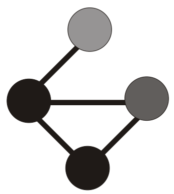
Location Models and Their Cell Phone Applications

May 31st, 2005
Seminar: Distributed Systems

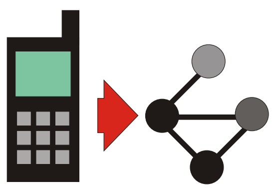
Gabor Cselle
gabor@student.ethz.ch

Advisor: Christian Frank

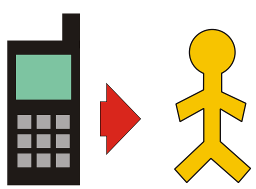
Overview



1. An introduction to location models



2. Automatic identification of locations on cell phones



3. Detecting human behavior patterns with cell phone data

1. Introduction to Location Models

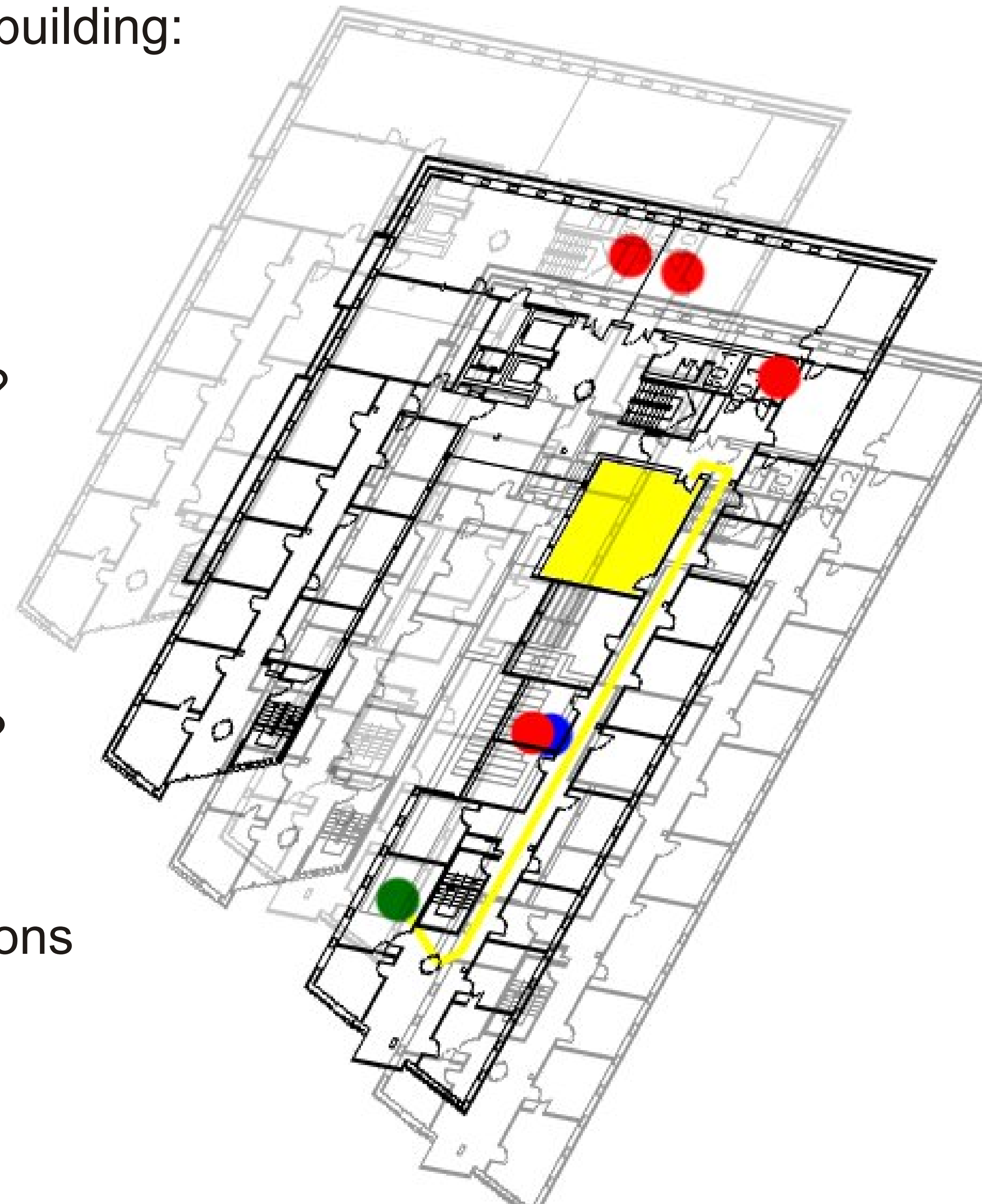
Questions we can ask in an office building:

Position queries:
Where am I?

Nearest neighbor queries:
Where is the nearest printer?

Navigation:
How do I get to room C42?

Range queries:
What printers are on floor C?



Challenge: Find data models so you're able to answer these questions quickly and efficiently.

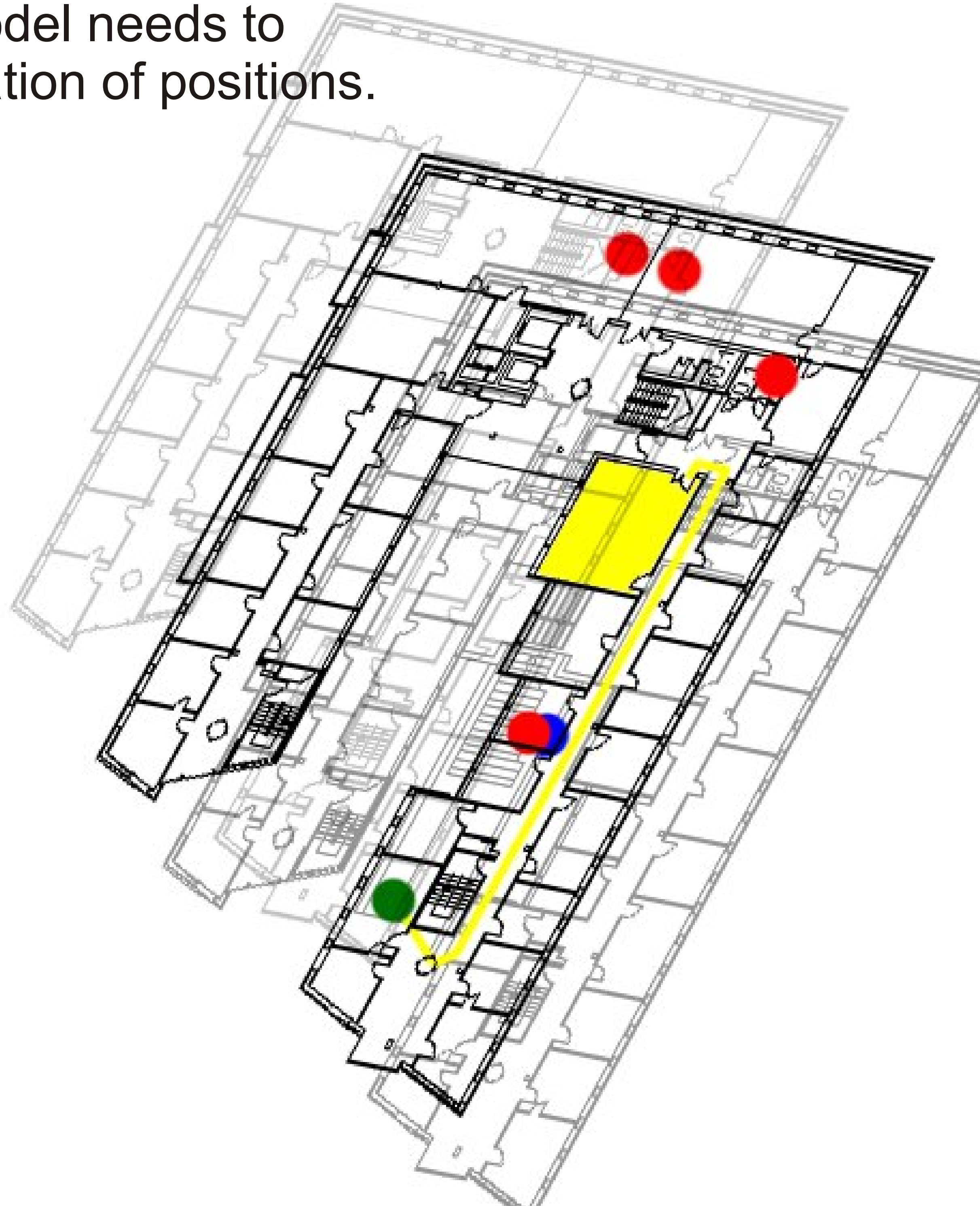
Requirements

For many common queries, the model needs to support more than simple identification of positions.

Nearest neighbor queries:
Where is the nearest printer?
➡ We need a notion of **distance**

Navigation queries:
How do I get to room C42?
➡ We need a notion of **connectedness**

Range queries:
What printers are floor C?
➡ We need a notion of **containment**



Why GPS isn't Enough

Let's ask:
Where is the nearest printer on floor C?

Global positioning could give us:

You are at: 8°15'E, 37°2'N, 424m

A database could give us nearest printers according to Euclidean 3D distance.

