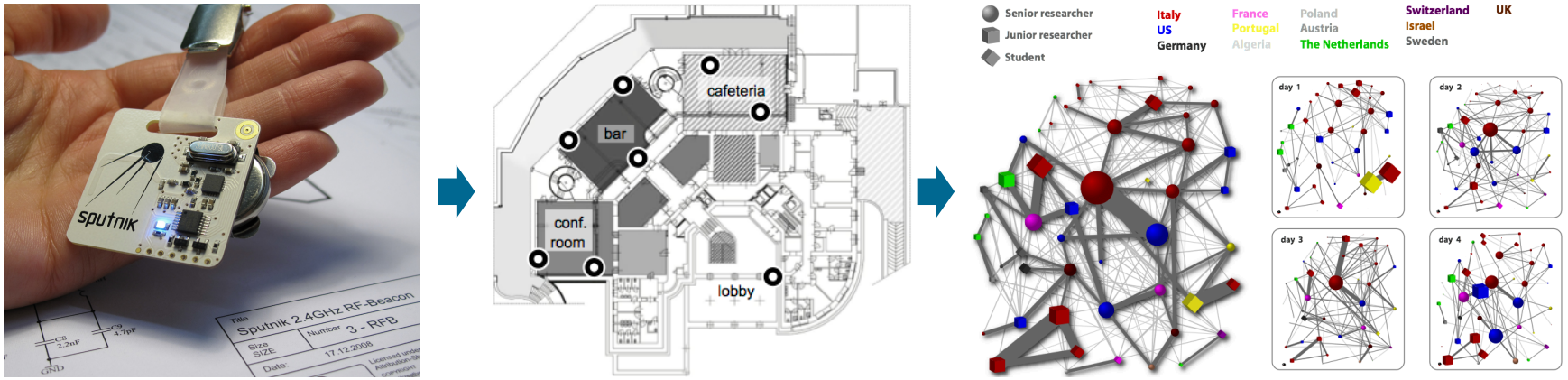
Inferring Social Networks from Sensor Data

Momin M. Malik, Jürgen Pfeffer, Yuvraj Agarwal

Sunbelt XXXIV
St. Pete Beach, Florida
February 22, 2014
Network Data Collection session, Saturday AM2

Motivation

- Sensors everywhere: consumer devices, infrastructure
- Rise of "social sensing"
- What are sensors?
- *How* do we use sensors to collect data?
- *Can* we really infer social networks from sensor data?



Cattuto et al., "Dynamics of person-to-person interactions from distributed RFID sensor networks," 2010
Barrat et al., "High resolution dynamical mappings of social interactions with active RFID," 2008

Method

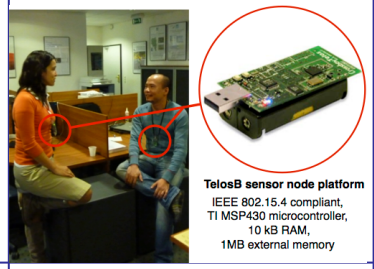
- Most literature on sensors within engineering
  - Literature search on IEEE Explorer, ACM Digital Library, for "sensor" + "social"
  - Specific search through relevant groups: Human Dynamics Group at MIT Media Lab, SocioPatterns
  - Focus on hardware, and what running a study with sensors practically involves
- Reviewed ~150 engineering papers
- Terms: social sensing, Social Signal Processing, ubiquitous computing, pervasive computing
- Reviewed classic SNA methodology literature for connections

What are sensors?

Engineering basics

- Sensors are devices that measure something physical and convert it into a signal
- Sometimes the measured physical quantity is simply the presence of another sensor. Bluetooth devices thus qualify as sensors
- By putting sensors on people (or on people plus in fixed locations), or using the sensor capabilities in devices they already have like smartphones, we can infer proximity (or co-location)
- The typical study undertaken with sensors focuses either on the hardware or the infrastructure of the sensors,
  - Some do deal with substantive results (Madan, Moturu, Lazer, Pentland, "Social Sensing," 2010)

Different sensors, pros and cons

Device	Examples	Protocols/ Channels	Pros	Cons
Wireless sensor nodes	iMote; RFID tag; WSN430 sensor node; Ubi-coin; TelosB nodes 	802.15.4 Wireless; Bluetooth; RFID; Radio	Relatively cheap; simplicity allows researcher control; power level allows calibration	Noisy; need to purchase; need to get people to wear; high power gets loose proximity but low power misses all but face-to-face; either bulky with battery, or need base stations
Smartphones	iPhone, HTC One, Samsung Galaxy, etc.	Bluetooth, Wireless, cell towers, GPS, accelerometer	Ubiquitous	Not necessarily as powerful/ controllable as custom devices
Sociometric badges		Infrared; audio; accelerometer; Bluetooth	Ready-made; redundancy of multiple channels increases accuracy	Privatized

