Study of Wireless Epidemic Spread in Dynamic Human Connectivity Traces

Eiko Yoneki and J. Crowcroft
University of Cambridge Computer Laboratory

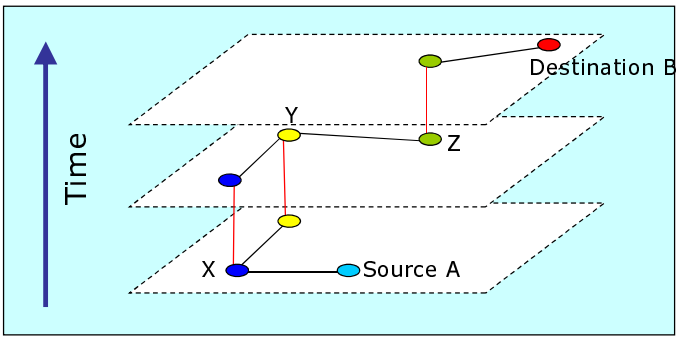
Motivation: Delay Tolerant Networks (DTN)

- Haggle EU FP6: New communication paradigm using dynamic interconnectedness
  - http://www.haggleproject.org
- Pocket Switched Networks (PSN) – A type of DTNs
  - Human-to-Human Communication exhibits Characteristics of Social Networks
- *Store-Carry-Forward* Paradigm
  - Nodes buffer and carry data when disconnected
  - Nodes exchange data when met

2

Time Dependent Networks

- Data paths may not exist at any one point in time but do exist *over time*



3

Use Epidemic: Forwarding Protocols

- Epidemic Information Dissemination: updates spread from source through local interaction
  - Highly robust against disconnection, mobility, and node failures; simple, decentralised, and fast
- However, careful tuning to achieve reliability and minimise network load
  - Control Flooding (e.g. Location, Count-base, Timer, History)
- Logical Connection Topology: Backbone Structure (e.g. Social Networks)
- Understanding Network Structure is important

4

Human Connectivity Traces

- Capture Human Interactions

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

- iMotes (ARM processor with 64K flash memory)
- BATH: plus Brief Location Information

5

Proximity Detection by Bluetooth

- Bluetooth usage (e.g. Bath, UK 7.5%, San Francisco, USA 13.5% among all pedestrians)
  - Scanning Interval
    - 2 mins iMote (one week battery life)
    - 5 mins phone (one day battery life)
    - BATH Trace: Continuous scanning
  - BT inquiry can only happen in 1.28-second intervals. 4x1.28 (5.12 seconds) gives >90% chance of finding device. But no data available when many devices and people are around
  - *Need Higher Fine-Grained Trace !*

6

Human Connectivity Traces

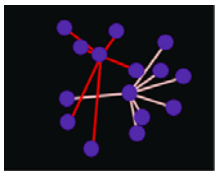
- Average Hops and Cluster Coefficient

Experimental traces	Average Hop Count	Cluster Coefficient
MIT	1.6	0.44
UCSD	2.2	0.41
CAM	1.2	0.66
INFC06	1.5	0.52
BATH	3.3	0.45

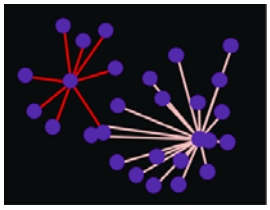
7

Dynamic Human Networks

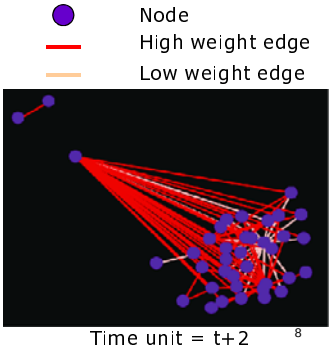
- Topology changes every time unit
- Need new measures to understand networks
  - Time (and space) -based degree, centrality, cluster coefficient...