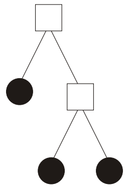
Printer 1: **8°16'E, 37°4'N, 427m**
Printer 2: **8°12'E, 37°2'N, 421m**
Printer 3: **8°15'E, 37°3'N, 424m**

But how would we know:
how easy it is to get to printers?
- lacking distance/connectedness data

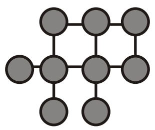
if they're really on floor C
- lacking containment data

Image source: NASA

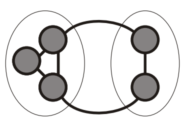
Symbolic Location Models



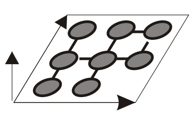
A. Hierarchical models



B. Graph-based models

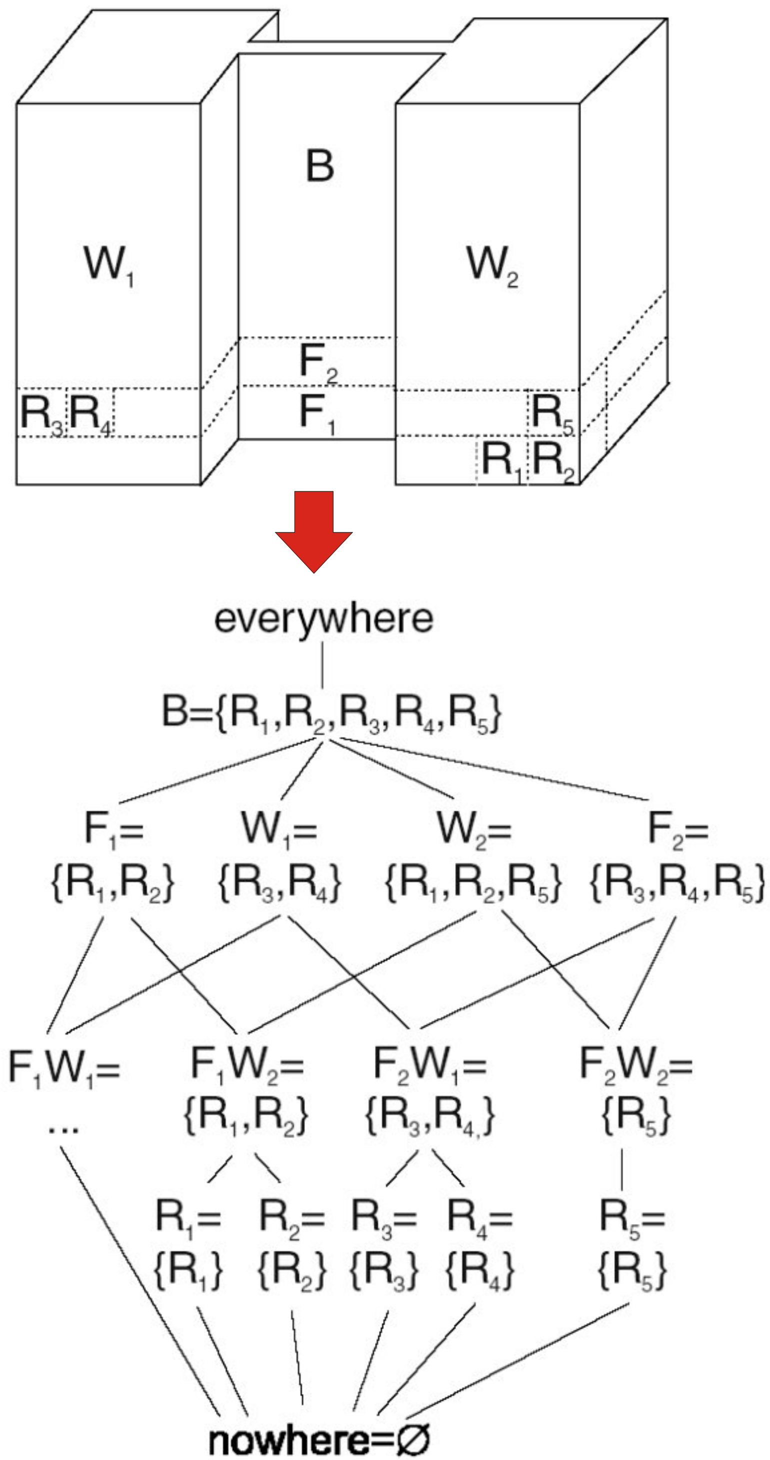


C. Graph- and set-based models



D. Subspace models

A. Hierarchical Location Models



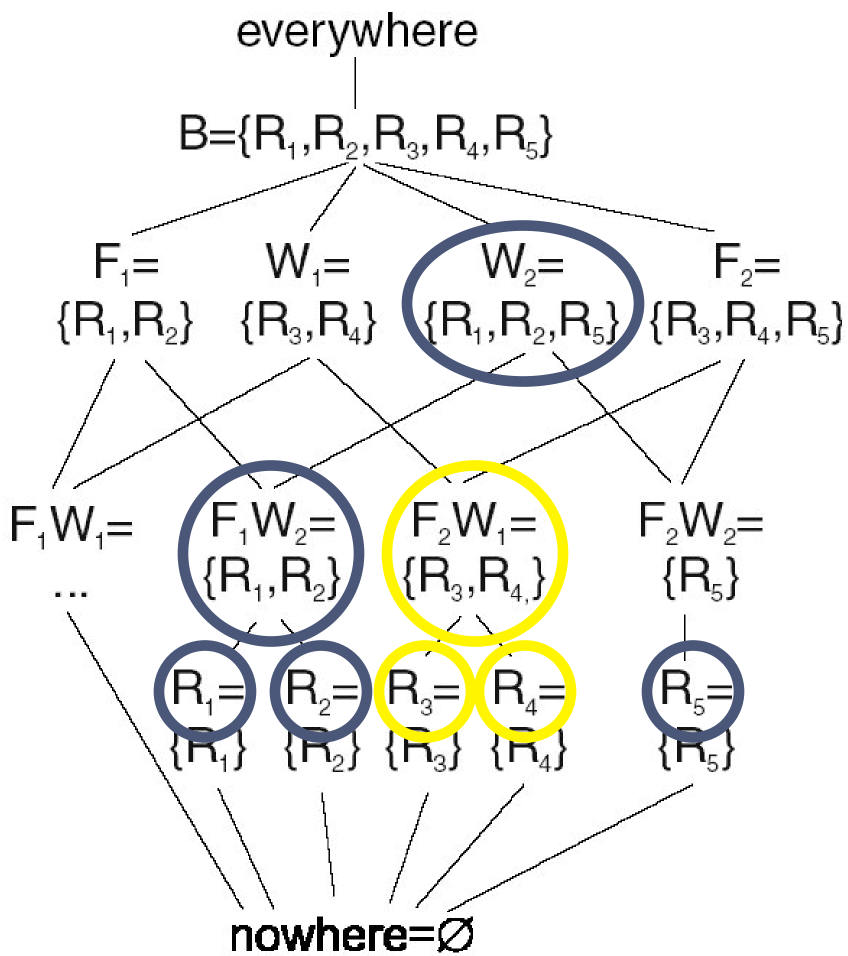
We group rooms R_i by:

- Building B
- Wing W_1 / W_2
- Floor F_1 / F_2 / ...

Create sets for each group: Add all rooms contained in them.
For overlapping groups, we need to a set for every combination of them. (F_1W_1, F_1W_2, ...)

This results in a lattice with the property:
A location l1 is an ancestor of a location l2 if l2 is spatially contained in l1.

A. Hierarchical Location Models



Evaluation:

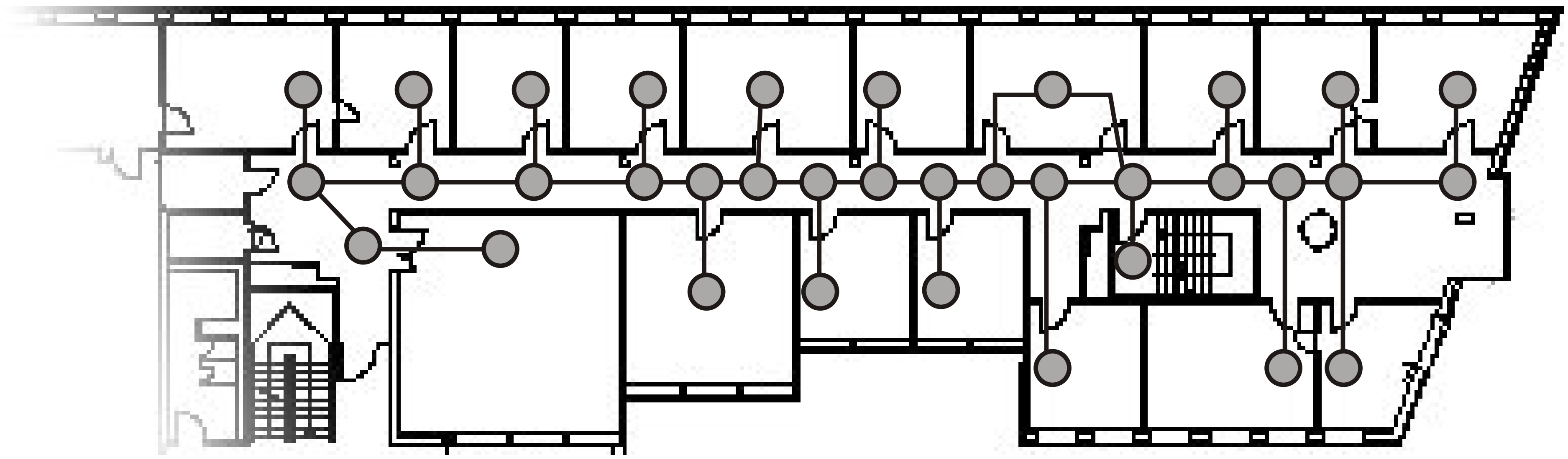
Unreliable distance queries only:
R_1, R_2 have closer common ancestor than R_1, R_5
➔ R_1 closer to R_2 than to R_5

