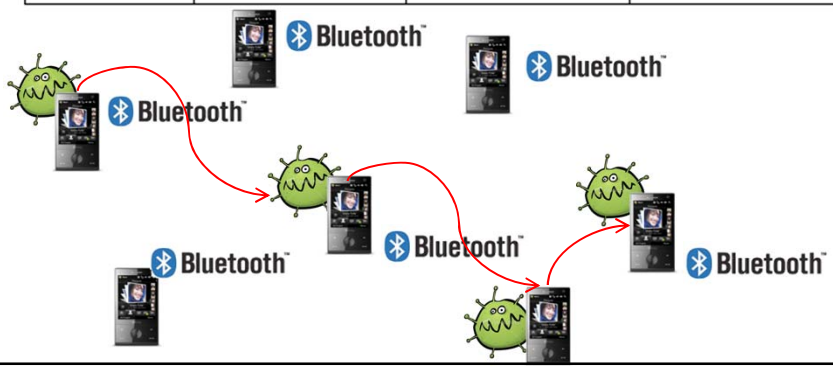
17

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Virtual Disease: Nexus One

- Spread virtual disease via Bluetooth communication
- Today: 3 seed nodes with base and 1 with all diseases

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



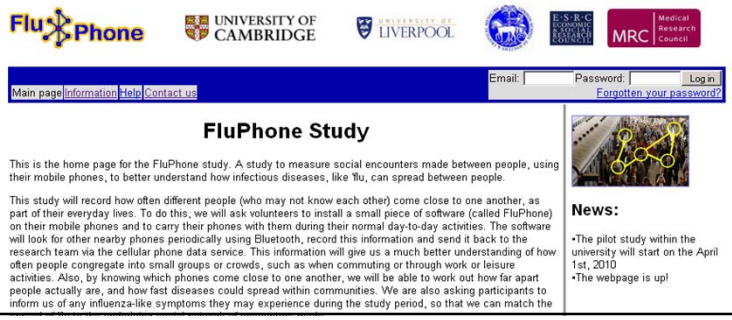
18

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

FluPhone Project

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

https://www.fluphone.org



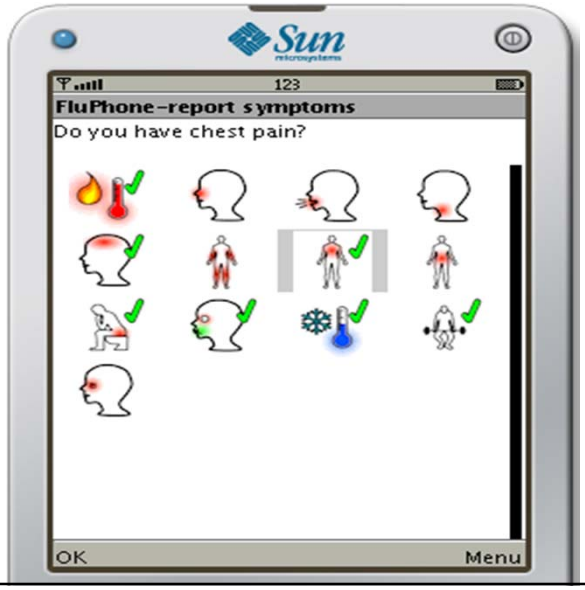
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

FluPhone: Main Screen

- Scan Bluetooth devices every 2 minutes (today's experiment)

