From Community Detection to Group Communication in DTNs

Eiko Yoneki
University of Cambridge Computer Laboratory

Joint work with
Pan Hui
Jon Crowcroft

Outline

Haggle Project:

- Empirical Approach with Real World Human Mobility Traces
- Further Decentralised Community Detection/Inference
- Socio-Aware Overlay for Many-to-Many Communication
  - Support DTN Applications (e.g. Smart Caching for Content Share → Ad Hoc Google)

Human Mobility Traces

- Capture of Human Interaction: Mobility Data
  - MIT Reality Mining: 100 nodes for 9 months (**MIT**)
  - UCSD: 300 devices for 3 months (**UCSD**)
  - U. Cambridge: e.g. 40 devices for 11 days (**CAM**)
  - U. Bath: 8 gateways for 5 days - inferring 7000 devices
  - more…

  Archive: http://crawdad.cs.dartmouth.edu/data.php
- Proximity Detection by Bluetooth
  - Bluetooth usage (e.g. Bath (UK) 7.5%, Bremen (Germany) 3.5%, San Francisco (USA) 13.5% among all pedestrians)
  - Current Scanning Interval → Coarse-Grained
    - 2 mins iMote for one week and 5mins phone for one day (power consumption)
    - Importance of Random Interval: When Device is Inquiry mode, it is not discoverable → Sleep a random interval
  - BT inquiry can only happen in 1.28 second intervals. 4*1.28 (i.e. 5.12 seconds) gives you more than 90% chance of finding a device. However there is no data available when many devices and many human bodies around.
  - **Need Higher Finer-Grained Trace**
- Use of Zigbee? No Discovery Function like BT

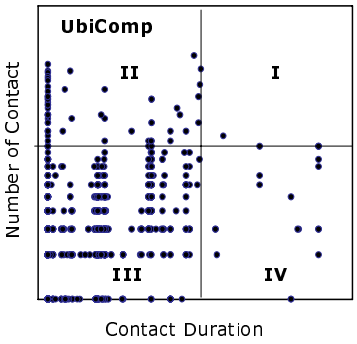
Uncovering Community Structure

- **Distributed Community Detection in Delay Tolerant Networks** SIGCOMM Workshop (MOBIARCH), August, 2007 (to appear)
- Community Structure behind Social Networks in Mobility Traces
- Mobility Trace in Form of Weighted Graph → Multi-Graphs
- Use of Community Detection Algorithms from Complex Network Studies
  - SIMPLE
  - K-CLIQUE
  - Modularity
- Use Contact Duration and Frequency (Nº of Contacts) for Defining Node Pair Relationship

Classification of Pairs

Pair Classification:

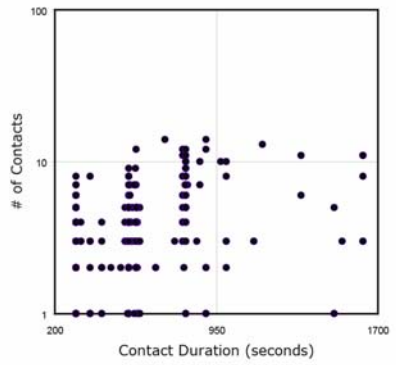
I: **Community**
High Contact Nº - Longer Duration:
II: **Familiar Stranger**
High Contact Nº - Short Duration:
III: **Stranger**
Low Contact Nº – Short Duration:
IV: **Friend**
Low Contact Nº - High Duration:



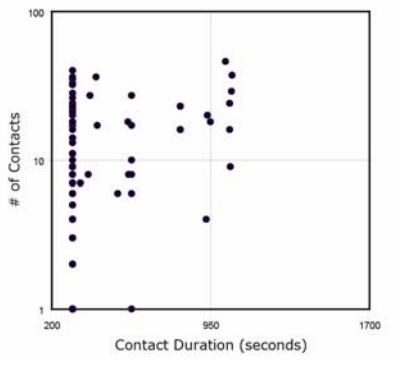
Mobile Phone w/Human and Tracking Station

