Z25 Adaptive and Mobile Systems
Dr. Cecilia Mascolo

Reality Mining: Sensing Complex Social Systems

N. Eagle and A. Pentland

MIT Media Lab

Motivation

- Social network (human contacts patterns) have been studied with models and simulation (surveys)
- You have seen that we use random models to validate our protocols
- No real data about social networks, movements of animals and people existed until now…

- Some traces are starting to emerge (CRAWDAD website at Dartmouth)

Aims of this project

- Study through mobile phone data collection of human patterns
- Both individual use and organizations

- Use of specific hardware to track only limited to specific subsets: now mobile phones are pervasive and more global data can be gained

Mobile phones proximity logs

- Use of short range (bluetooth) and long range (cellular) to augment location and activity info
- Cell and encounters can be logged

- Apps on Bluetooth to log (eg BlueAware and many others). A basestation would allow the transmission of contacts logged to a storage place
- Energy issues: depends on the scanning frequency

Privacy issues

- Yes :)

- Anonymized traces
- Manipulation/aggregation of the data on the phone itself

Identifying structure of routine

- People have daily, weekly and yearly routines
- Aim: build predictive classifiers

- Model of behaviour with 3 states: home, work, elsewhere
- Use data from bluetooth, cell towers and temportal info collected from phones
- Most users spend significant time close to static bluetooth devices (10m indication of position)

Accurate Location

- Difficult to achieve accurate location info using cell towers only
  - Sometimes depends on density, traffic, signal strength
  - The quantity of towers which can be seen by a phone varies substantially with small changes in location

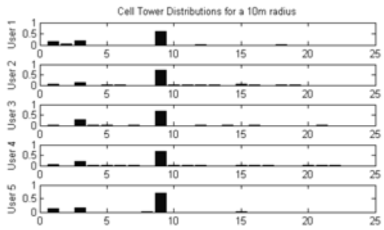

Fig 2. The probability distribution of seeing twenty-five cell towers from the third floor corner of an office building using 150 hours of data from each of five users. (Ranged was assured to 10m by the presence of a static Bluetooth device.)

Fusion of cell tower data with Bluetooth

- Static bluetooth ids used to refine location info

- Users were without cell phone receptions 6% of time but in reach of a bluetooth device or another mobile phone 21% and 29% of time

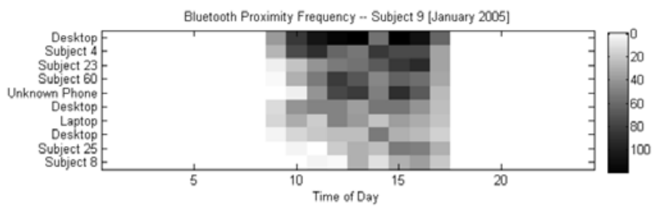

Fig 3. The number of Bluetooth encounters for Subject 9 over the month of January

Entropy metric: a high entropy subject

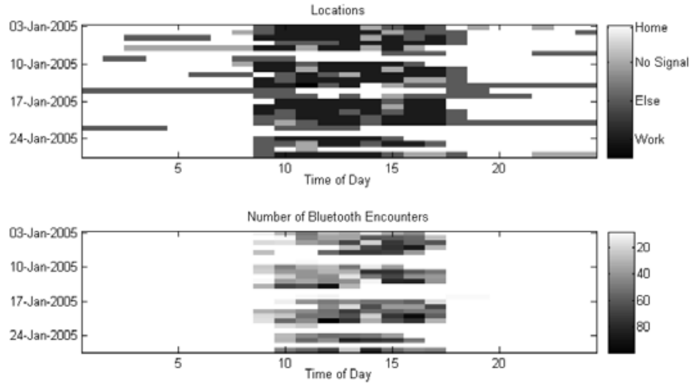

Fig 4. Subject 9's 'low entropy' daily distribution of home/work transitions and Bluetooth devices. The 'hot spot' in mid-day is when the subject is at the workplace.