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

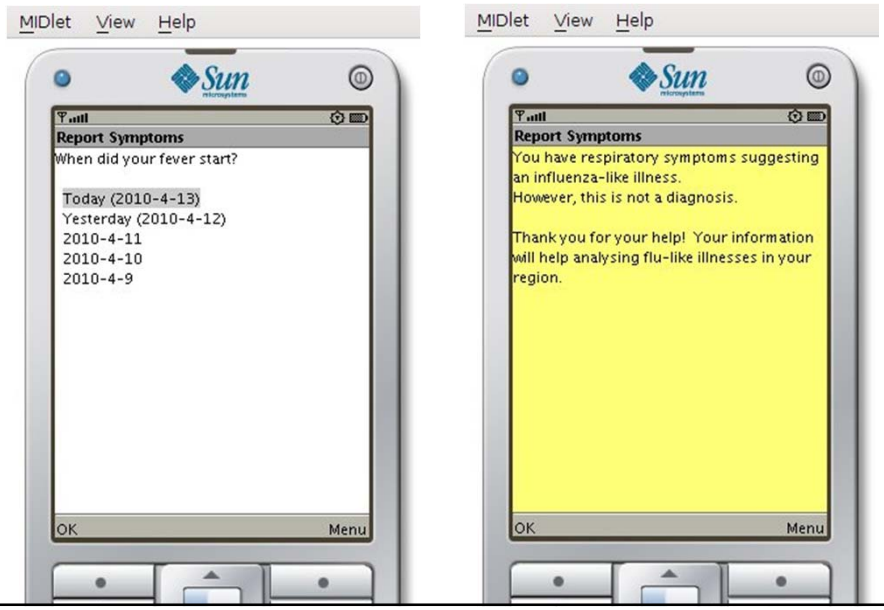
FluPhone: Report Symptom



21

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

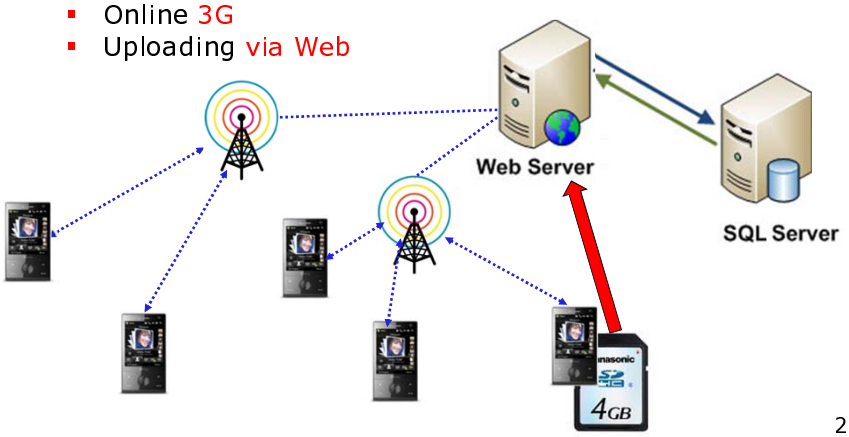
FluPhone: Report Time - Feedback



UNIVERSITY OF CAMBRIDGE
Computer Laboratory

FluPhone Server – Data Collection

- Via GPRS/3G FluPhone server collects data
- Collection cycle: ~real-time, day, or week?
- Collection methods:
  - Online 3G
  - Uploading via Web

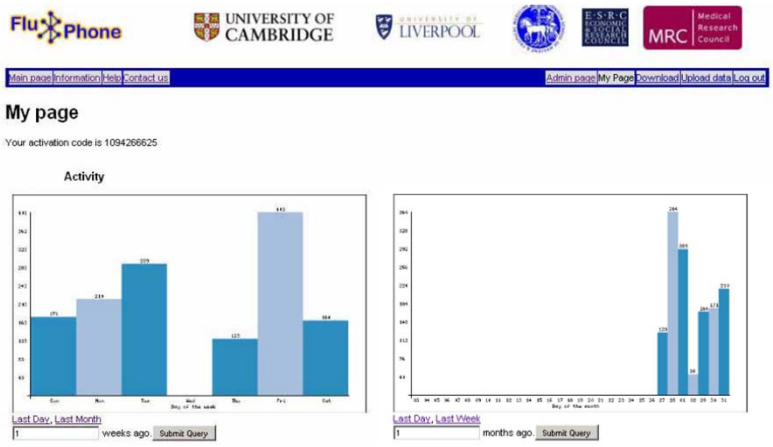


23

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

FluPhone MyPage

- FluPhone participants can login to personal page to see your activity



24

Human Connectivity Traces

- Capture potential human interactions
- ..thus far not too large scale – challenge to obtain info from general public
- CRAWDAD database at Dartmouth University

Experimental data set	MIT	UCSD	CAM	INFC0[illegible]	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

Cambridge Projects

25

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, betweenness)
  - Community evolution
  - Network measurement metrics
  - Patterns of interactions