Time unit = t



Time unit = t+1

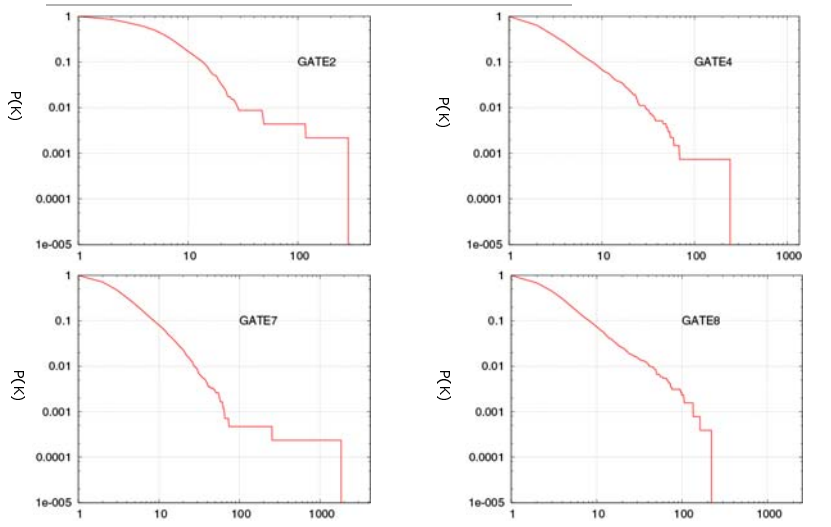


Time unit = t+2

UCSD Trace

8

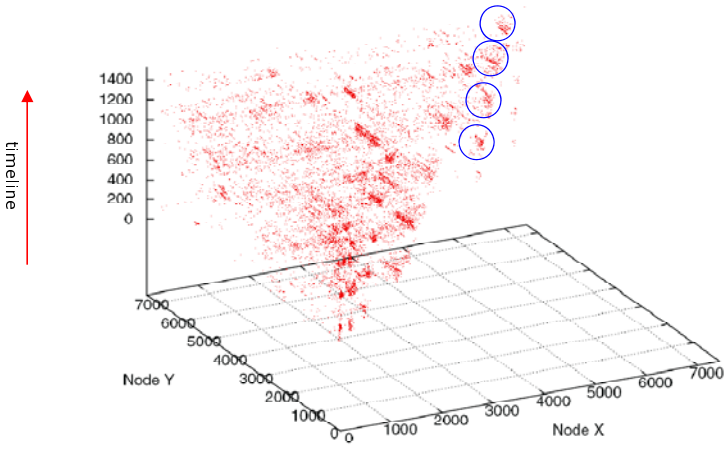
Node Degree Distribution – Location (BATH)



9

Encountering Pairs (BATH) – 5 Days

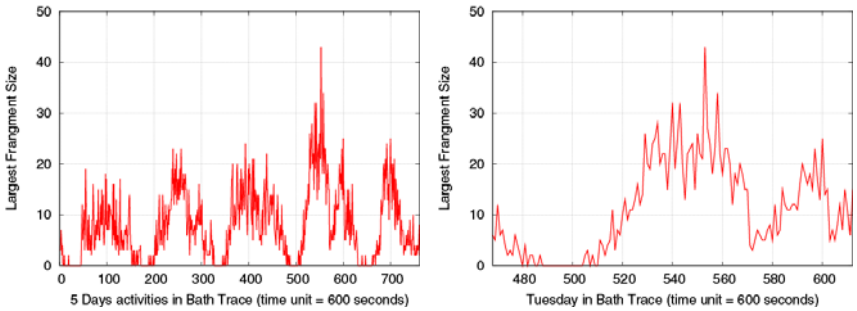
- *Regularity of Encountering*



10

Regularity of Network Activity (BATH)

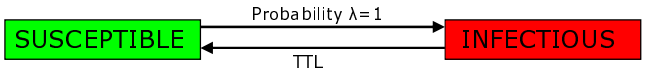
- Size of largest fragment shows network dynamics