Unreliable connectedness queries only:
R_1, R_2 in have common superset
➔ R_1, R_2 are neighbors"

Great for containment queries

B. Graph-Based Models

Use vertices to represent rooms
Use edges to represent connections
Edges may be weighted to model distances

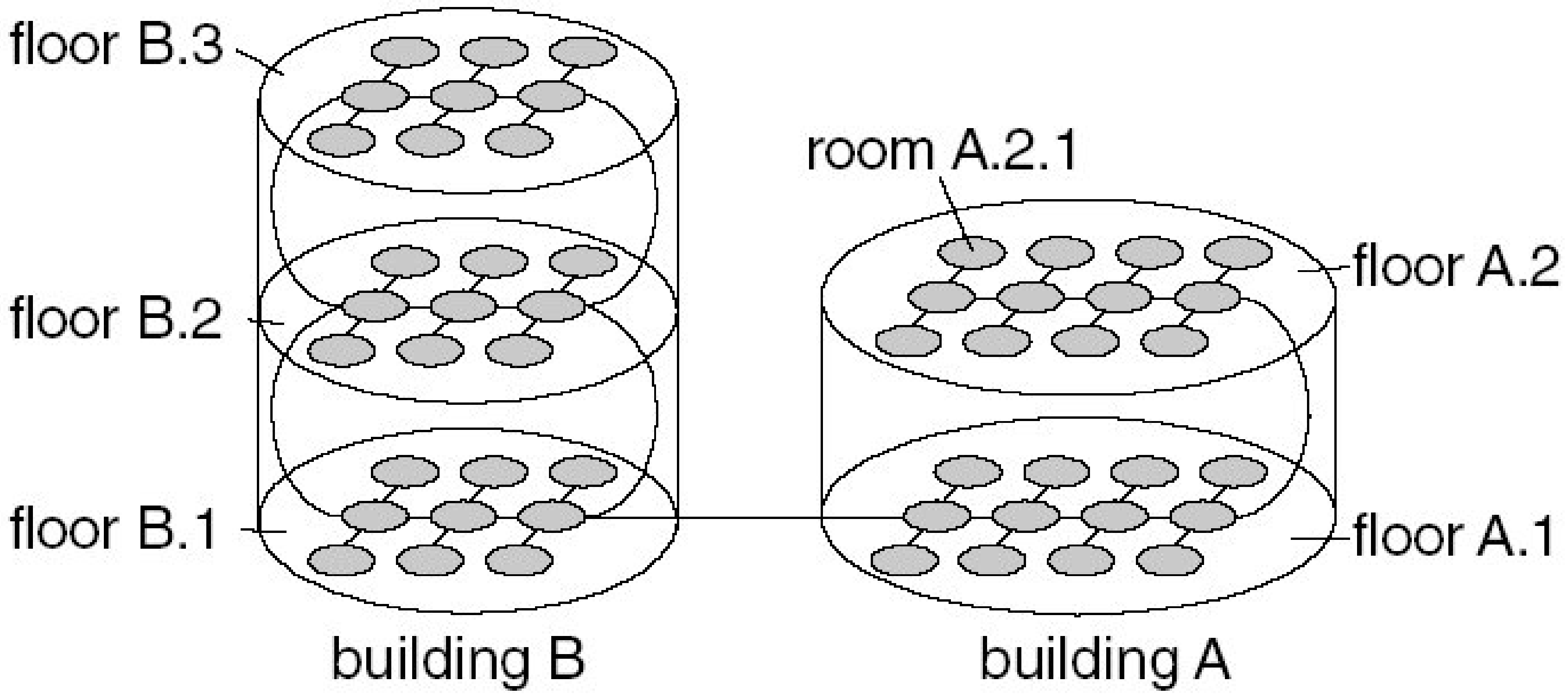


Evaluation:

- Distance queries are easy
- Connectedness queries are easy
- Containment queries hard:
  Given a room on C floor, we can find closeby rooms in graph:
  ➡ they are likely to be on C floor also

C. Graph- & Set-Based Models

Idea: Take subgraphs of the total location graphs, stick them into sets identifying related locations.

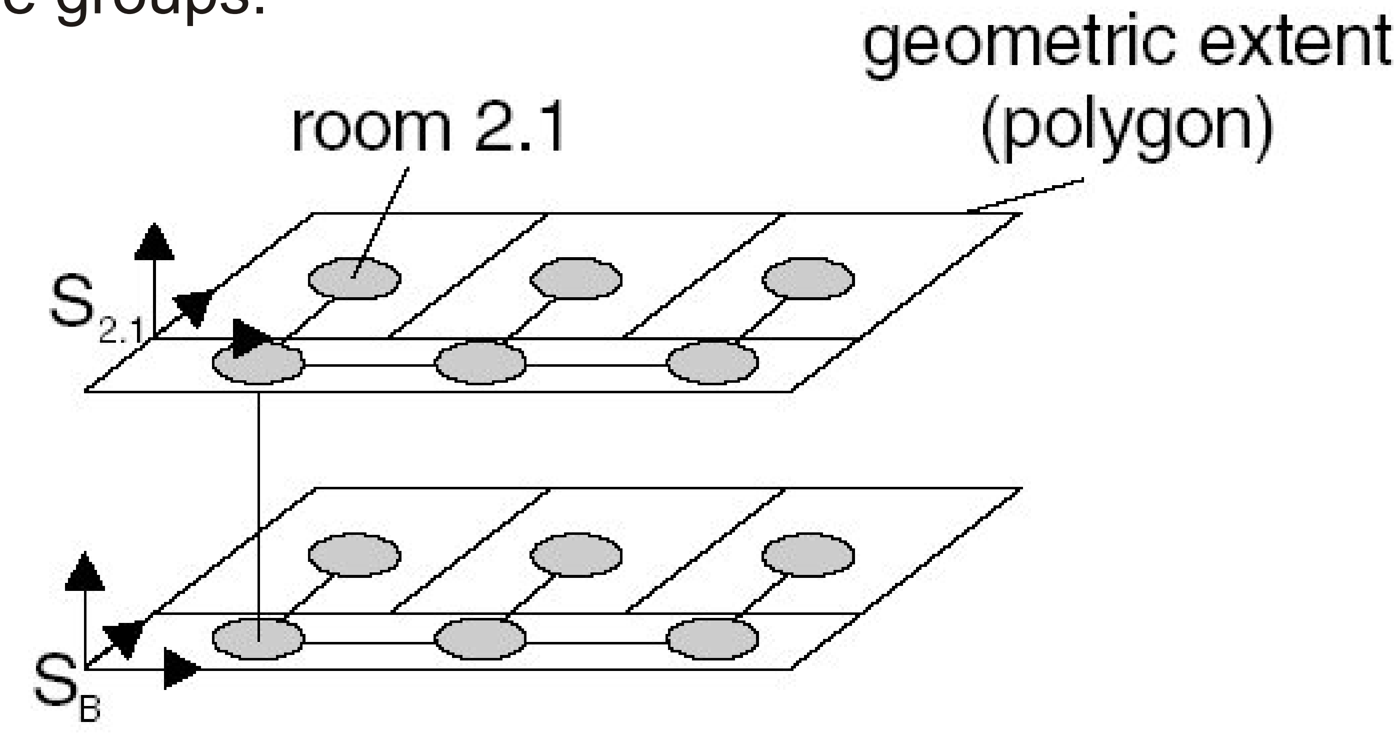


Evaluation:

- Containment queries much easier than with graph-based models

D. Subspace Models

Idea: Group into subgraphs as before, but attach geographic extent to each of the groups.



Evaluation:

- Distance queries are easy
- Connectedness queries are easy
- Containment queries are easy

+ A big plus: Can estimate position in space

Power Comes at a Price

	Distance support	Connectedness support	Containment support	Modelling effort
Hierarchical	⊖	⊖	⊕⊕	○
Graph	⊕	⊕	⊖	○
Graph+Set	⊕	⊕	⊕	⊖
Subspaces	⊕⊕	⊕	⊕	⊖⊖

As model's power grows ...

... so does the modelling effort

2. Automatic Location Identification on Cell Phones

With PlaceLab, we can see how mobile end devices can be used to get geographic coordinates using a base station database.

But:

Sometimes, there is no base station data for the current location.

Instead of coordinate data (8°15' E, 37°2' N), user would like to see its description:

"Home"
"Work"
"Coffee shop"

Input: Timestamps & Tower IDs

Install special software on cell phones that records changes of the primary cell tower along with a time stamp

We get:

Problems:

No one-to-one correspondence between physical location and cell used.

Cells can be very large or very small.