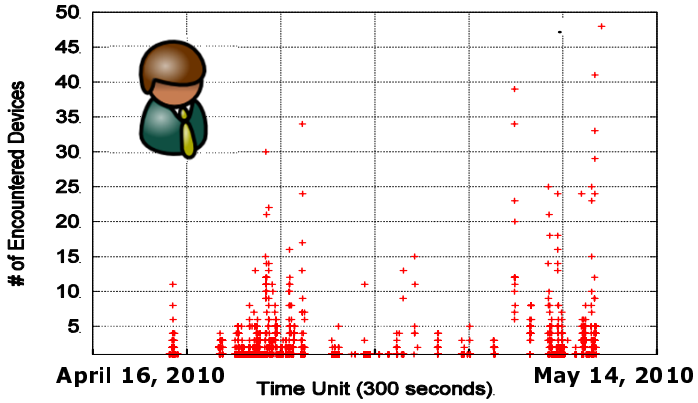
Publications at:
http://www.haggleproject.org
http://www.social-nets.eu/
http://www.cl.cam.ac.uk/~ey204

26

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Encountered Bluetooth Devices

- Encountering History
  - ~1500 unique devices per 10 days

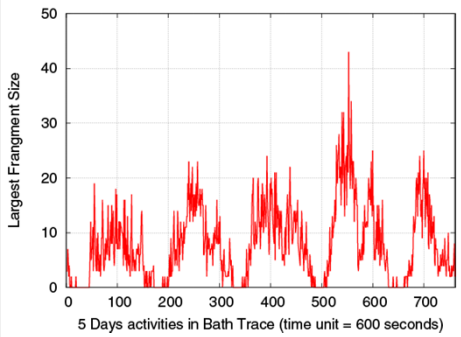


27

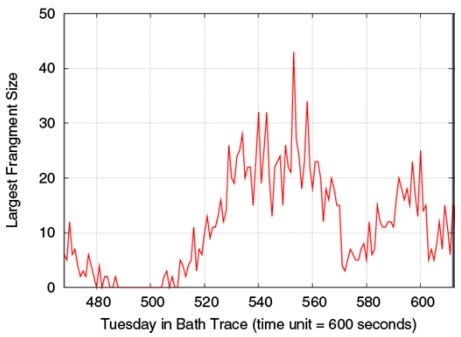
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Regularity of Network Activity

- Size of largest connected nodes shows network dynamics



5 Days



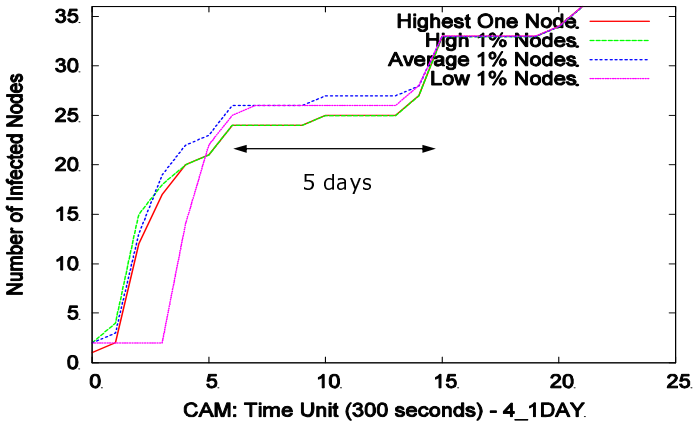
Tuesday

28

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Simple Flood (3 Stages)

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

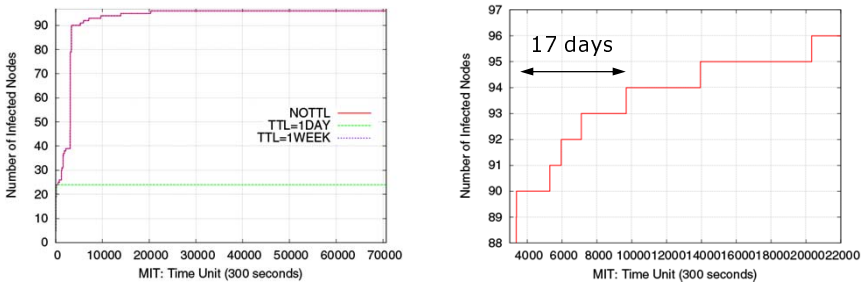


29

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Three Stages of Epidemic Dynamics