- Tracking Station – High Visibility but No Friends
- Mobile Device – No Familiar Stranger

Human Mobile Node

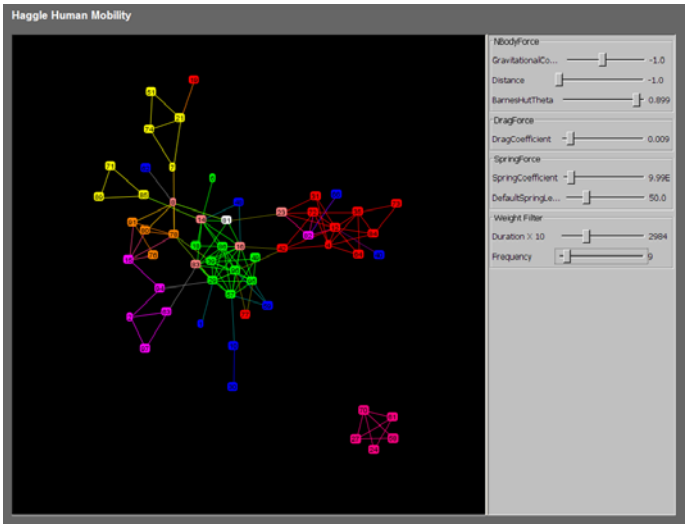


Stational Node



Visualisation Demo 1 – Detected Communities

http://www.cl.cam.ac.uk/~ey204/Haggle/Vis/mobility.html

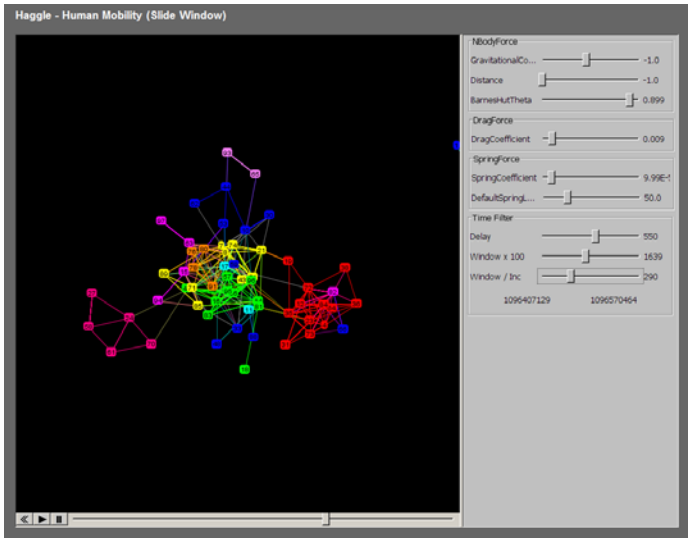


Community Definition ~= Membership

- Current Approach: Contact Duration and Frequency
- Needs Various Aspects:
  - Temporal Information
    - e.g. Minimum duration within certain time windows
    - e.g. Specific time of the day
    - e.g. Matching contact interval sequence
    - e.g. Suppress night/day time
    - e.g. large connected cluster within certain time window
  - Spatial Information
    - e.g. Specific location (and time)
  - Network Locality
    - Ego-centric network – only surrounded nodes
    - Socio-centric – whole network
  - Static Community vs Dynamic Community (surrounding common interests under specific condition)
    - e.g. Same affiliation vs queuing to see the show