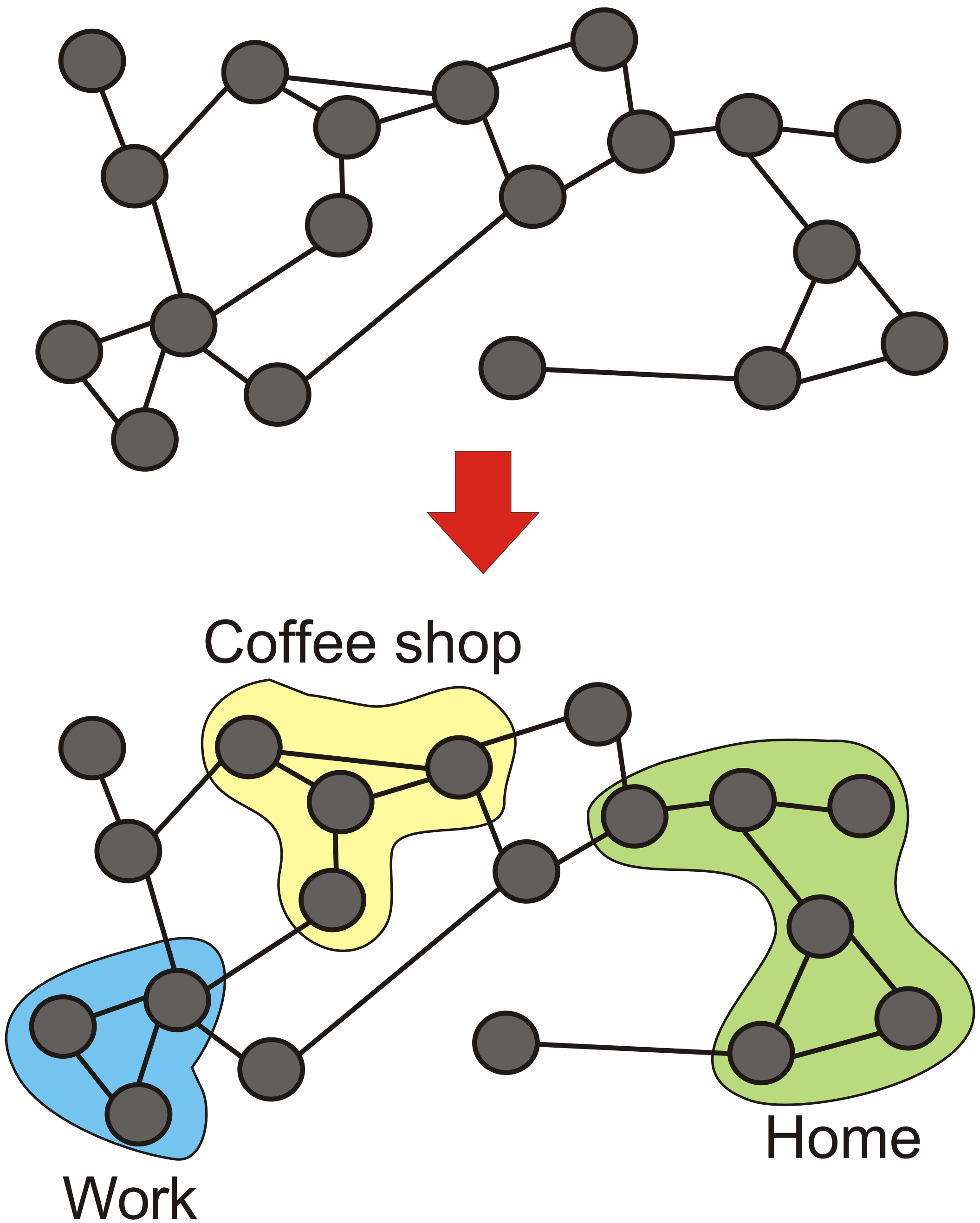
Areas covered by cells can overlap.

Cells can be non-contiguous areas.

Cell Graph

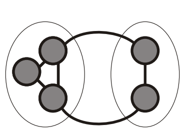
Create a graph:

- vertices = observed GSM cells
- edges = observed transitions between two GSM cells



The Goal:

- group GSM cells into sets representing "bases"
- each base represents a physical location where user spends a lot of time



We're building a graph & set-based location model

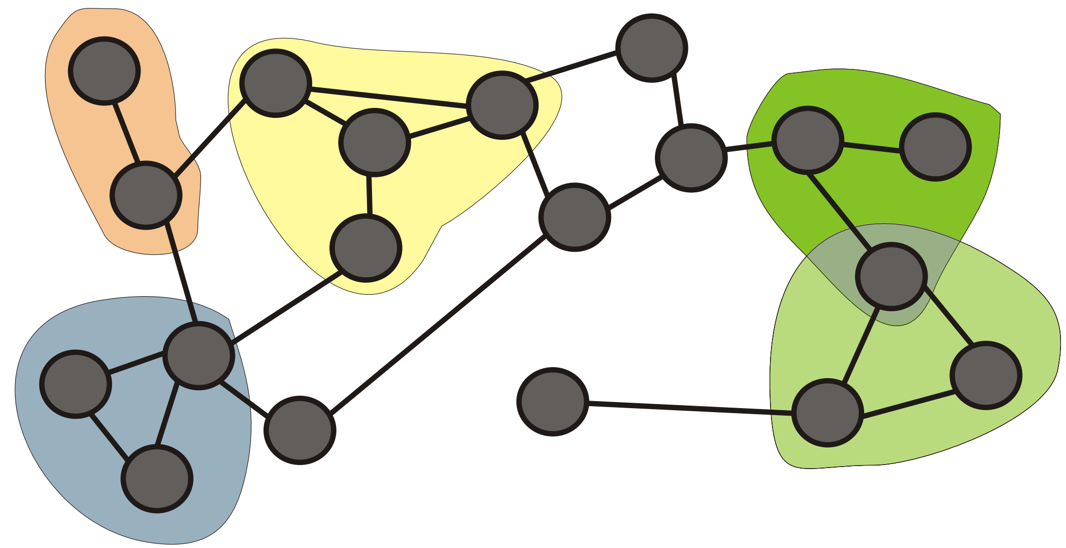
Identifying Bases

Step 1: Find Clusters

Required properties:

- subgraphs with max. diameter 2
- average time spent visiting a cluster is larger than sum of individual visit times

=> Fulfilled only when user oscillates between cells in cluster



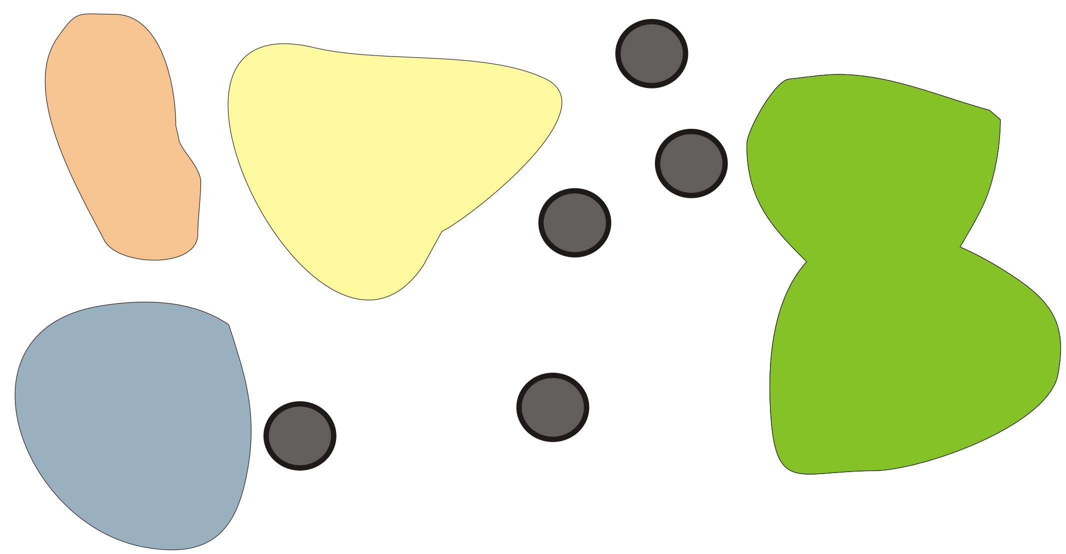
Step 2: Create Location Set $\mathfrak{L}$

Merge overlapping clusters

Location set $\mathfrak{L}$ now contains:

Merged clusters

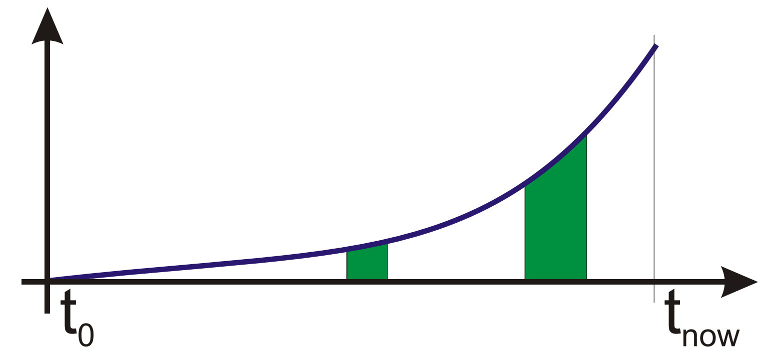
+ Individual vertices not contained in clusters



Identifying Bases

Step 3: Calculate (weighted) time spent in each location L

$$time(L) = \int_{t_0}^{t_{now}} at_L(t)\, r^{t_{now}-t}\,\mathrm{d}t$$



$at_L(t)$: indicator function: 1 if user is in location L at time t, 0 else

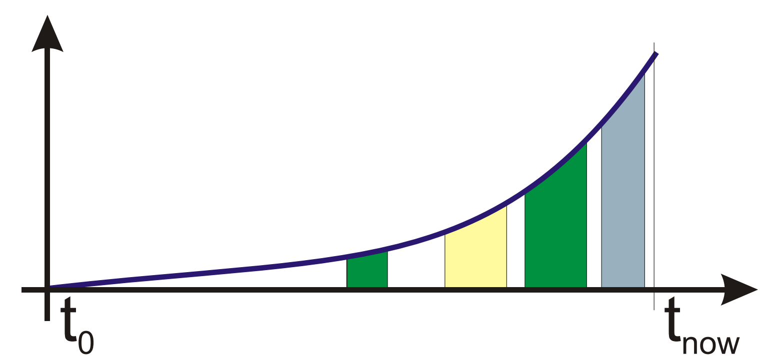
r: aging factor: 0.95

Exponential weighting of past times when we were at a location

Step 4: Identify minimal set of locations

These locations must cover fraction p of time

$$B = \arg\min_{B' \subseteq \mathfrak{L}} |B'| : \sum_{L \in B'} time(L) \geq p \int_{t_0}^{t_{now}} r^{t_{now}-t}\,\mathrm{d}t$$