5 Days *Tuesday*

11

Epidemic Model – SIS Model

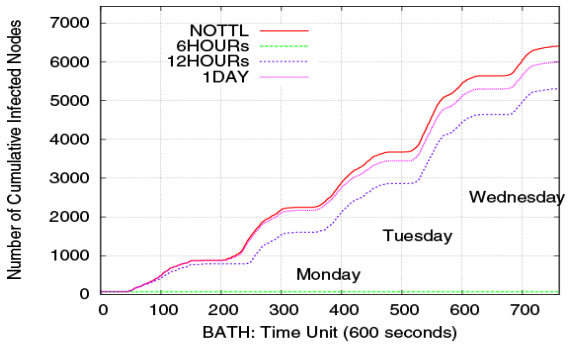


- Two states on each node:
  - SUSCEPTIBLE (currently not infected)
  - INFECTIOUS (infected)
- Disease/virus has TTL (time-to-live) property; node turns into SUSCEPTIBLE state when TTL expires
- At every time unit, a node with INFECTIOUS state infects other nodes in proximity

12

Population of Infected Nodes

- TTL sets scale for dynamics
  - Varying TTL shows Different Time-Dependent Networks

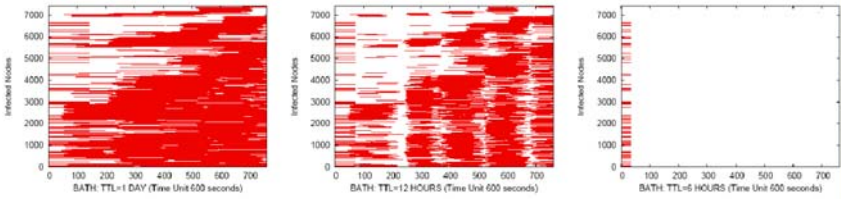


Cumulative Infectious Nodes

13

Evolution of Infectious Nodes

- Human Epidemic: TTL close to order of day

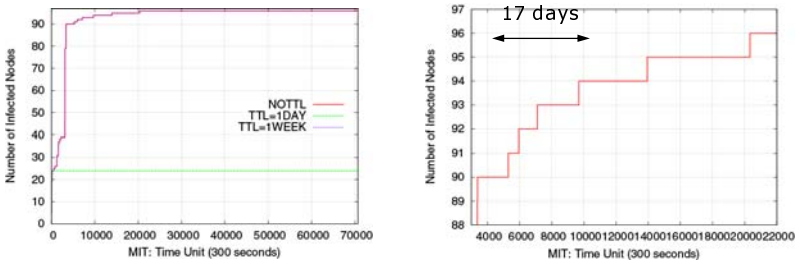


1 DAY | 12 HOURS | 6 HOURS

14

Three Stages of Epidemic Dynamics

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

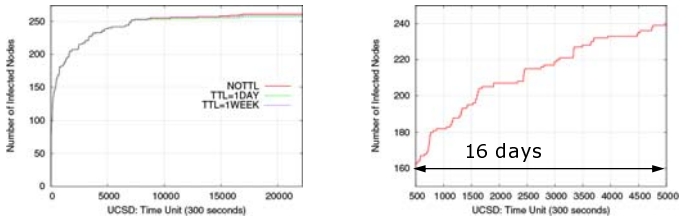


MIT Trace

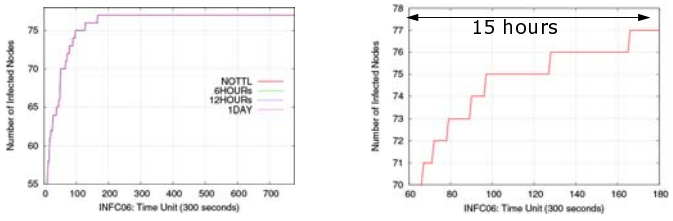
15

Three Stages of Epidemic Dynamics (continued)

- UCSD



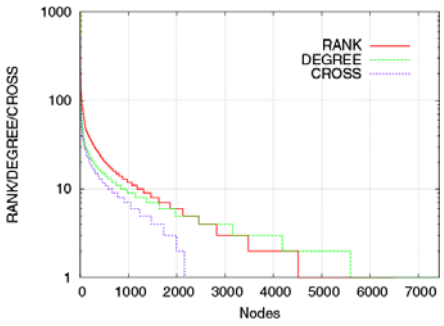
- INFC06



16

Influential Nodes - Hubs

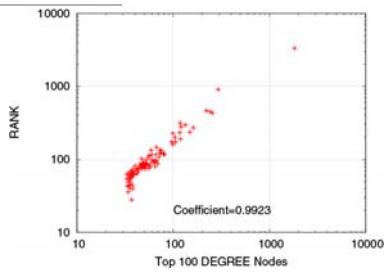
- **DEGREE**: Number of connecting nodes (Degree Centrality)
- **RANK**: How often is node used to relay data to other nodes (Betweenness Centrality)
- **CROSS**: How often node appears at different location (Mobility Centrality)
- Extract top 100 nodes