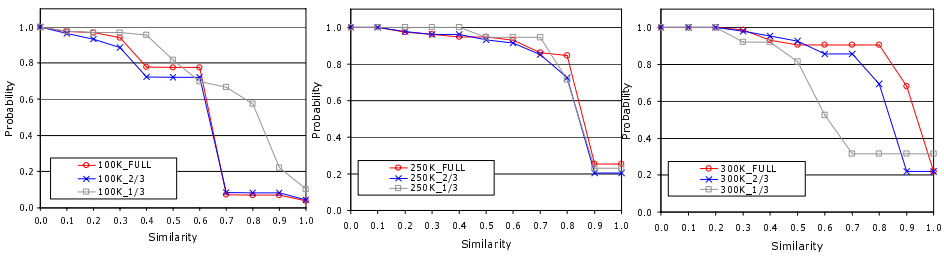
Visualisation Demo 2 – Time-Dependent Detection

http://www.cl.cam.ac.uk/~ey204/Haggle/Vis/mobilitySlide.html



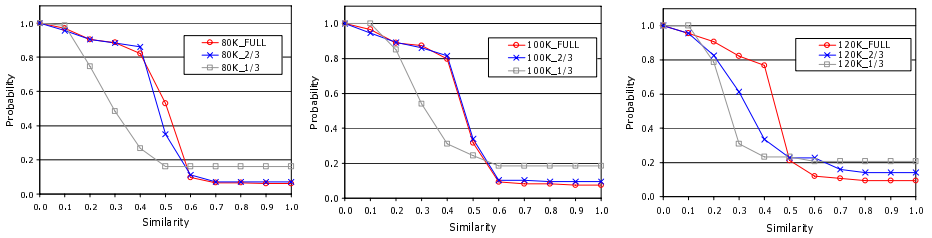
K-CLIQUE with MIT Data

- With Certain Threshold (e.g. 250K seconds contact duration), Detection with 1/3 of time shows comparable result to centralised approach
- Add message passing to improve community detection
  - Similarity improves 0.87 → 0.88 (1/3 position forcing to interact two betweenness nodes)

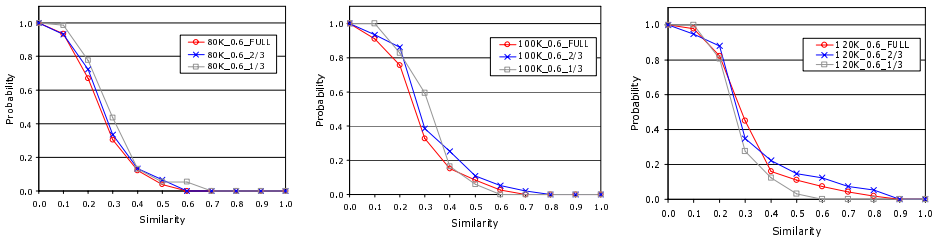


K-Clique and Simple with UCSD Data

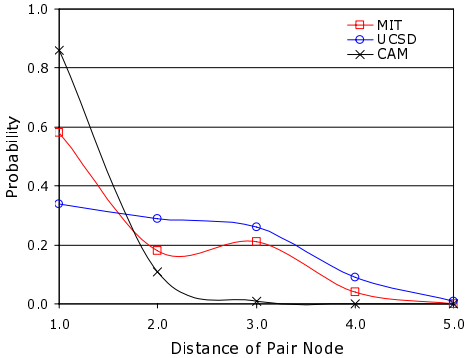
K-Clique



Simple



Time-Dependent Networks: Distance of Pair Nodes



Average Hop Counts:

- MIT 1.6
- UCSD 2.2
- CAM 1.2

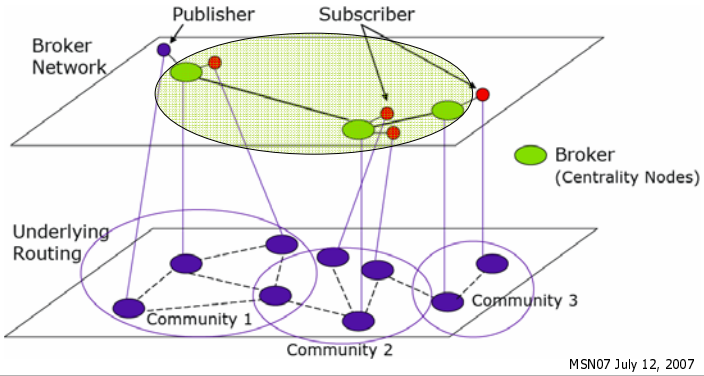
Cluster Coefficient:

Probability of *B connects C*, when *A connects B* and *A connects C*

- MIT 0.44
- UCSD 0.41
- CAM 0.66

Overlay over Communities for Publish/Subscribe

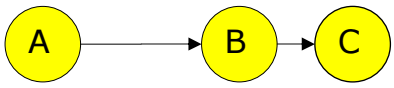
- **A Socio-Aware Overlay for Multi-Point Asynchronous Communication in Delay Tolerant Networks** ACM/IEEE MSWiM, October, 2007 (to appear)
- Subscription Propagation during Community Detection
- Closeness Centrality Nodes Creates Overlay
  - Closeness Centrality = 1.0 (MIT, UCSD, CAM)
  - Multiple Centrality Nodes Coexist → Resource/Load Sharing



Publication State

- Publication State:

A: Publication Created
B: Publication gets first Contact
C: Subscriber Received Publication



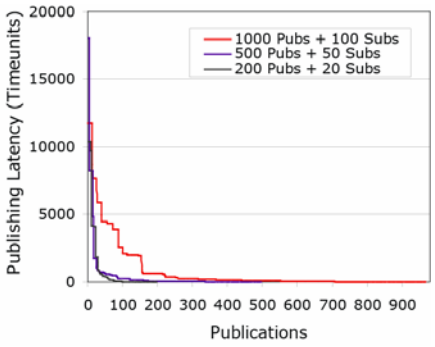
- Haggle – Interactive Interface + Multiple Connectivity
  - Alert to connect to direct connection media
  - Stop at stational nodes nearby…
  - Controlling/Accelerating Information Flow

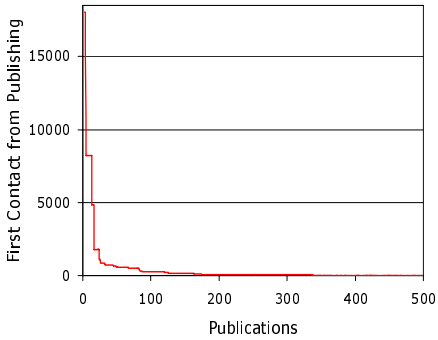
# Pub/Sub	Average Hops	Contact to Sub	Pub to Sub	Latency	Undelivered	Total Hops
1000/100	1.28	5.6 units	631.6 units	5.26 mins	261(26%)	6431
500/50	1.34	4.6 units	828.5 units	6.90 mins	242(48%)	1373
200/20	1.32	4.3 units	831.4 units	6.93 mins	115(58%)	204
1000/100C	1.35	2.7 units	449.4 units	3.75 mins	33(3%)	-

B to C A to C

Latency of Publication

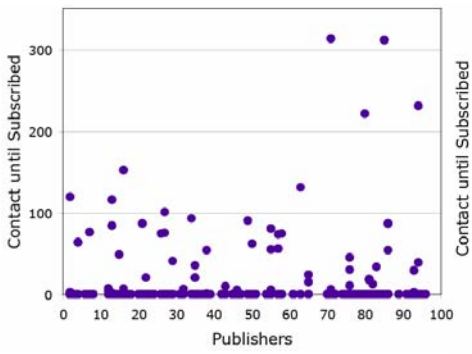
- Most disseminations have short durations