Entropy metric: a low entropy subject

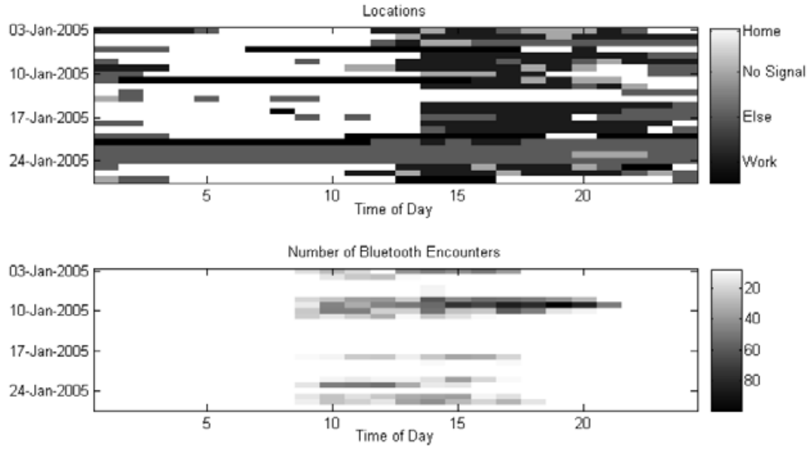

Fig 5. Subject 4's 'high entropy' daily distribution of home/work transitions and Bluetooth devices.

Model of patterns: inference about activity

- Used Hidden Markov Chains to model location transitions (based on time) and probabilities inferred
  - Expectation-maximization inference used to infer probabilities
  - 95% accuracy (states: home/work/elsewhere, 1 month of date on several subjects)

Experiences in data gathering

- Data corruption: storing on flash memory card, finite number of read-write cycles
  - Initial prototype wrote on same cell: failure after 1month
- Bluetooth errors:
  - people close to each others but on different sides of a wall :)
  - Frequency of scanning impacts type of encounters logged
  - Small crash percentage
- Human errors:
  - Turn phone off/ battery exhaustion
  - Forgot phone somewhere :)..difficult to extract these from data sets
    - What is different between forgetting a phone home and being sick?
  - Move to another office without the phone (difficult to detect)
- Use of surveys to complement

Social Network Modelling

- Logging can lead to models of social behaviour of a network of users
  - Application usage
  - Epidemiology
- location prediction of low entropy subjects can be done with 90% accuracy
- Most likely time and place to meet colleagues or friends
- [also non ethical uses]

Relationship inference

- Time of proximity may mean different kind of relationships
- However also just proximity conveys some info
- Surveys were used to ask about friendship info
- Gaussian mixture model used to detect patterns in the proximity of users
- Identification of friends with 90% accuracy

Friends and proximity nets

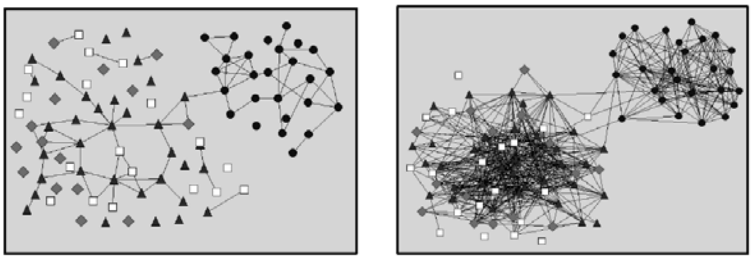

Fig 10. Friendship (left) and daily proximity (right) networks share similar structure. Circles represent incoming Sloan business school students. Triangles, diamonds and squares represent senior students, incoming students, and faculty/staff/freshman at the Media Lab.

Friends vs Acquaintance

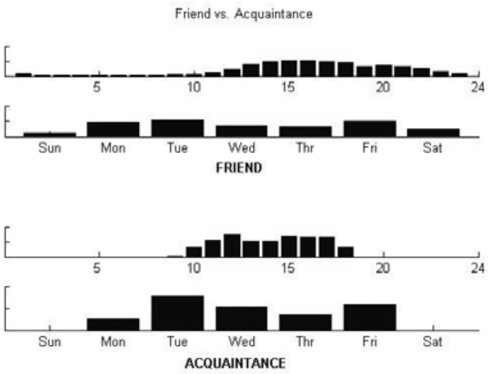

Fig 11. Plotted is proximity frequency data for a friend and a workplace acquaintance of one subject.