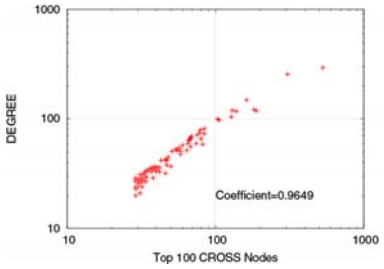


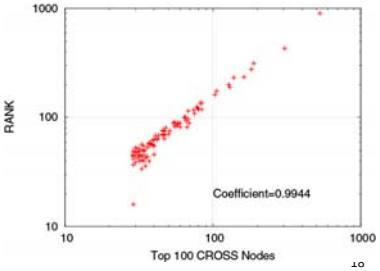
17

Correlation DEGREE, RANK and CROSS

- Top 100 Nodes
- CROSS/RANK: Highest
- CROSS/RANK > CROSS/DEGREE







Hub Nodes

- Correlation of Hub Nodes

Category	All Nodes	Top 100 Nodes	Top 50 Nodes	Top 30 Nodes
Rank/Degree	0.99	0.99	0.99	0.99
Degree/Cross	0.97	0.96	0.96	0.96
Corss/Rank	0.99	0.99	0.99	0.99

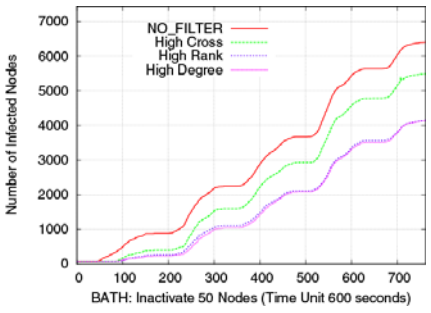
- Membership Similarity (Jaccard Coefficient)

Top n Nodes	Rank/Degree	Rank/Cross	Degree/Cross
100	0.79	0.43	0.44
70	0.92	0.41	0.41
50	1.00	0.43	0.49
30	1.00	0.46	0.46
10	1.00	0.33	0.33

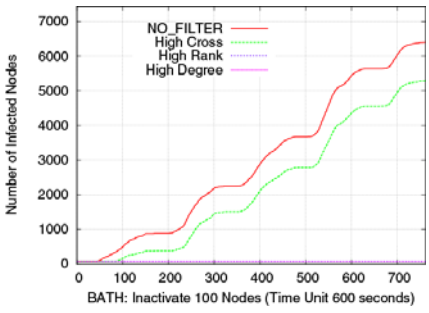
19

Inactivation of Hub Nodes in Epidemic Spread

- DEGREE and RANK give higher impact than CROSS
- Inactivation of 100 High DEGREE nodes among 7431 nodes kills epidemic



Top **50** Nodes

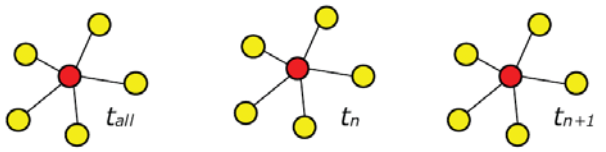


Top **100** Nodes

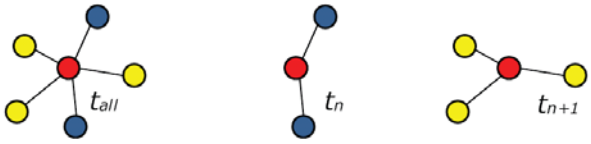
20

Party and Date Hubs

- High Degree Distribution: Party Hub connects to the same set of nodes, while Date Hub changes the neighbourhood nodes



Party Hub: Same Time and Space



Date Hub: Different Time and /or Space

21

Neighbourhood Similarity Rate

- Find High RANK Hub locally
- Neighbourhood Similarity Rate (NSR)*

$$NSR = N_t \cap N_{(t+1)} / N_t \cup N_{(t+1)}$$

N is a set of neighbourhood nodes

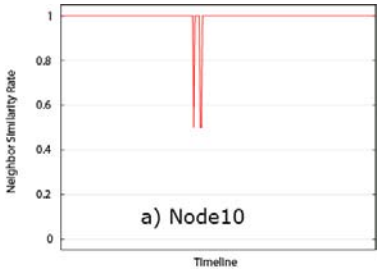
- Neighbourhood plus Neighbourhood Similarity Rate (NNSR)
- Can NSR and NNSR characterise Party and Date Hubs?