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



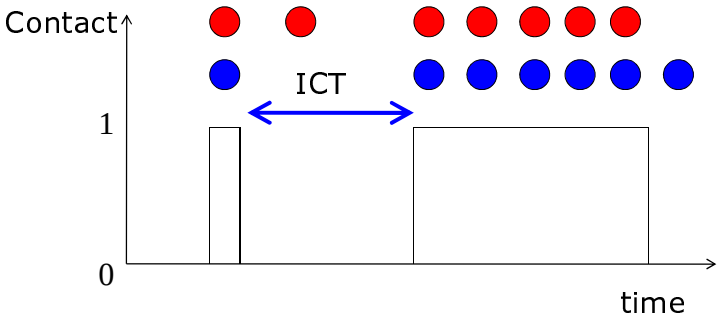
MIT Trace

30

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Inter-Contact Time (ICT)

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

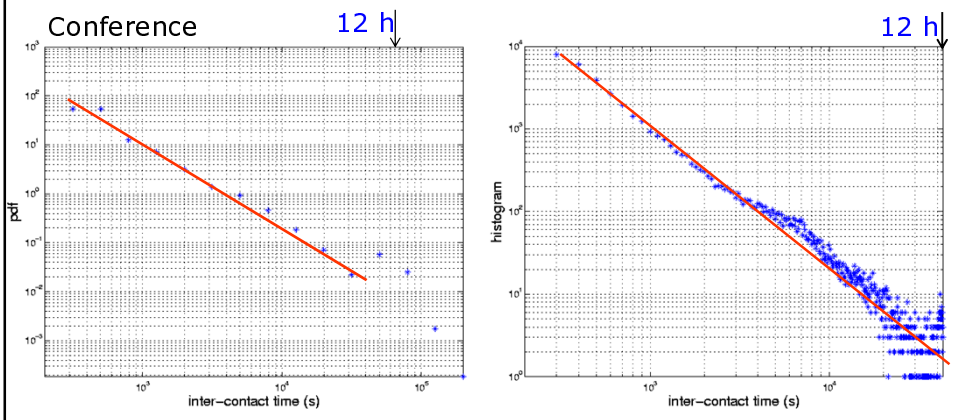


31

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

ICT: Random and Scale-free

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



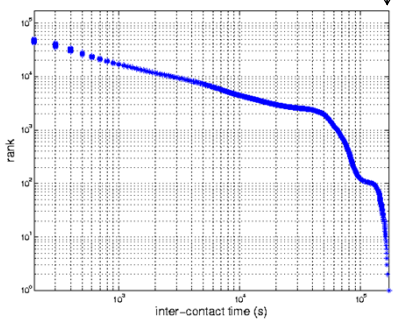
32

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

ICT: Truncated

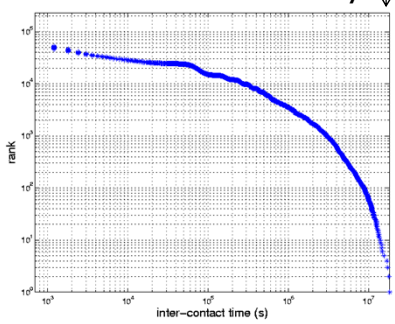
- At some time scale the power law component is truncated by a constraint on inter-contact time
- One artificial constraint is the experiment itself which prohibits recording ICTs longer than the experiment duration

~3 days



Conference

~230 days



MIT

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

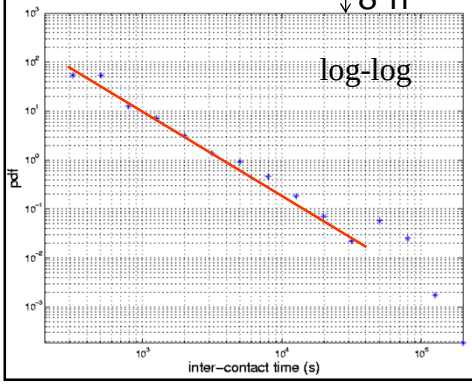
ICT: Truncated

- Another constraint is the removal of nodes from the contact domain. An example of this is movement from work to home which suppresses ICTs between agents in the same work group on times scales beyond the working day. This truncates the power law component at ICT ~ 8 h