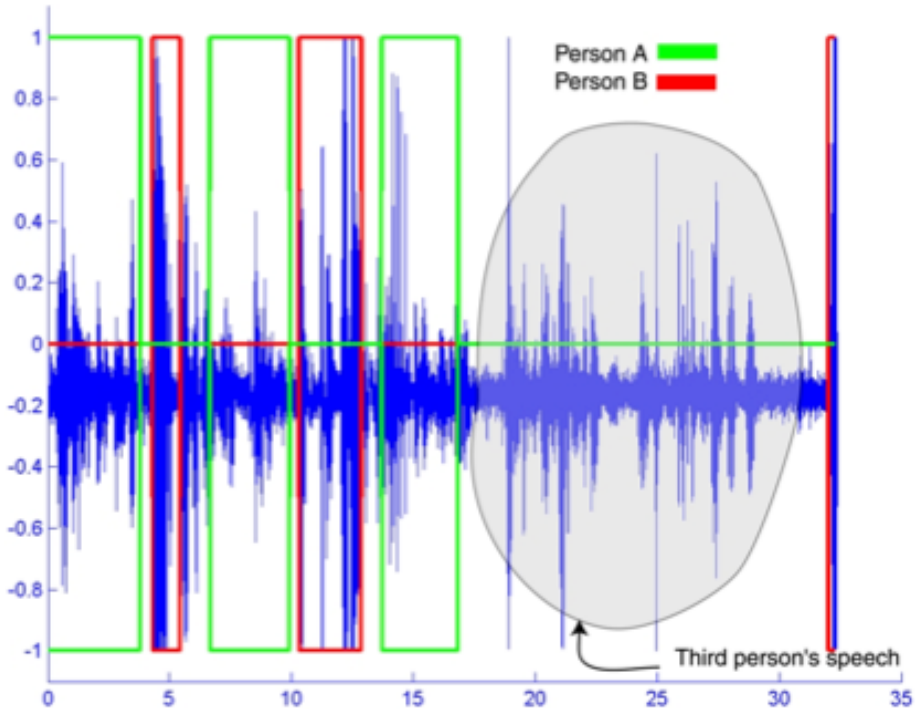
What about GPS?

Ákos et al., *PLOS Comp Bio*, 2013

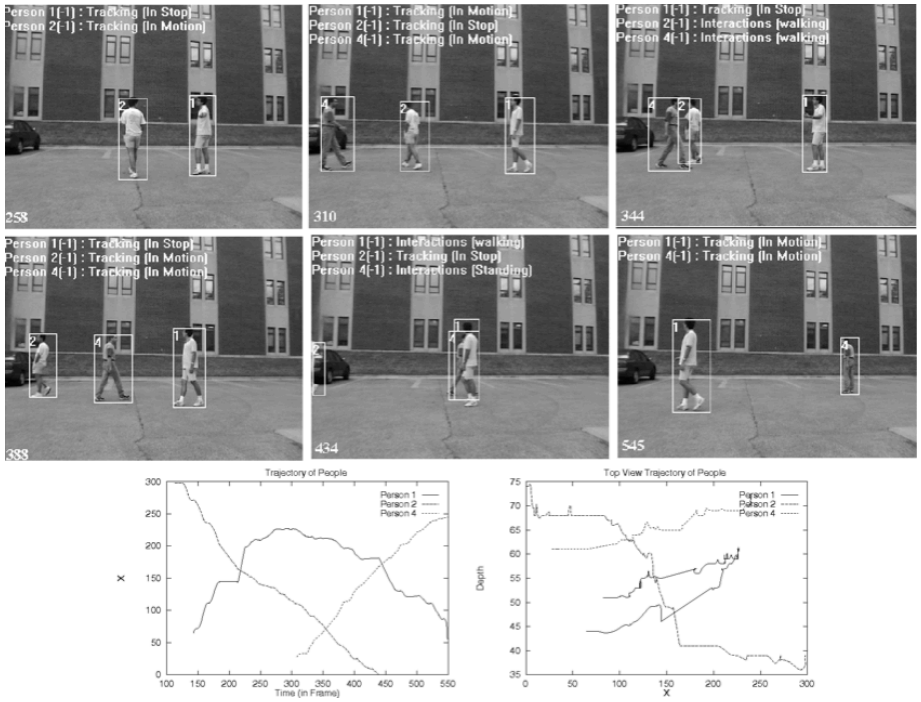
What about Wifi?

Wen et al., "On assessing the accuracy of positioning systems in indoor environment," 2013

Audio and video?



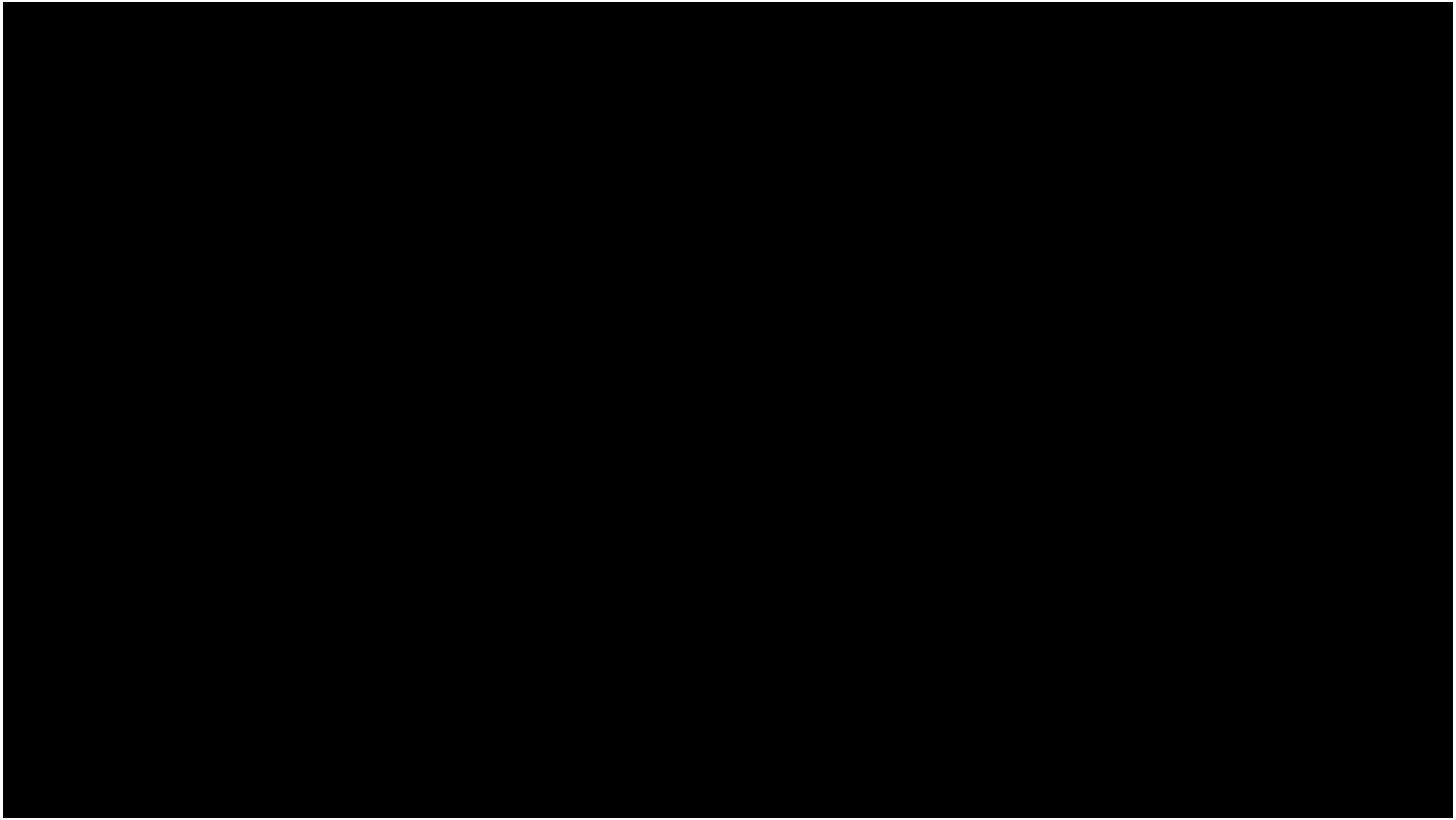
Choudhury & Pentland, "The Sociometer," 2002