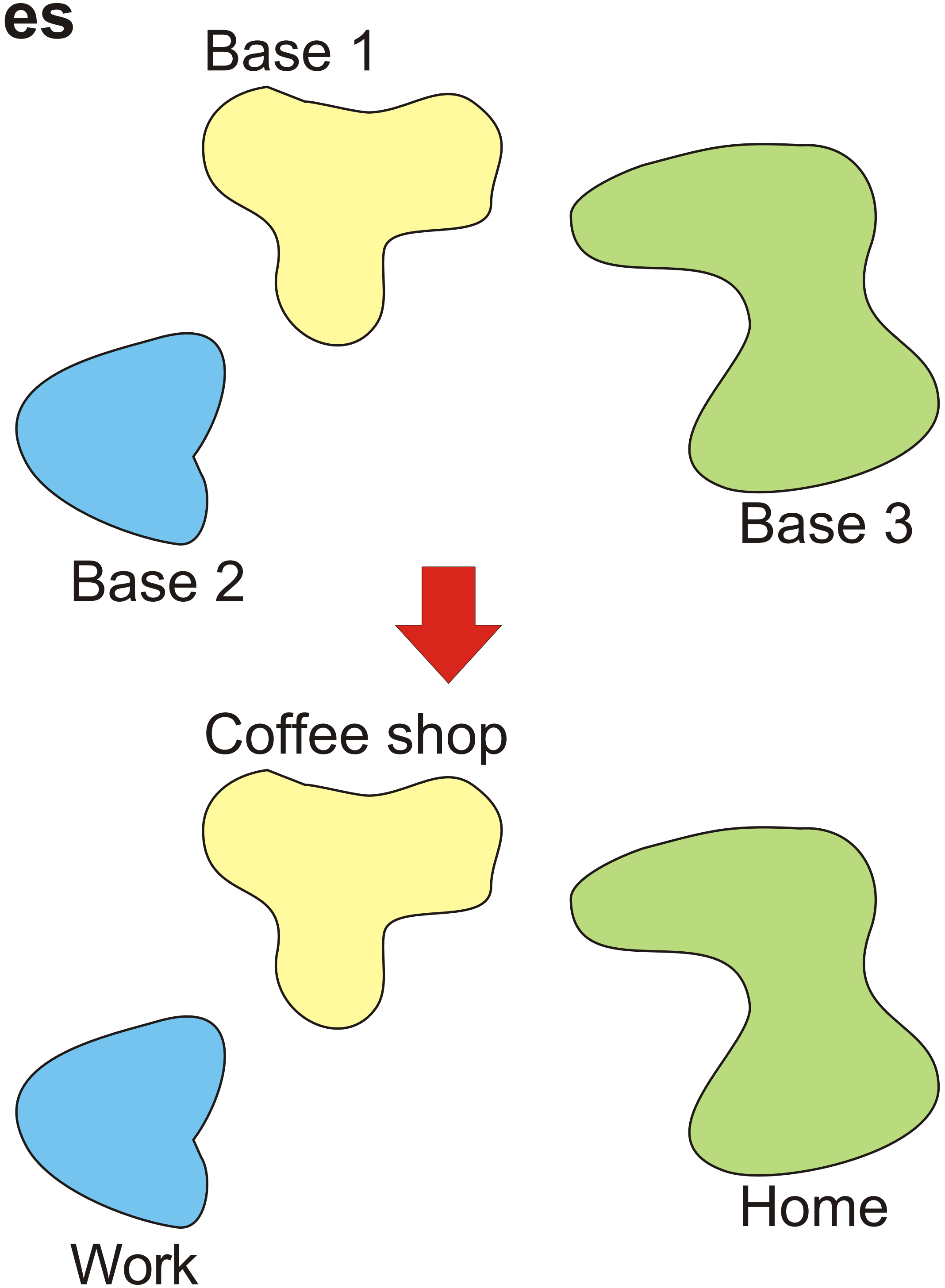


Identifying Bases: Naming

Step 5: User must name bases

We now have identified bases where the user spends a lot of time.

However, we don't know the meaning of these bases. The user must **manually assign names**.



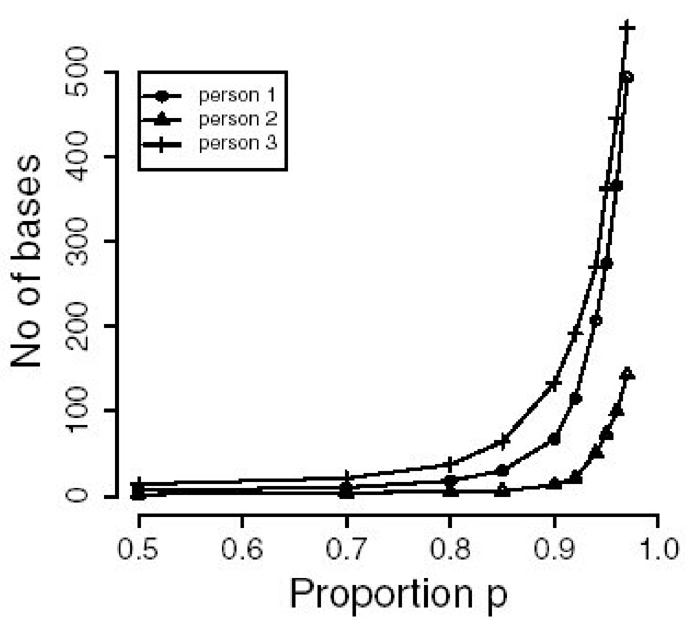
Base Identification Results

Identified bases for one of the test users.

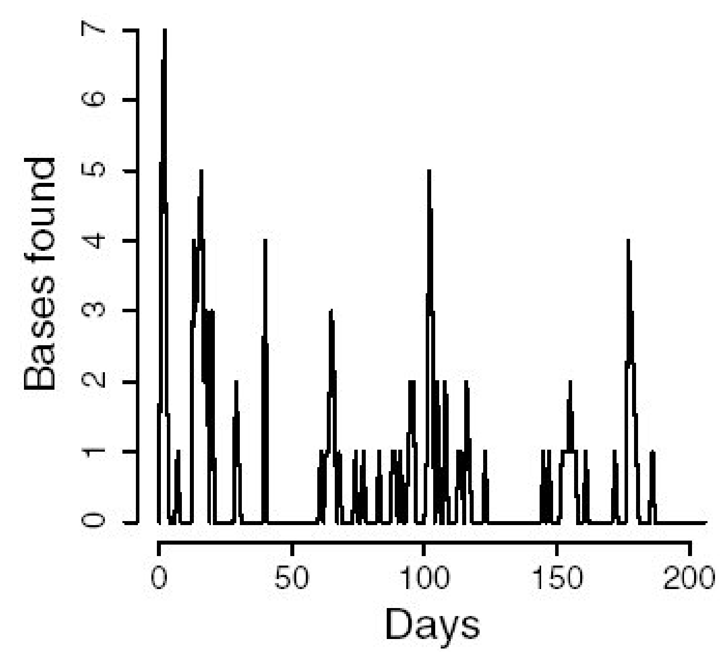
Home	Work
Friends' home	Girlfriend's home
Parents	Girlfriend's parents
Shopping center	Vammala town center
Girlfriend's summer cottage	Student association house
Viitasaari accommodation	Family firm office
Summer cottage	Helsinki center west

• • •

Number of bases found with for different p



Number of bases to manually name per day during test



Possible Uses

Reno: Answering a location request by curious wife.