INFOCOM 2006

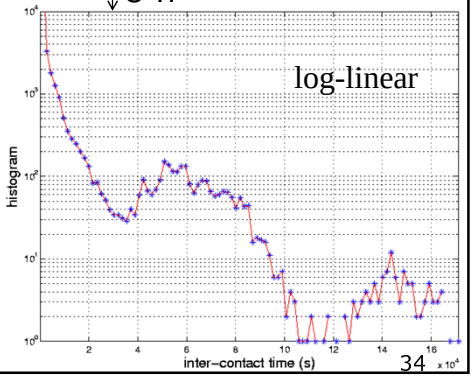
8 h

log-log



8 h

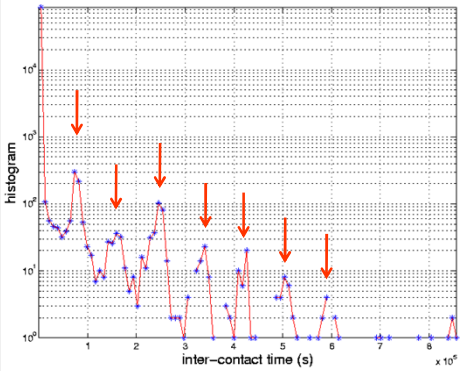
log-linear



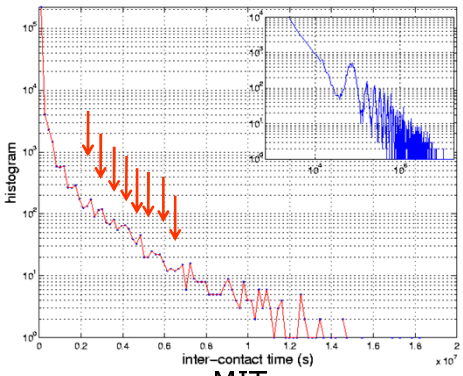
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

ICT: Periodic

- Environmental, biological, and social constraints may have rhythms that encourage repeated encounters such as the daily to-ing and fro-ing between work and home. This gives ICT separated by 24 hours



Cambridge

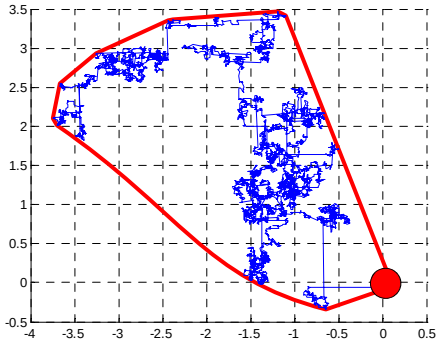


MIT

35

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Example of open Lèvy flight



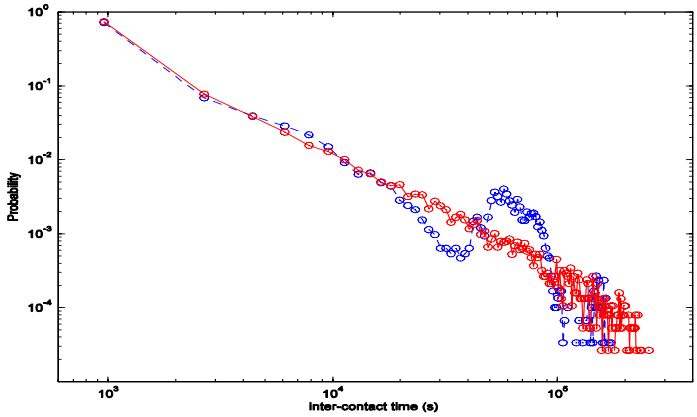
- Example of unconstrained Levy flight, stability index=1.6
  (red circle at lower right denotes start)

36

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Simulation without Periodicity

- Assuming simply a truncated Lèvy flight (red) only roughly describes the actual INFOCOM 2006 distribution (blue)

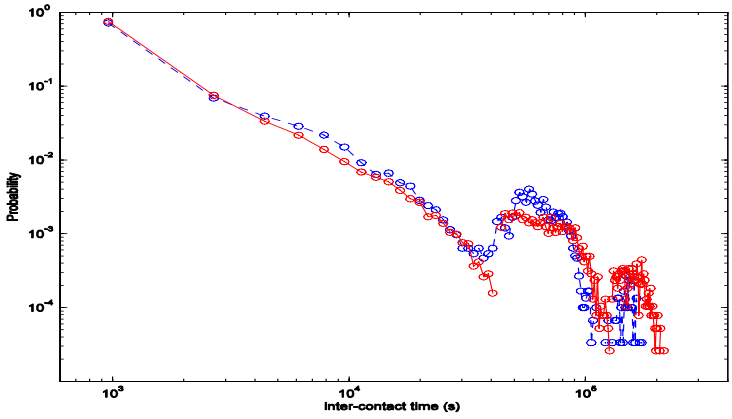


37

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Simulation with Periodicity

- Omitting those contact times outside the working day gives a much better fit, showing the importance of this circadian rhythm