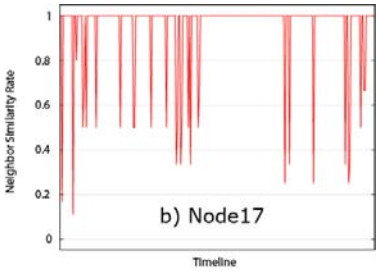
* time unit without connectivity is suppressed in sparse networks

22

Evolution of NSR (MIT Trace)

- Node10: Continuous High NSR
- Node 17: Change of Neighbourhood

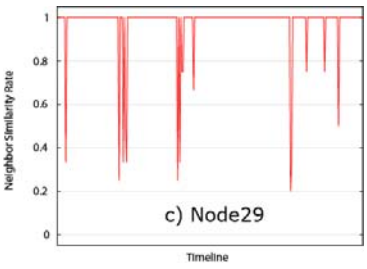


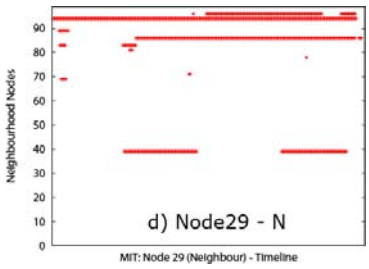


23

Evolution of NSR and Actual Neighbourhood

- NSR shows occasional changes of Neighbourhood
- Map to Actual Change of Neighbourhood Nodes





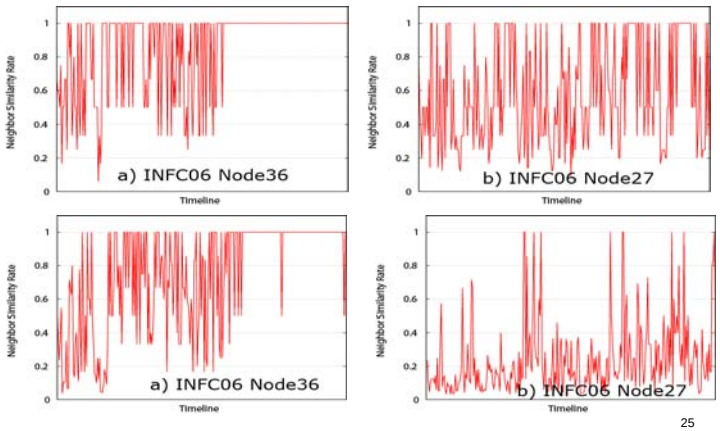
24

Evolution of NSR (INFC06 Trace)

- NNSR amplifies NSR

NSR

NNSR



25

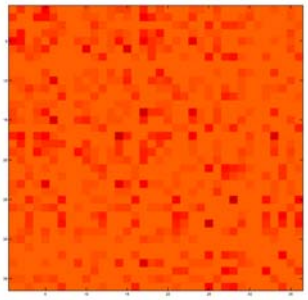
Party and Date Hubs: Outlook

- On average, 30% (Party Hub), 40% (Date Hub), 30% (Combined)
- BATH trace results 0: too sparse
- Current results still need to be refined
- Limitation: NSR shows evolution rate but not neighbourhood size

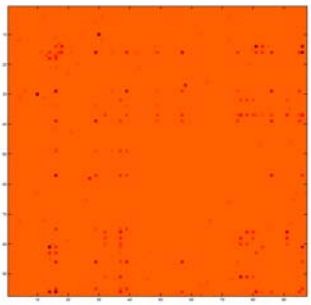
26

Pair Node Relation by SVD

- Use of SVD (Singular Value Decomposition)
  - Eigenvectors corresponding to its dominant modes of behaviour