Automatically generate list of likely current locations

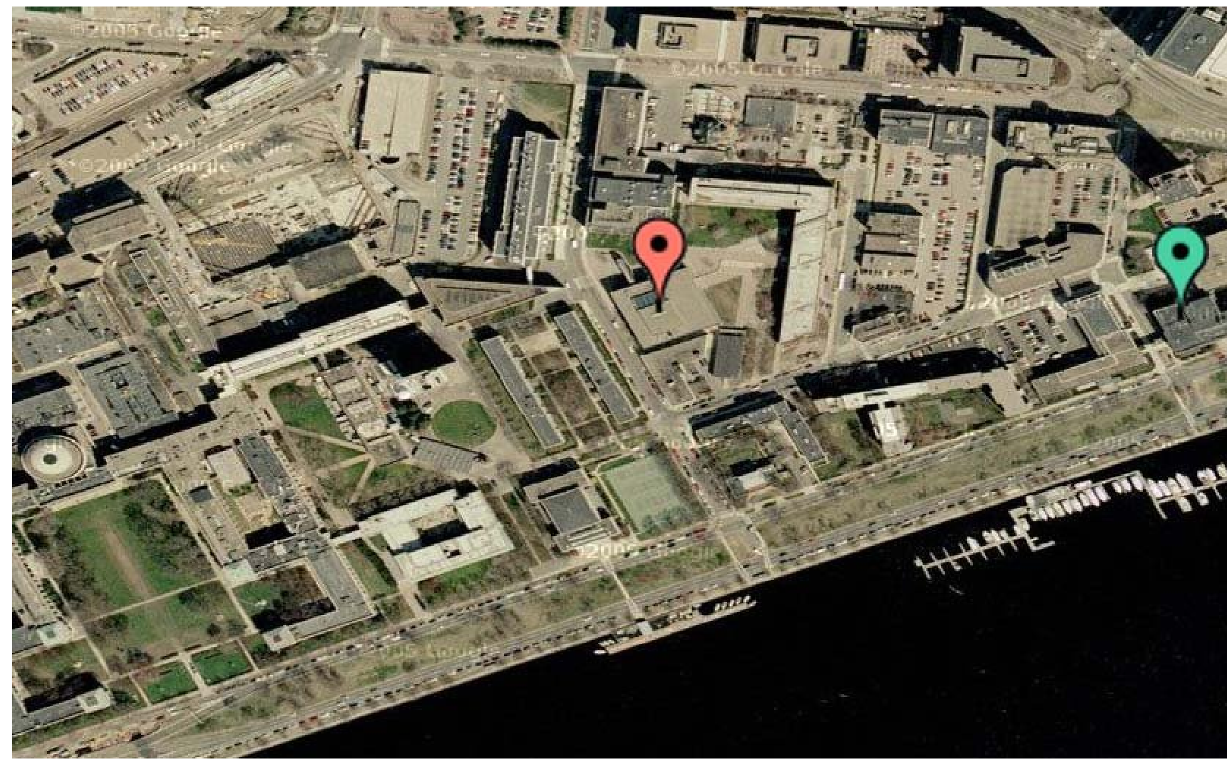
Dodgeball / Google: Instead of your having to send a manual login SMS, we could automatically infer which bar you're at.

3. Detecting Human Behavior Patterns with Cell Phone Data

Big data collection experiment with 100 cell phones:

MIT Media Lab students / faculty

MIT Sloan School (business school) MBA students

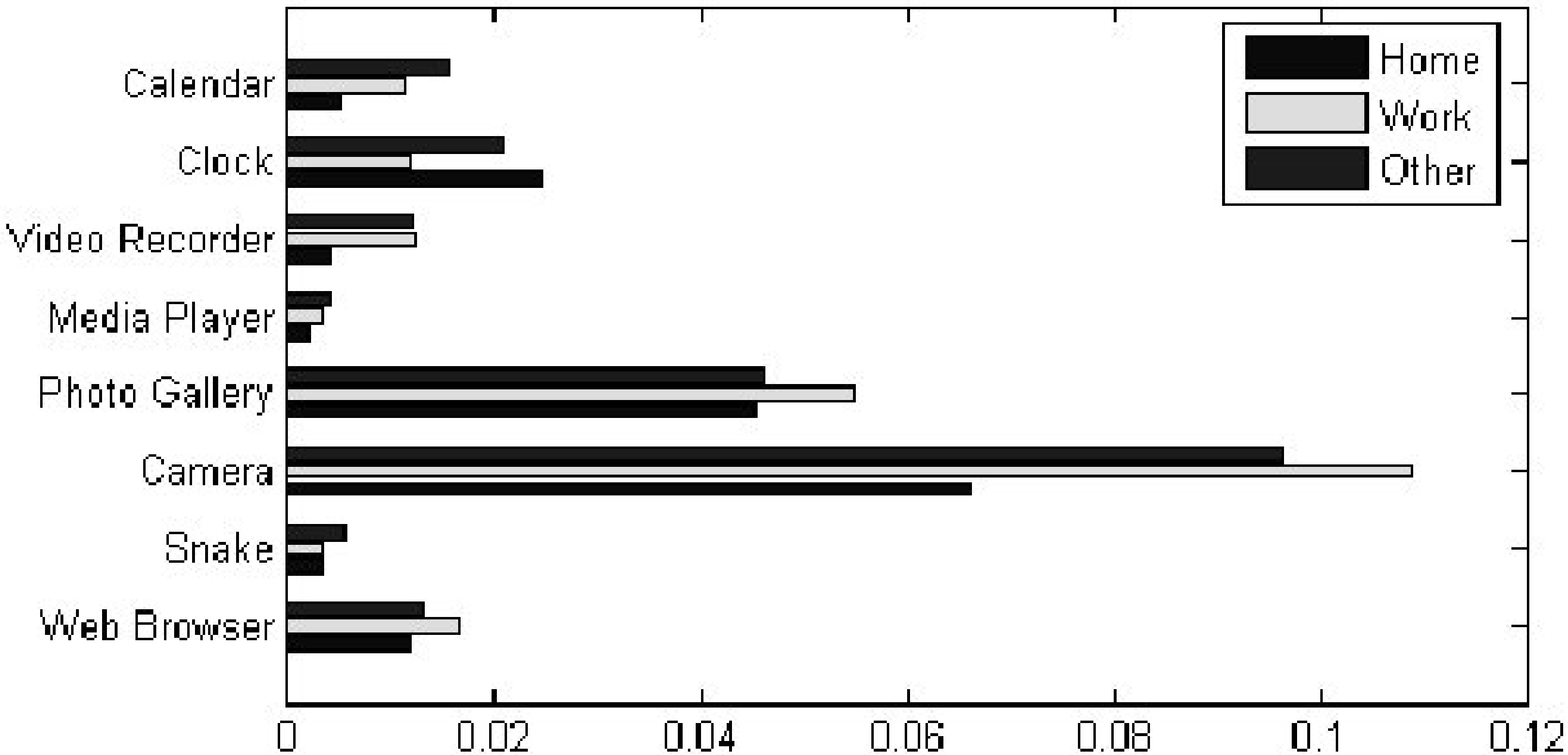


Satellite image source: maps.google.com

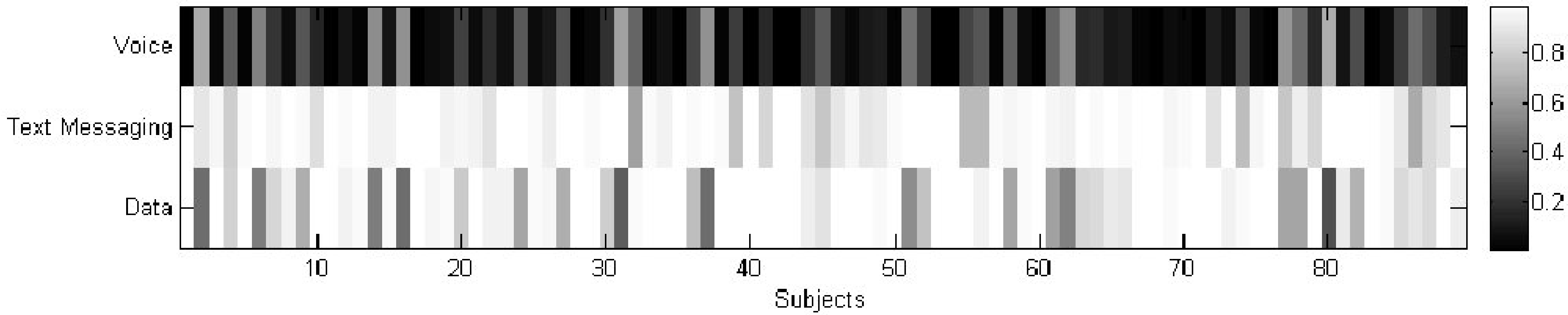
Locations determined using cell tower ID and Bluetooth.
Recorded on phone's memory card.
What can we find out using collected data?