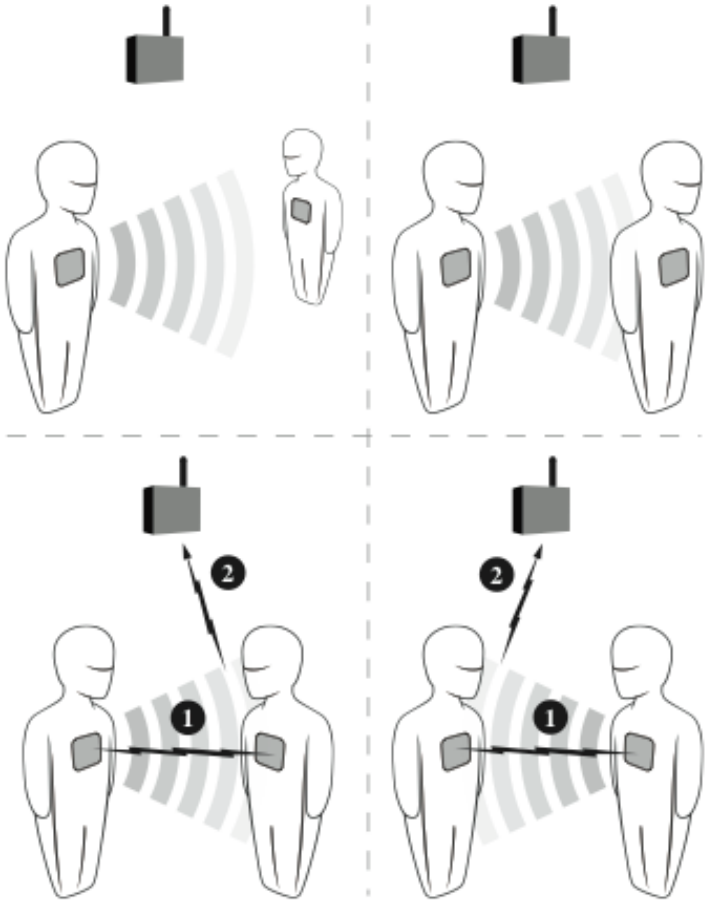
Haritaoglu et al., "Real-time surveillance," 2000

What do we do with sensors?

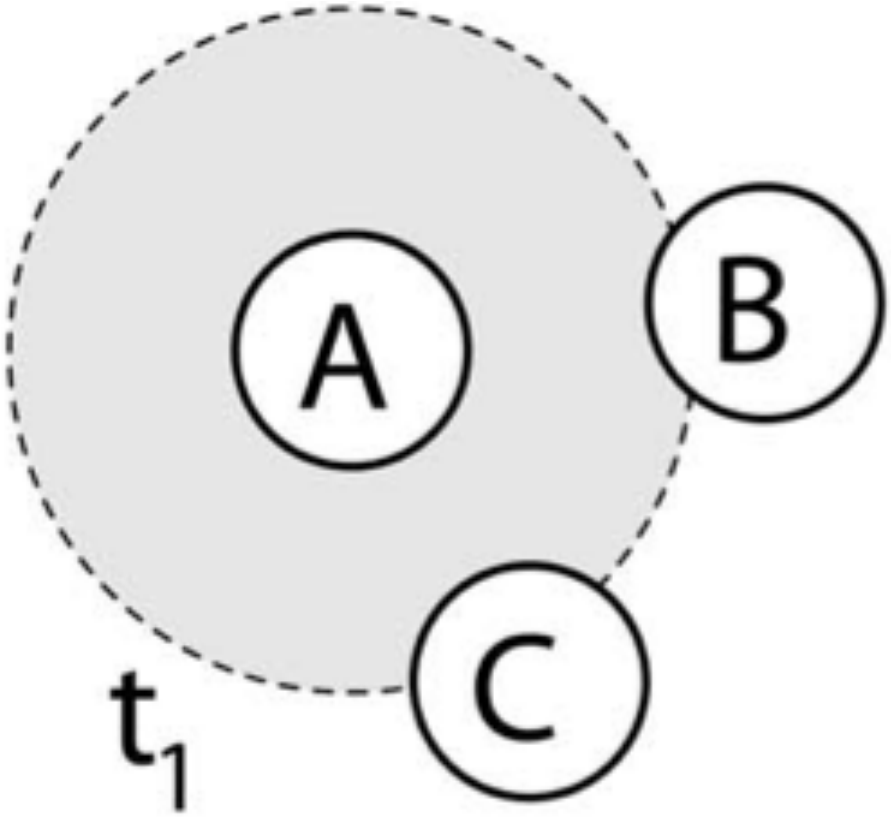
What it looks like in action



Major problem: What is a link?