Network Dynamics

- Change in behaviour
  - Due to a deadline the patterns of some people (75) changed for a month
  - Time at night to finish off something ;)

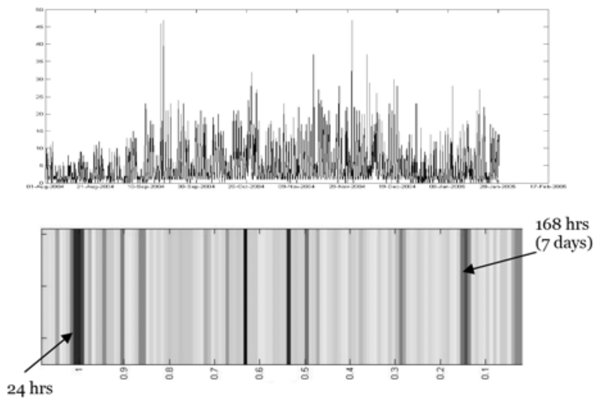

Fig 14. The total number of edges each hour in the Media Lab proximity network from August 2004 to January 2005. Below is its corresponding Fourier transform confirming the two most fundamental frequencies of the dynamic network to be (not surprisingly) 1 day and 7 days.

Discussion

- Lots of data collection happening
  - Humans via Bluetooth
  - Animal encounters
  - Transport traces
- Analysis of these traces of connectivity/mobility is difficult
- The actual gathering of the traces can be conditioned by some of the sensing choices made
- The kind of information gathered might be highly sensitive

Inter-contact time analysis

- Traces of encounters of bluetooth devices or wireless ones

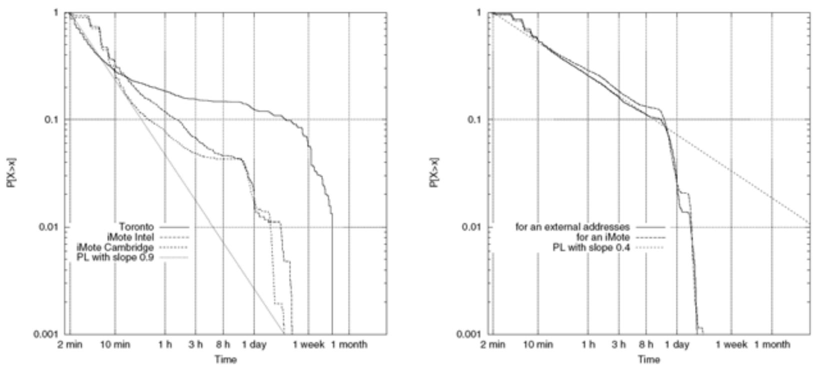

Fig. 1. Tail distribution function of the inter-contact time in six experiments: iMote-based experiment at Intel and Cambridge, and Toronto experiment (left), iMote-based experiment at INFOCOM (right).

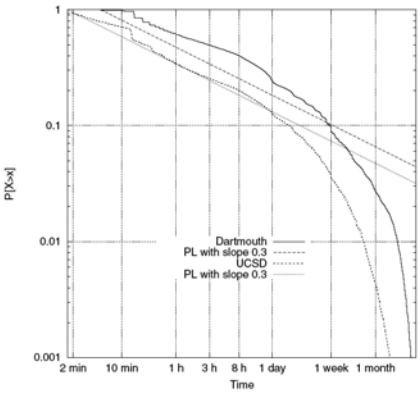

Fig. 2. Tail distribution function of the inter-contact time: data collected in Dartmouth and UCSD.

Mobility Model

- Random mobility model exhibit exponential decay behaviours
- Different from traces!!
- Most forwarding protocols assume exponential decay in the inter-contact time.
- In case of heavy tail inter contact time behaviour they have infinite expected delay

Other sources

- http://crawdad.cs.dartmouth.edu/
- Augustin Chaintreau, Pan Hui, Jon Crowcroft, Christophe Diot, Richard Gass, and James Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM), Barcelona, Spain, April 2006.
- Tristan Henderson, David Kotz, and Ilya Abyzov. The changing usage of a mature campus-wide wireless network. Proceedings of the Tenth Annual International Conference on Mobile Computing and Networking (MobiCom), pages 187-201, Philadelphia, PA, USA, September 2004. ACM Press.

Social Mobility Model

- Used social theory to build a highly clustered mobility model
  - People tend to stay in cluster
  - Sometimes they are attracted to other clusters and move there
- Evaluated that the behaviour of the traces generated by the model exhibits same power law behaviour of the real traces

Problems

- Problems we have with the model
- What are the inputs? Social attractivity matrix?

- Maybe use the reality mining work to extract values for inputs