On-Phone Application Usage

Aggregate Application use in Context

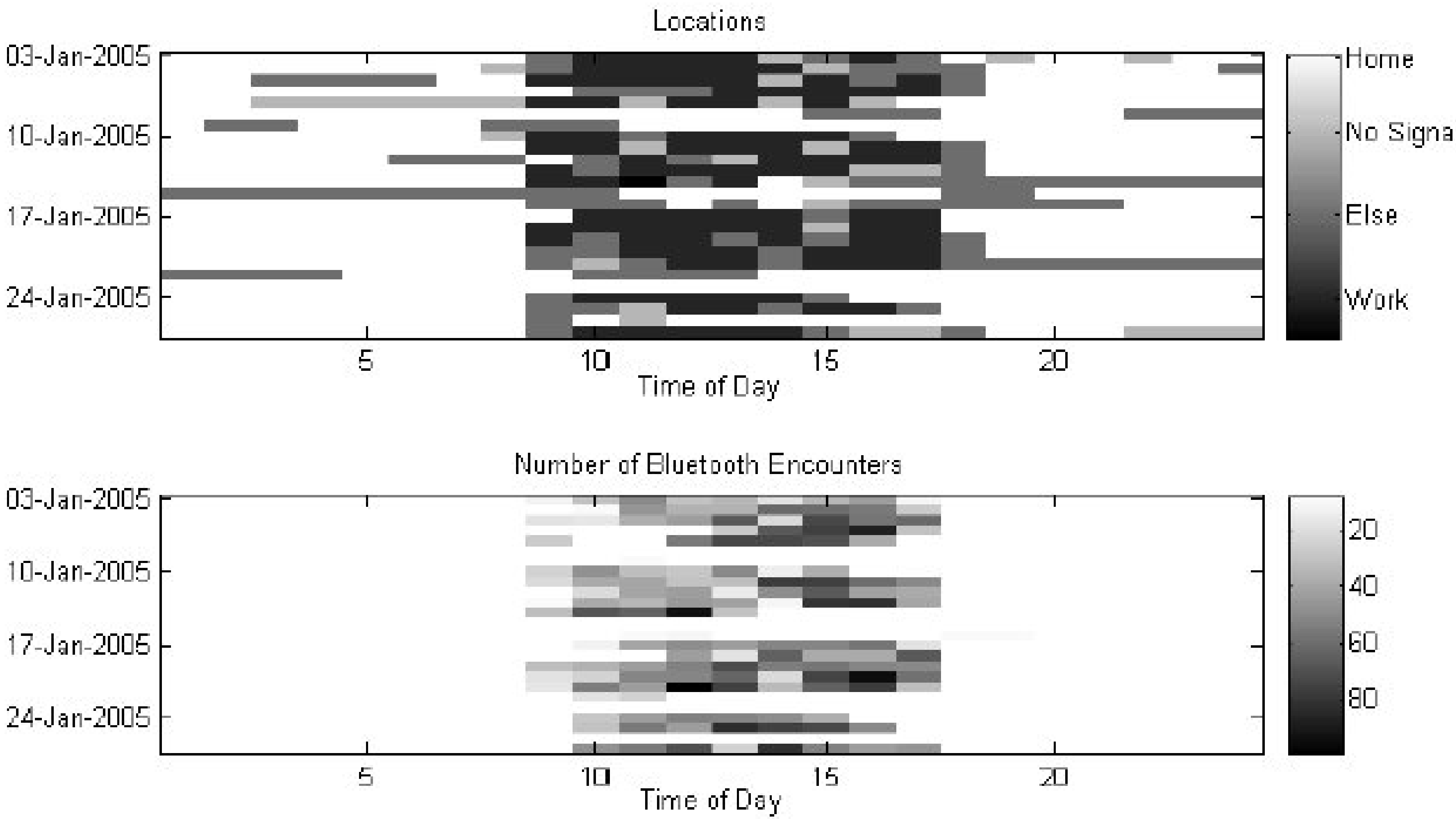


Communication Usage Patterns (%)



Location Patterns of Users

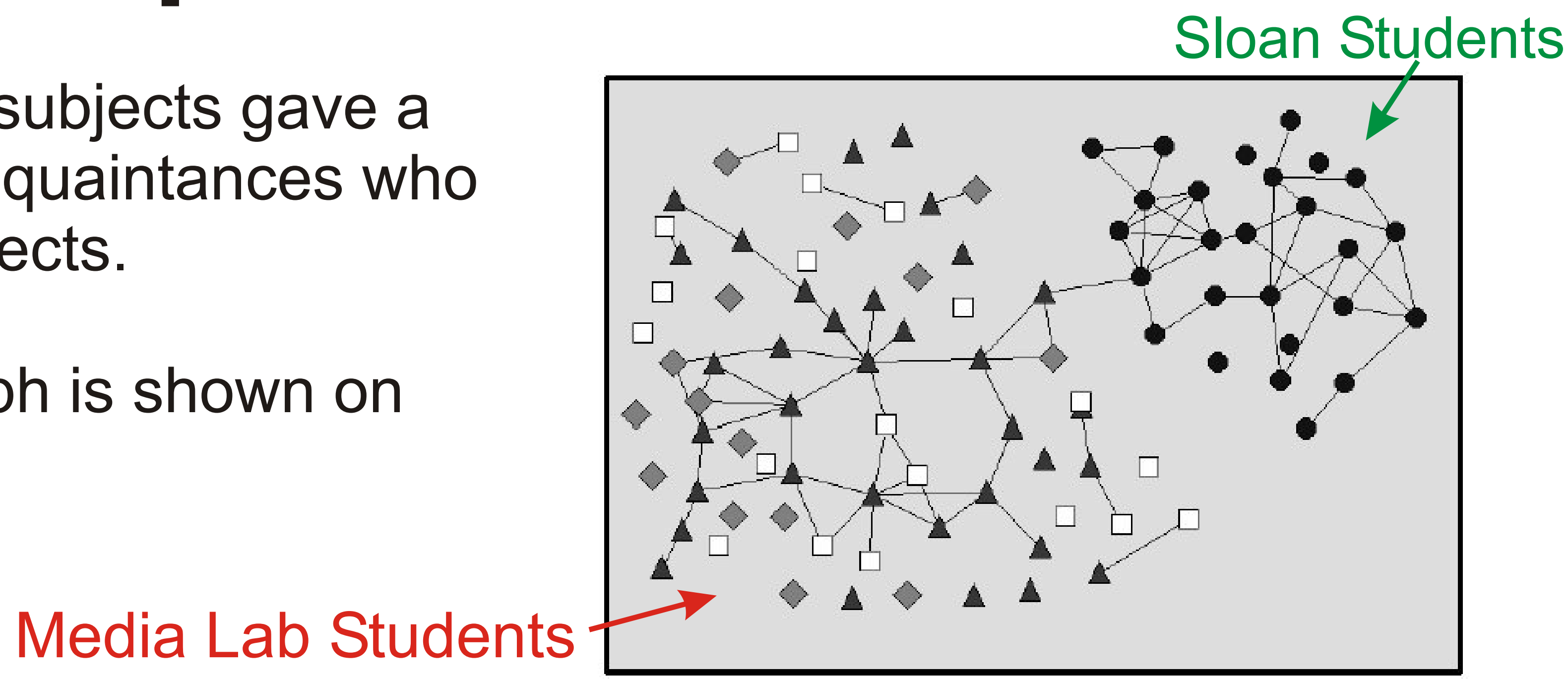
Daily distribution of home/work transitions and Bluetooth encounters for a 'low-entropy' user.



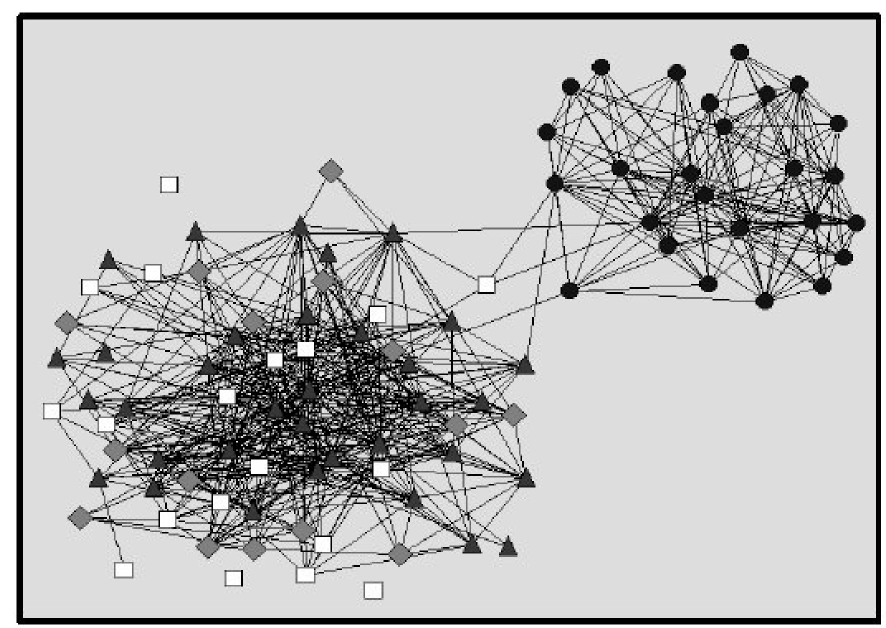
Relationship Inference

For the study, test subjects gave a list of friends and aquaintances who were also test subjects.

The friendship graph is shown on the right.

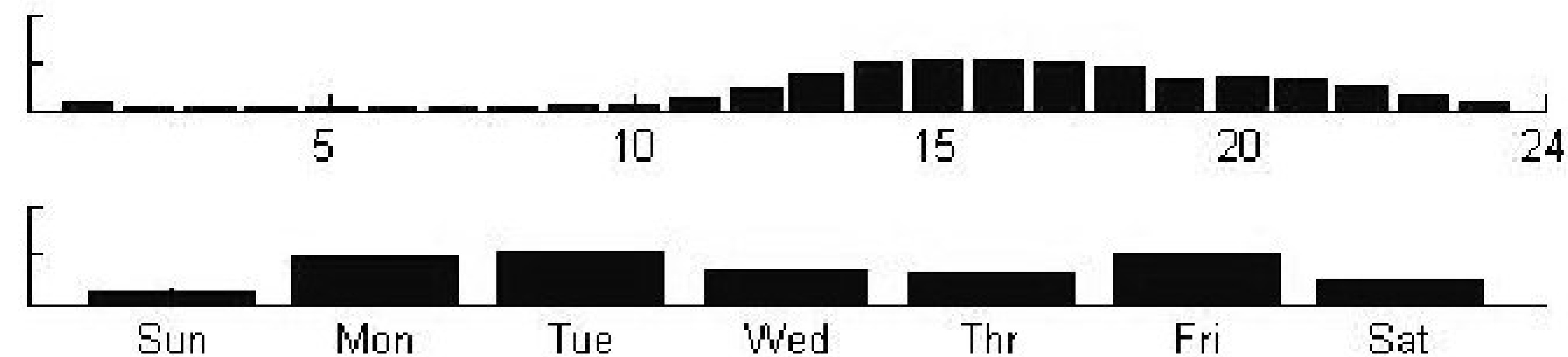


The proximity pattern graph has a similar structure to the friendship graph.