Panisson et al., "Fingerprinting temporal networks of close-range human proximity," 2013

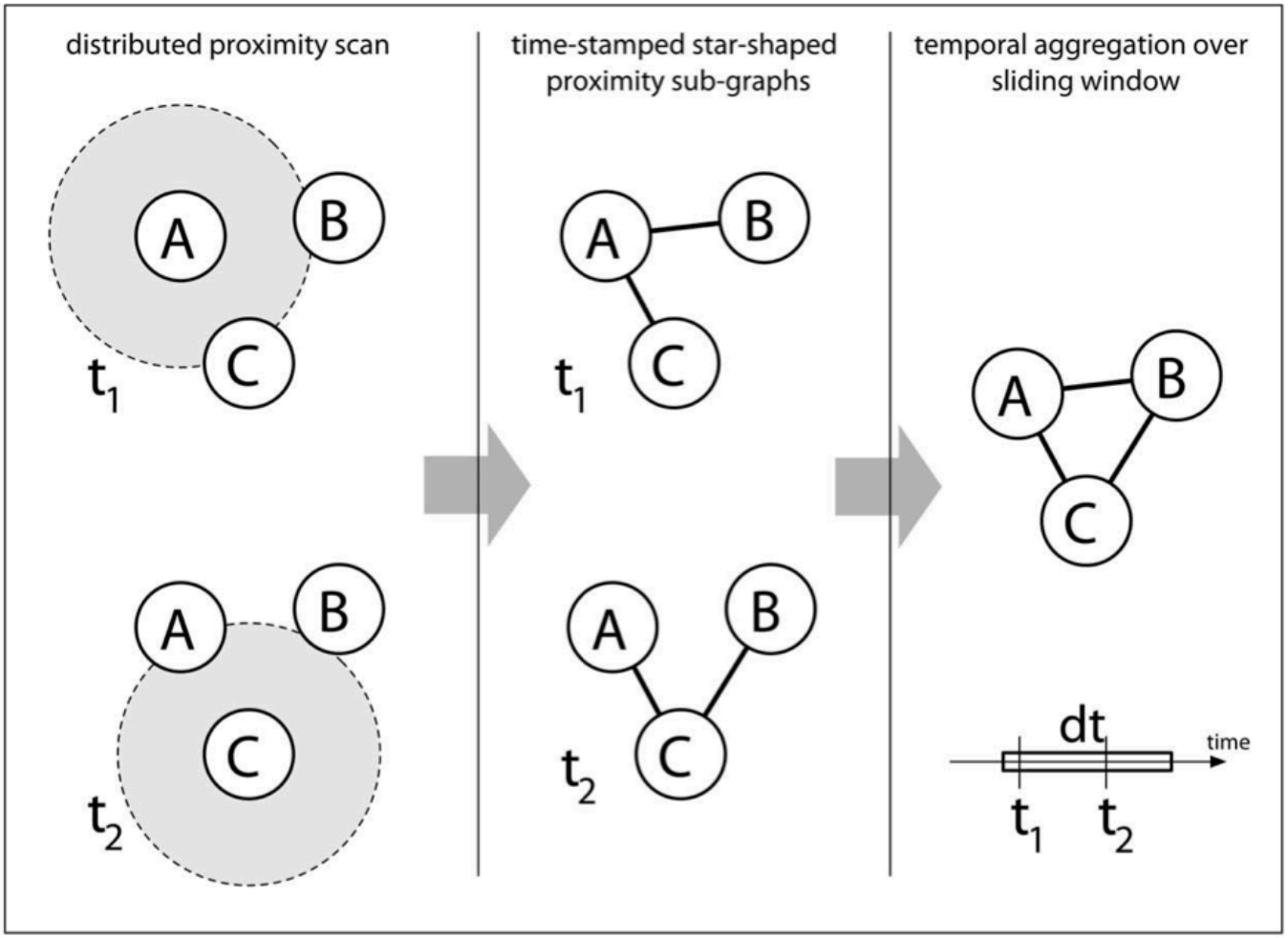


Cattuto et al., "Dynamics of person-to-person interactions from distributed RFID sensor networks," 2010