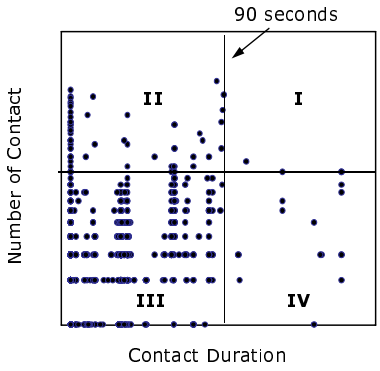


38

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Edge Weight→ Community Detection

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

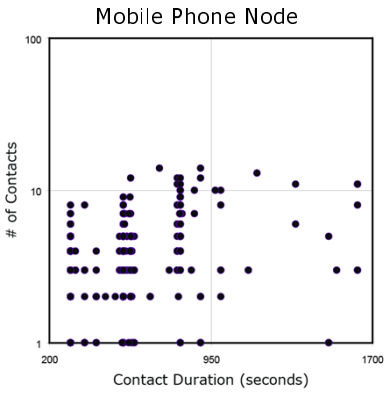


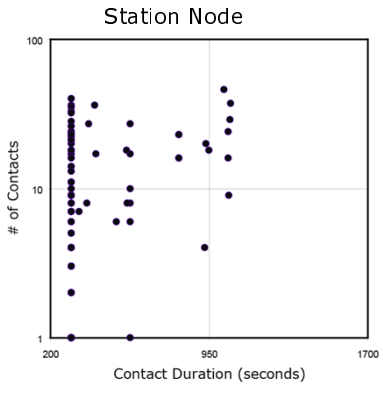
39

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Classification of Node Pairs

- Stationary Device – High visibility but no friends
- Mobile Device – No familiar stranger





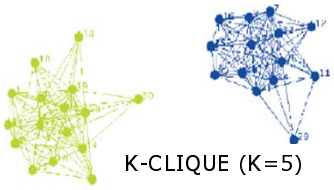
40

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Uncovering Community

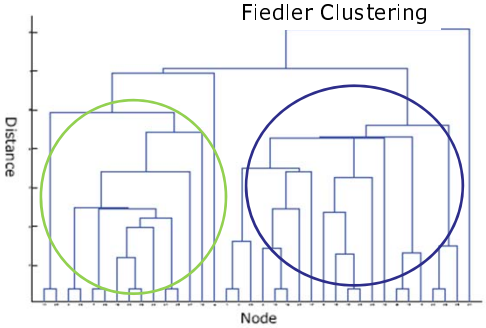
- Contact trace in form of weighted (multi) graphs
  - Contact Frequency and Duration
- Use community detection algorithms from complex network studies
  - K-clique [Palla04], Weighted network analysis [Newman05], Betweenness [Newman04], Modularity [Newman06], Fiedler Clustering etc.

Cambridge Trace



K-CLIQUE (K=5)

Fiedler Clustering



41

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

K-CLIQUE Detection

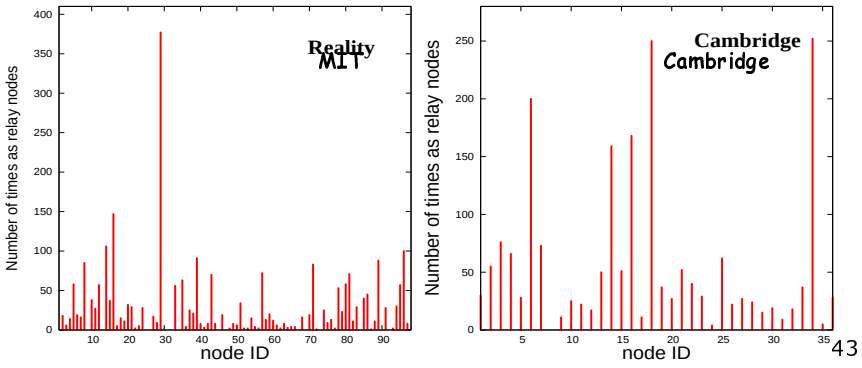
- Union of k-cliques reachable through a series of adjacent k-cliques
- Adjacent k-cliques share k-1 nodes
- Members in a community reachable through well-connected subsets
- Examples
  - 2-clique (connected components)
  - 3-clique (overlapping triangles)
- Overlapping feature