CAM



MIT

  - Dark colour indicates a pair nodes do not correlate/activate at the same time (e.g. MIT node 17)

27

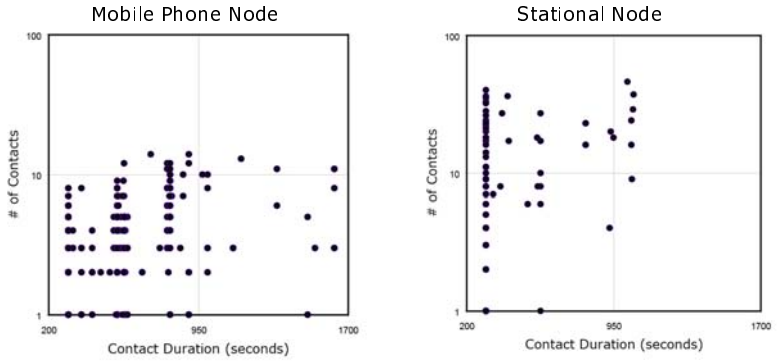
Uncovering Community

- Social Networks: Community in Mobility Traces
- Mobility Trace in Form of Weighted (Multi) Graphs
  - Use contact duration and frequency for defining node pair relationship
- Both Centralised and Decentralised Detections
  - Use community detection algorithms from complex network studies
    - K-clique detection[Palla04]
    - Weighted network analysis[Newman05]
    - Betweenness [Newman04]
    - Modularity [Newman06]
    - Fiedler Vector[Fiedler75]

28

Classification of Node Pairs

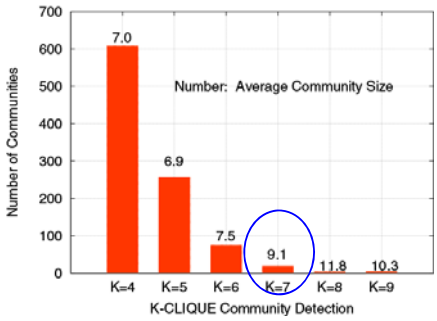
- Stational Device – High visibility but no friends
- Mobile Device – No familiar stranger



29

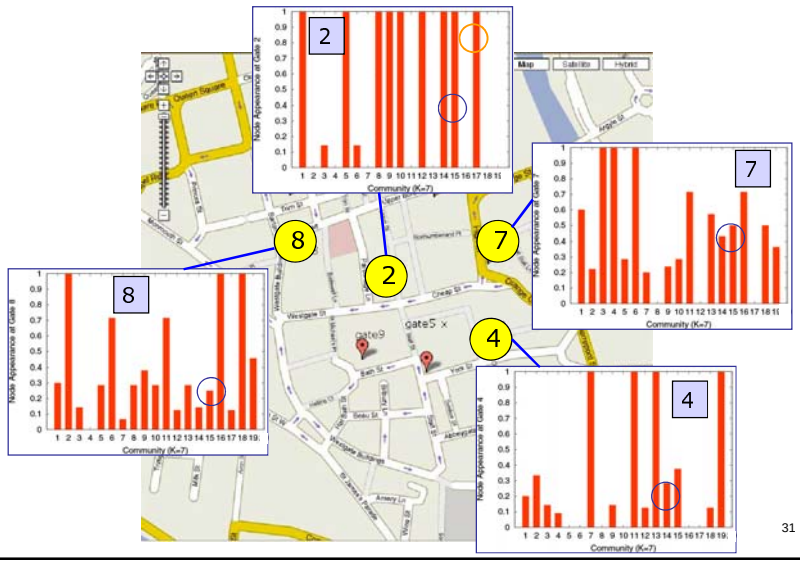
Communities in Bath Trace

- Hard to identify the optimal number of communities
- K=7: 19 communities



30

Inter-Community and Intro-Community



31

Conclusions and Future Work

- Dissemination dynamics is affected by complex interplay between network structure and spreading process
- Further experiments/analysis
  - Directional graph
  - Community tie with location and time
  - Membership evolution

Thank You !

Email: eiko.yoneki@cl.cam.ac.uk

32