Major problem: Accuracy

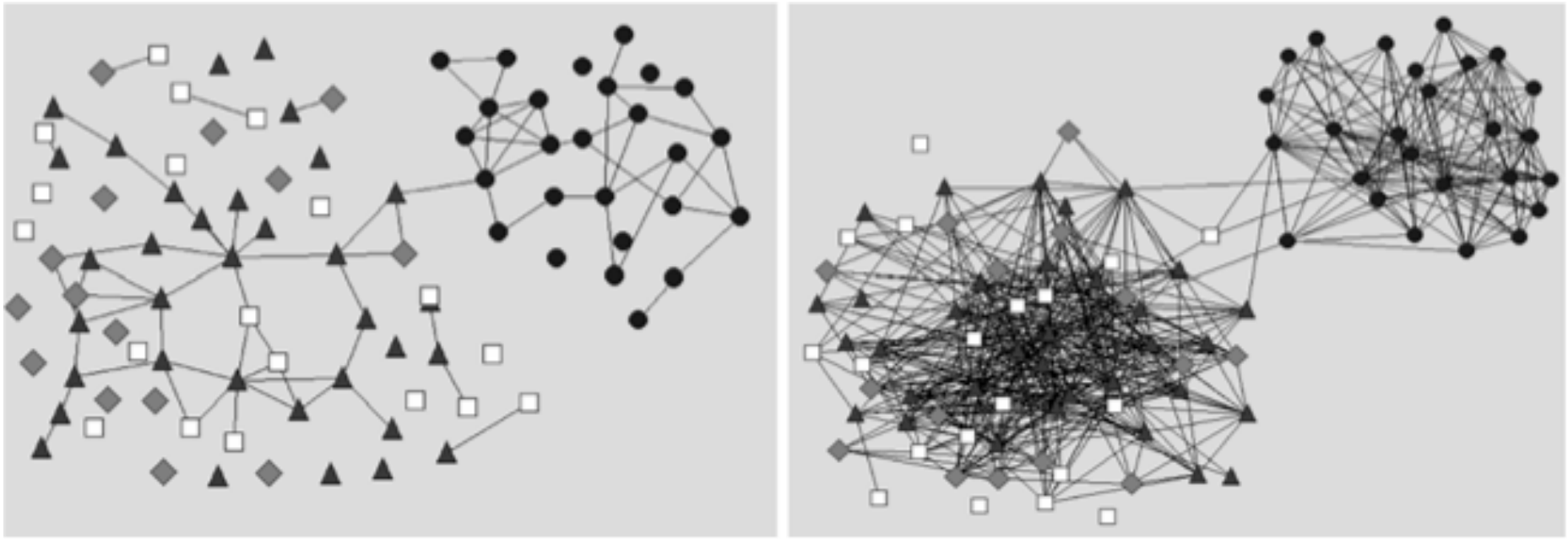
- Infrared only has a 39% base accuracy compared to observation; only when aggregated is it usable at 86% (Choudhury & Pentland, "Sensing and modeling human networks using the sociometer," 2003)
- Bluetooth detection is only 50% in 10min intervals, although asymmetric detection raises that to 75% (Do & Gatica-Perez, "Human interaction discovery in smartphone proximity networks," 2011)
- Signal strength fluctuates wildly within detection range, NOT a reliable measure of distance

Approaches: Aggregation, thresholds, etc.



Cattuto et al., "Dynamics of person-to-person interactions from distributed RFID sensor networks," 2010

Inference



"Friendship (left) and daily proximity (right) networks. Circles represent incoming Sloan business school students. Triangles, diamonds and squares represent senior students, incoming students, and faculty/staff/freshman at the Media Lab. While the two networks share similar structure, inferring friendship from proximity requires the additional information about the context (location and time) of the proximity." (Eagle & Pentland, "Reality Mining," 2006)

Can we really infer social networks?

Data collection methods: Core

- Observation (Gibson, *AJS*, 2005)
  - Most accurate for capturing interaction, but tedious (Gibson: 100 hours of observation time for 2 observers)
- Survey and Interviews
  - Classic problem of construct validity: **informant accuracy** (Bernard, Killworth & Sailer, 1979)
  - Can consider survey data for **cognitive structures** (Krackhardt, 1987; Freeman, Romney & Freeman, 1987)

Data collection methods: Alternative

- Archives and historical materials
  - Limited
- Call logs
  - "several billion ties… dwarves anything that an academic social scientist could garner." (Savage & Burrows, 2007)
  - Single people with multiple cell phones and single cell phones with multiple people are nontrivial (Drew Margolin's talk)
- Social media: go to Jürgen's session later today!! (Sabal, 4:20pm)
  - Many opportunities, but big problems that are under-addressed
- No such thing as a pure 'social network'

Data collection methods

- Sensors
  - Like observation, sensors induce a *realist* boundary specification (Laumann, Marsden & Prensky, 1983)
  - But technical constraints mean we only get directly facing interaction or co-location/proximity
  - So: NOT an alternative to observation for interaction data (yet)
  - We would like to suggest: rather than trying to recover interaction (observational data), affiliation (trace/ organizational data) or communication (archive/trace data), we can think of using [noisy] proximity/co-location