42

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Betweenness Centrality

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

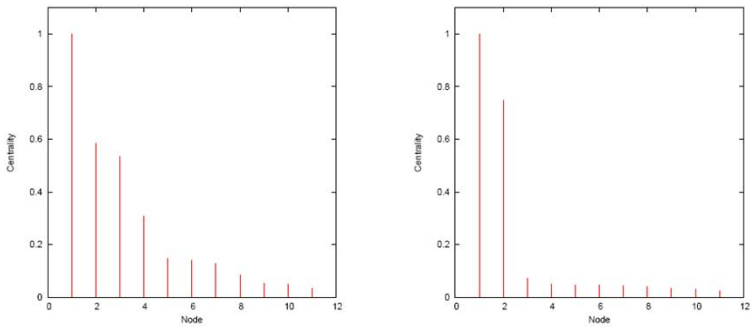


43

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Betweenness Centrality

- Centrality in two groups in Cambridge
  - Group A: Undergraduate year1
  - Group B: Undergraduate year2



Cambridge Group A

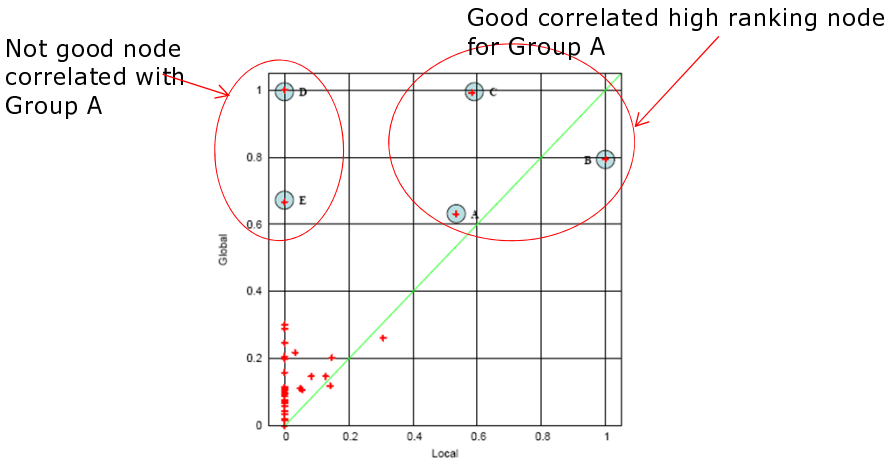
Cambridge Group B

44

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Local centrality and Global Centrality

- Correlation of centrality of Group A and global centrality

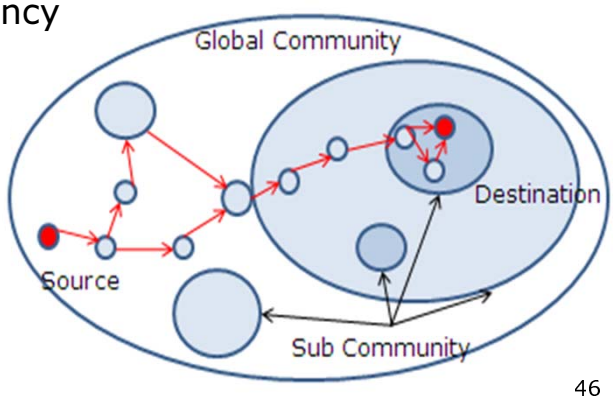


Cambridge

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Social Structure for Communication

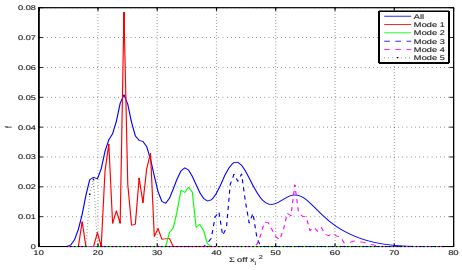
- Robust epidemic routing
- Use of social hubs (e.g. celebrities and postman) as betweenness centrality and combining community structure for improved routing efficiency



UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Joint Diagonalisation

- Build average interaction graph by combining many of spanning tree based samples of a network
- Use of Joint Diagonalisation
- Distribution of deviation from average graph is multi-modal → different behaviour of network
- Change of mode corresponds with transition to infectious state



47

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Simulation of Disease – SEIR Model

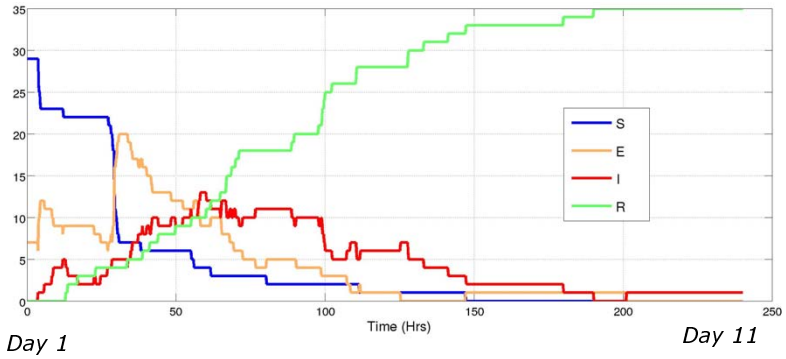
- Four states on each node:
  SUSCEPTIBLE→**E**XPOSED→**I**NFECTED→**R**ECOVERD
- Parameters
  - p: exposure probability
  - a: exposed time (incubation period)
  - t: infected time
- Diseases
  - D1 (SARS): p=0.8, a=24H, t=30H
  - D2 (FLU): p=0.4, a=48H, t=60H
  - D3 (COLD): p=0.2, a=72H, t=120H
- Seed nodes
  - Random selection of 20% of nodes (=7) among 36 nodes

48

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

SARS Simulation

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



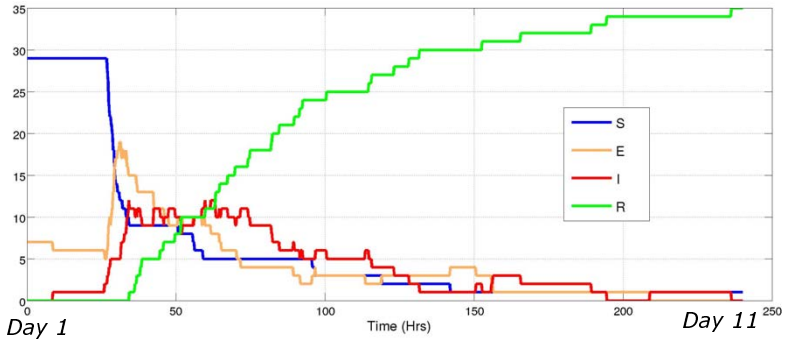
Day 1 *Day 11*

49

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Flu Simulation

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

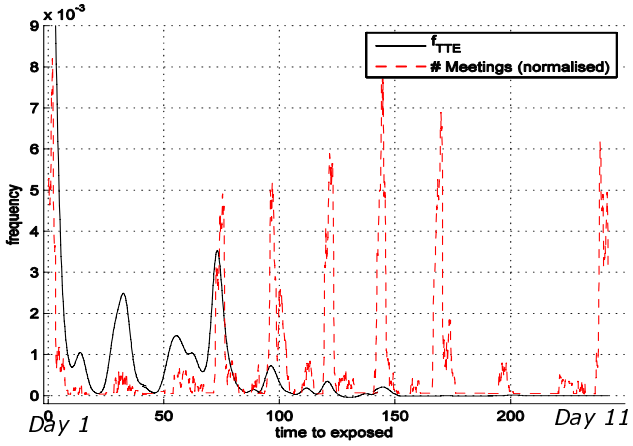


Day 1 *Day 11*

50

Time to Exposure vs #of Meetings

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



51

Summary

- Data Driven Approach: Data is useful from building communication protocol to understanding disease spread
- Post-facto analysis and modelling yield insight into human interactions
  - How does community structure affect epidemic spread?
  - How do hubs and weak links influence temporal or spatial effects, and how does this affect the transmission characteristics of disease?
  - How does community topology of interpersonal connections and its hierarchical nature yield a multi-level structure?
- Where to exploit such Social Structure to Computer Systems beyond Communication and Epidemiology?

52