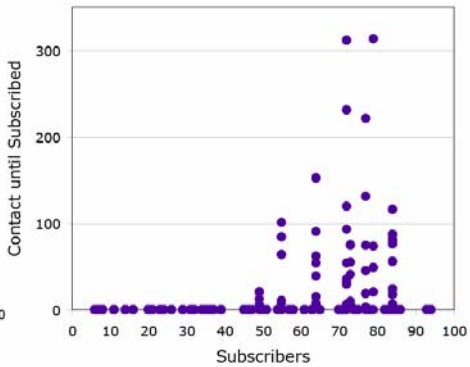




Latency within Network

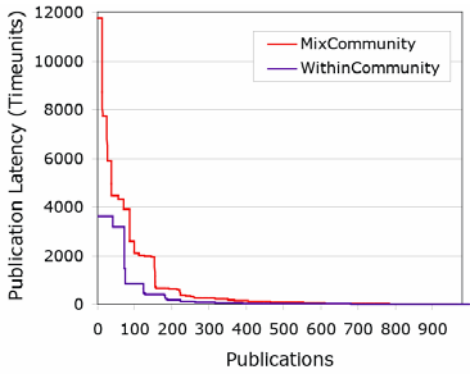
- Once publication has a contact with any node, latency of reaching subscriber is low
- Subscribers 70-80 show longer durations
  - Experiments: 1000 Publications and 100 Subscriptions
  - Possibly Distant from Centrality Node





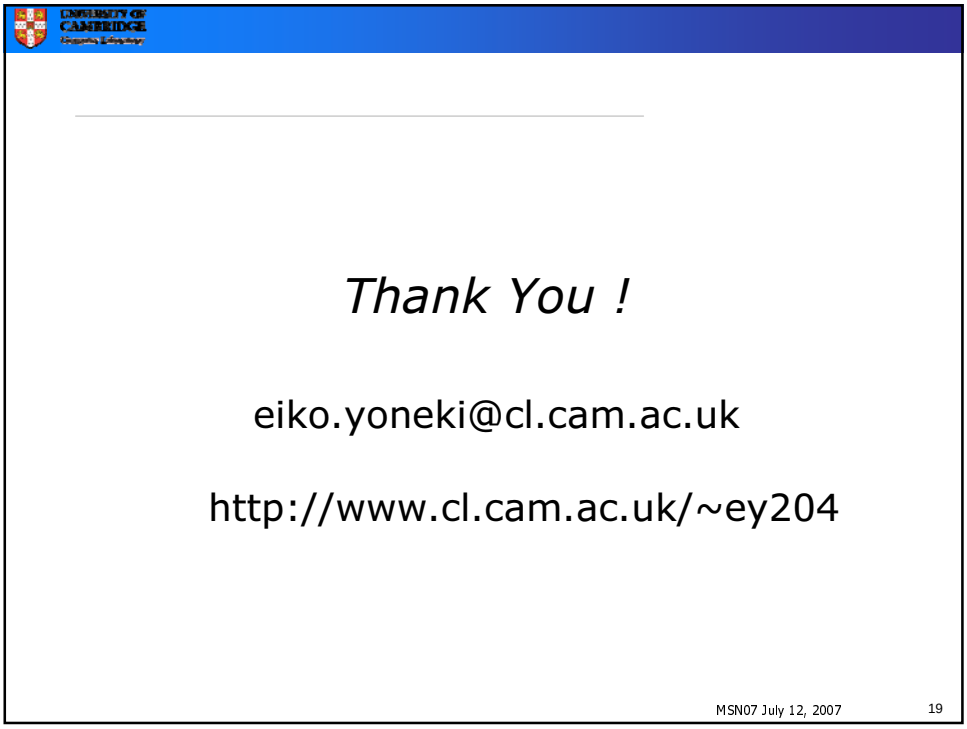
Publish/Subscribe within Community

- Intra-Community Pub/Sub shows Low Latency
  - Mix Community: Publishers and Subscribers evenly spread across Communities
  - Within Community: 90% of publishers and subscribers reside in same community



Future Work: Recall Membership

- Topics may often map to community: Sharing same interest forms community
- Not Yet Membership Management: Membership refresh, expiration, permanent/temporary membership
- Group Communication in pervasive computing tends to be smaller and dynamic
  - Selective dissemination based on contents
  - Extend publish/subscribe semantics


UNIVERSITY OF CAMBRIDGE
Thank You !
eiko.yoneki@cl.cam.ac.uk
http://www.cl.cam.ac.uk/~ey204
MSN07 July 12, 2007
19