Historical examples

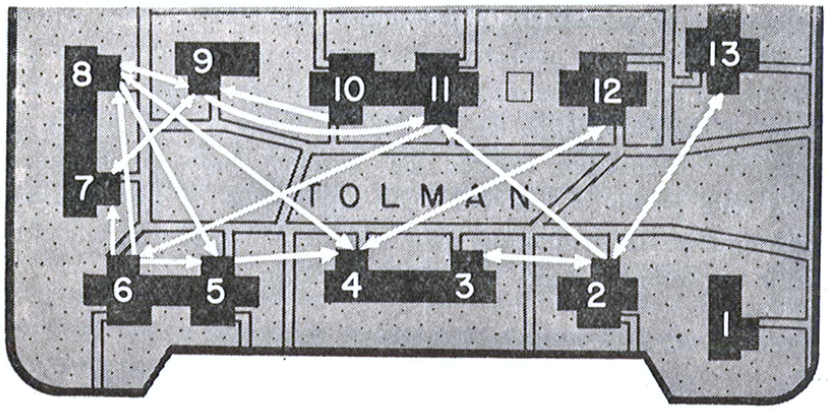


FIG. 9a. Pattern of Sociometric Connections in Tolman Court

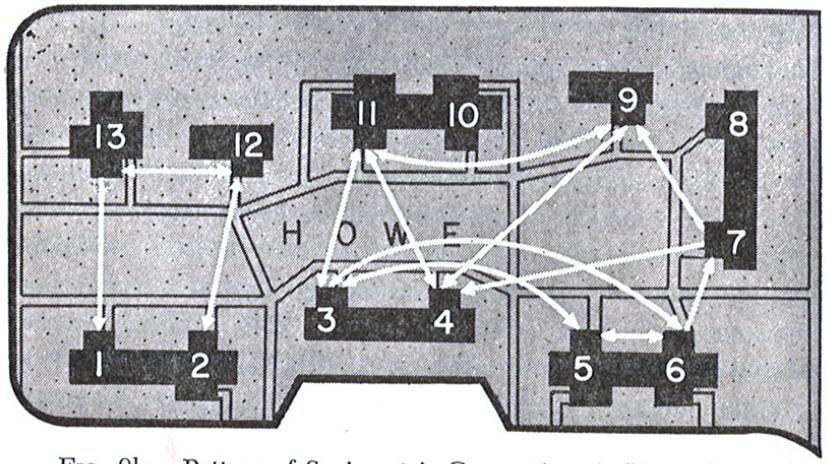


FIG. 9b. Pattern of Sociometric Connections in Howe Court

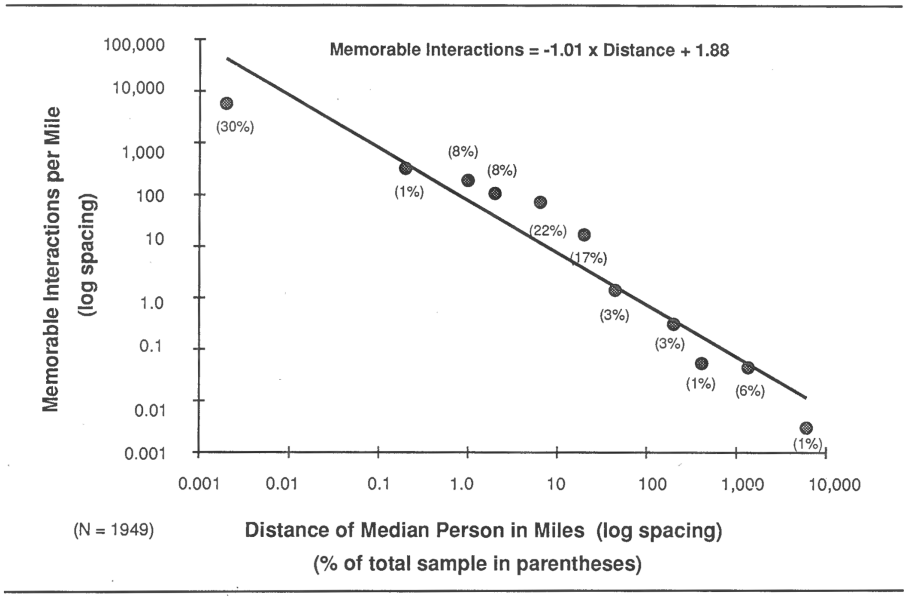
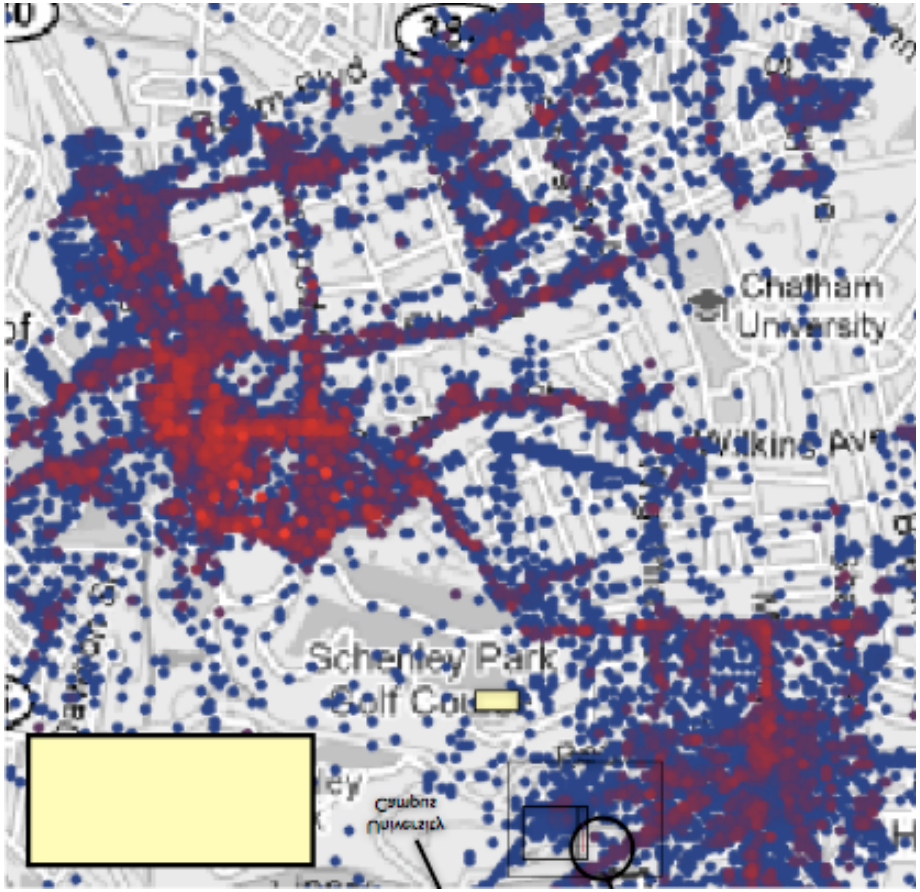


Figure 1 Memorable interactions as a function of physical distance (logarithmic scale): United States.