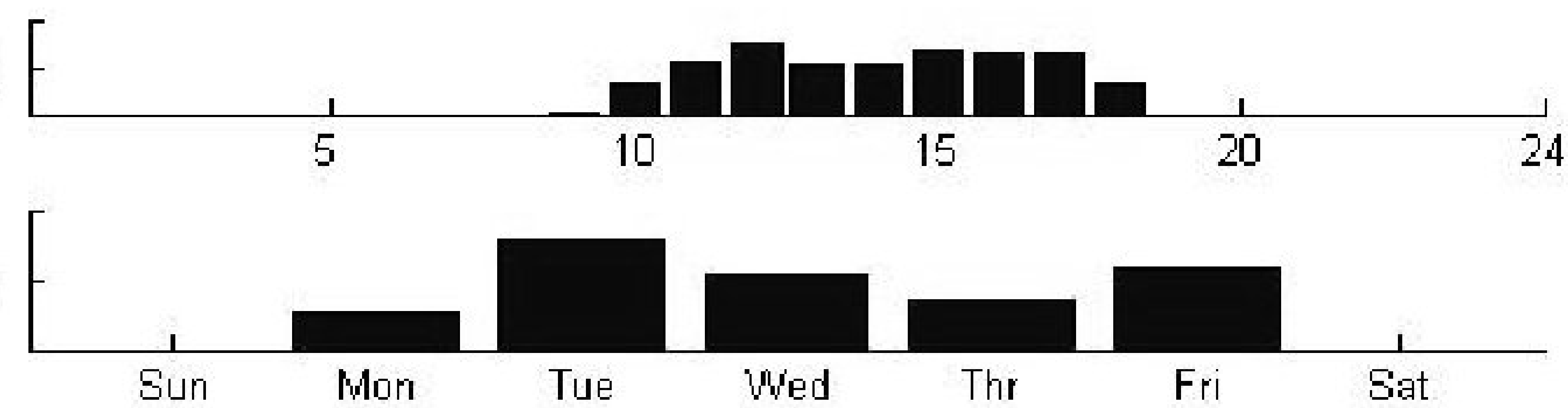


Friends vs. Acquaintances

Proximity frequencies depending on time, weekday and relationship.



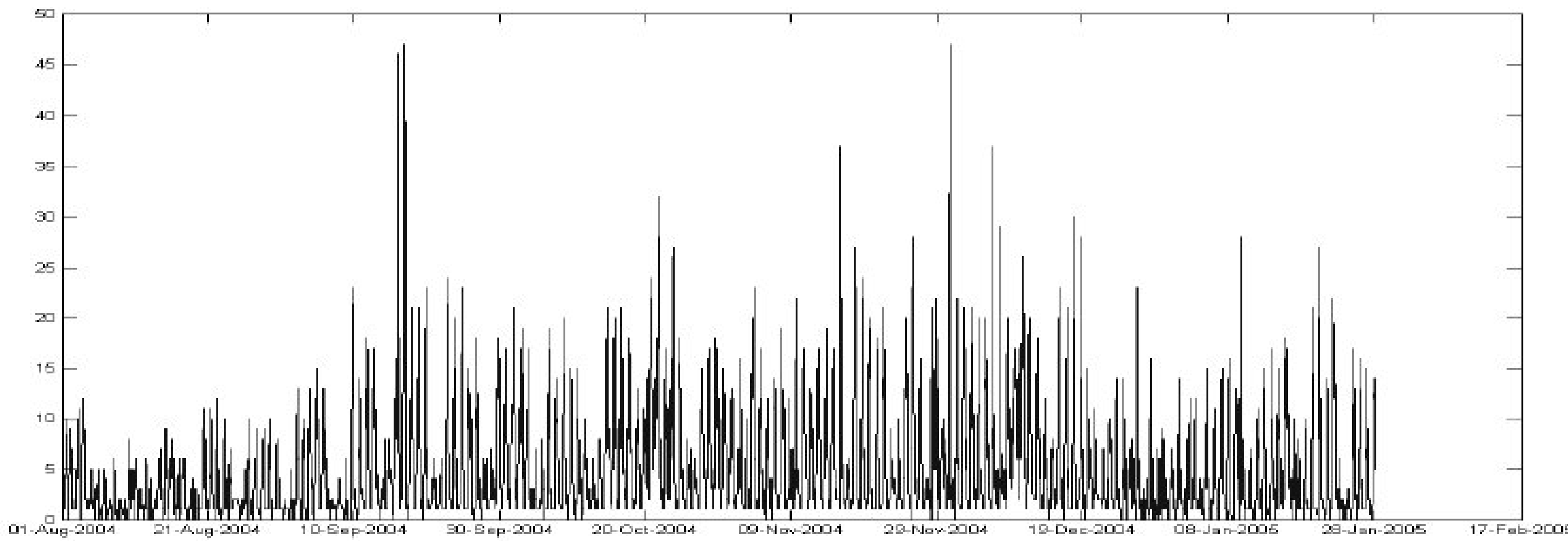
Friend



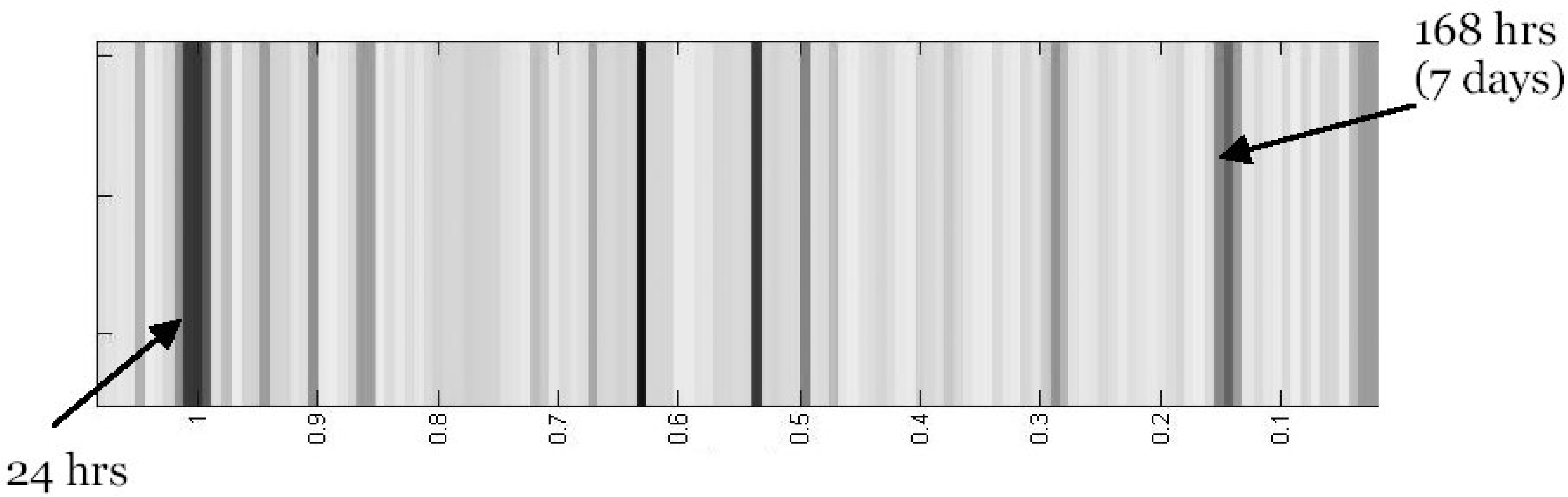
Aquaintance

Human Behavioral Patterns

Time series of maximum number of links in Media Lab proximity network during every one hour window.



And its Fourier transform ...



What do Participants Think?

From: "-----@sloan.mit.edu" <-----@sloan.mit.edu>
To: "gabor@student.ethz.ch" <gabor@student.ethz.ch>
CC: "-------@sloan.mit.edu" <-------@sloan.mit.edu>
Subject: RE: Do you know any reality mining participants?
Date: Mon, 30 May 2005 18:30:17 -0400

Hey Gabor,

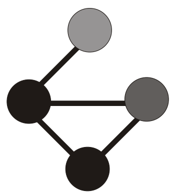
I participated in the cell phone study for the past two semesters. [...] As for as your questions:

I didn't mind any of the privacy ideas but I'm a pretty open gal. Also, keep in mind we received a brand new, top of the line, Nokia cell to participate so bit of an incentive to forgo any hang-ups on privacy.

We were never told about any of the data collected. We dropped the phones off once a month to do a "data dump" and were asked to fill out an on-line survey about every 3 months.

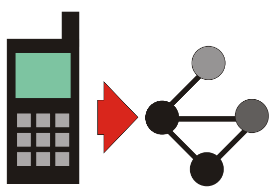
[...]
Best,

What We've Seen



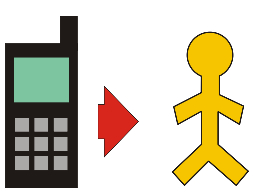
1. Location models

Powerful location models are available.
But: high modelling effort.



2. Automatic identification of locations on cell phones

Possible to infer location model for cell phone users.
Good accuracy of identified locations.



3. Detecting human behavior patterns with cell phone data

Once locations are identified and user's moves are recorded, interesting analyses can be performed.
But: privacy concerns.

References

[1] Summary of common location models:
Becker C, Durr F: **"On Location Models for Ubiquitous Computing"**
Personal and Ubiquitous Computing, Volume 9, Issue 1 (Jan 2005)

[2] Inferring bases from GSM tower switch data:
Laasonen K, et al: **"Adaptive On-Device Location Recognition"**
Pervasive 2004, Vienna, Austria

[3] Inferring human behavior from cell phone data:
Eagle N, Pentland A: **"Reality Mining: Sending Complex Social Systems"**
Personal and Ubiquitous Computing, to appear: June 2005

[4] Source of Reno usage example:
Smith I, et al: **"Social Disclosure of Place: From Location Technology to Communication Practices"**
Pervasive 2005

[5] Source of Dodgeball usage example:
http://www.dodgeball.com