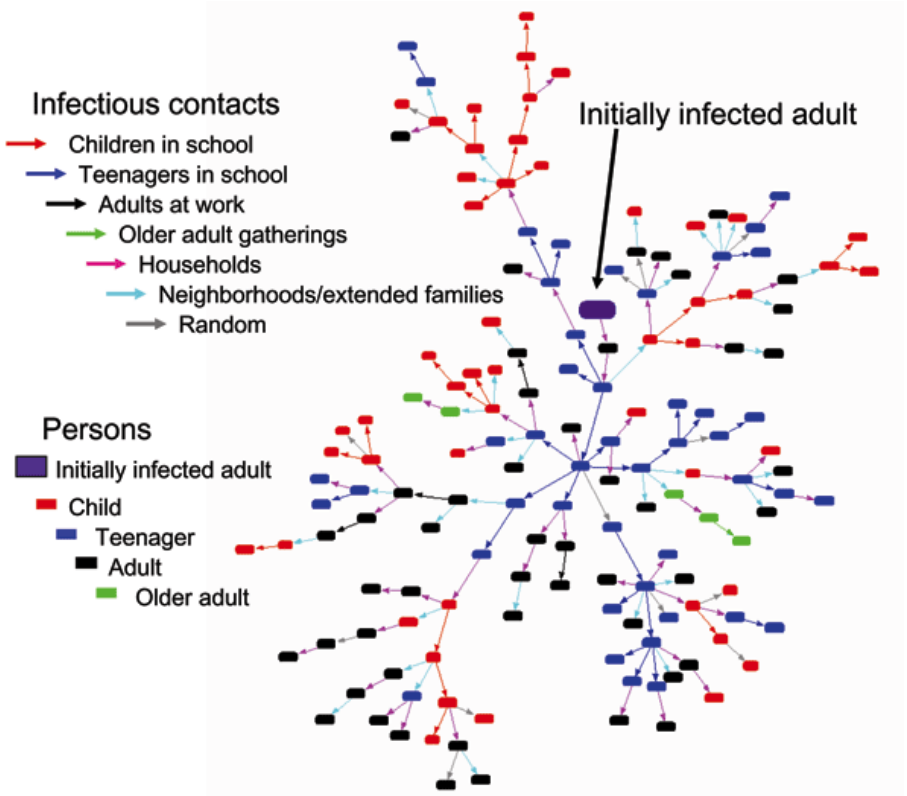
Latané et al., "Distance Matters: Physical Space and Social Impact," 1995

Festinger, Schachter, & Back, *Social Pressures in Informal Groups: A Study of Human Factors in Housing*, 1950

Urban networks? Infectious disease?



Pittsburgh. Cranshaw et al., "Bridging the gap between physical location and online social networks," 2010



Simulated interaction network. Glass et al., "Targeted social distancing designs for pandemic influenza," 2006

Conclusions

- Sensors are a new data collection method, with benefits and drawbacks like every other method
- Sensors are best for proximity or co-location data, rather than interaction or absolute location
  - Sometimes what we want
- Working within their constraints, we can potentially start asking new sorts of questions

Questions?

Extra: SNA and Engineering Collaborations

PSYCHOMETRIKA—VOL. 14, NO. 1
MARCH, 1949

A METHOD OF MATRIX ANALYSIS OF GROUP STRUCTURE

R. DUNCAN LUCE AND ALBERT D. PERRY
GRADUATE STUDENTS, DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Matrix methods may be applied to the analysis of experimental data concerning group structure when these data indicate relationships which can be depicted by line diagrams such as sociograms. One may introduce two concepts, *n*-chain and clique, which have simple relationships to the powers of certain matrices. Using them it is possible to determine the group structure by methods which are

Extra: SNA through radio signals in history

3) Hams. Our third set of data comes from a group of amateur radio operators, commonly called 'hams', living in West Virginia, western Pennsylvania, and eastern Ohio. The hams belong to the Monongalia Wireless Association (MWA), which owns and maintains WR8ABM, a 2-meter, FM repeater station. Virtually all the 2-meter communication of the members of the MWA passes through the repeater station (perched atop a convenient hill) which receives local signals and sends them out over a wide area. WR8ABM is used by hams in a 30,000 mile^2 area.

With the cooperation of the MWA, all conversations on WR8ABM were monitored around the clock for 27 days, using a voice-operated relay between a receiver and a tape recorder. By law and by convention, hams identify themselves with their 'call', the letters and numbers combination issued by the FCC. Thus, all communicants could be monitored, and the length of their conversations (in minutes) could be recorded. For current purposes, only the frequency of communication was used, for similarity with the rest of the data sets.

Bernard, Killworth & Sailer, 1979;
Henry Radio TempoOne (1979), http://soakland5.wix.com/w1al

Sensors within larger data infrastructures

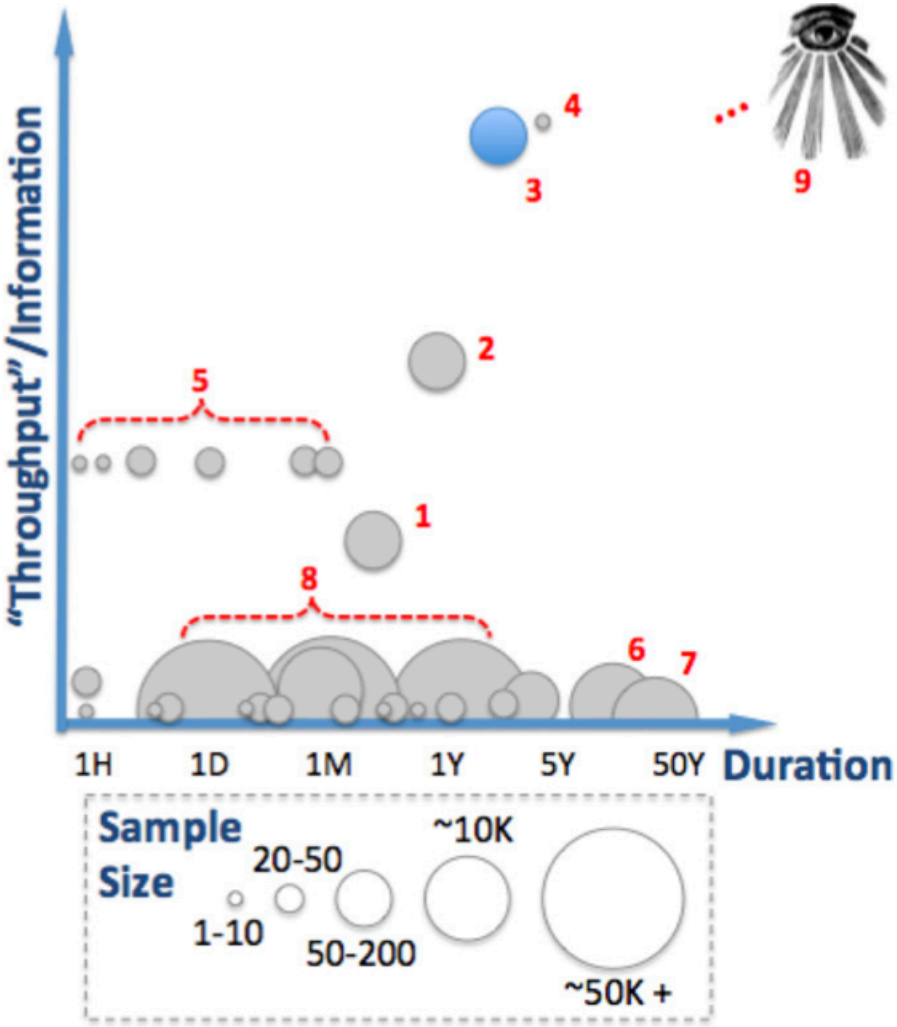


Fig. 1 from Aharony, Pan, Ip, Khayal & Pentland, "Social fMRI," 2011

(1) Reality Mining (Eagle & Pentland, 2006)

(2) Social evolution (Madan, Farrahi, Perez & Pentland, 2011)

(3) Friends and Family dataset (Ahrony et al., 2011)

(4) Rich-data pioneers (Roy et al., "Speechome Project," 2006)

(5) Sociometric Badge studies (Lu et al., *SenSys*, 2010)

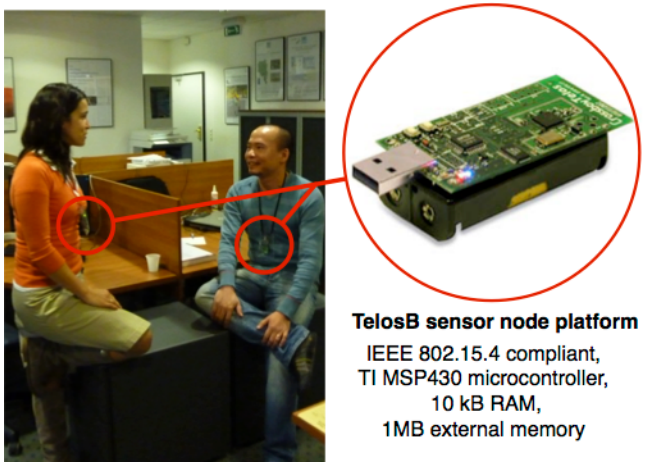
(6) Midwest field station (Barker, 1968)

(7) Framingham Heart Study (Dawber, 1980)

(8) Large call record datasets

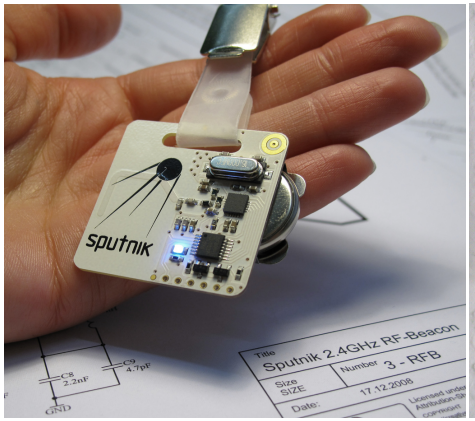
(9) "Omniscient"/all-seeing view

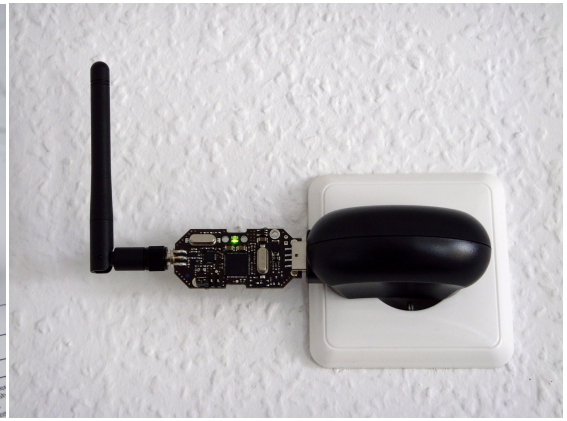
Examples



TelosB node: broadcasts beacons 2-3m every 10 milliseconds.
Förster et al., "On context awareness and social distance in human mobility traces," 2012

Sociometric badge: Bluetooth, infrared, accelerometer, microphone.
Fischbach et al., "Analyzing the flow of knowledge with sociometric badges," 2009



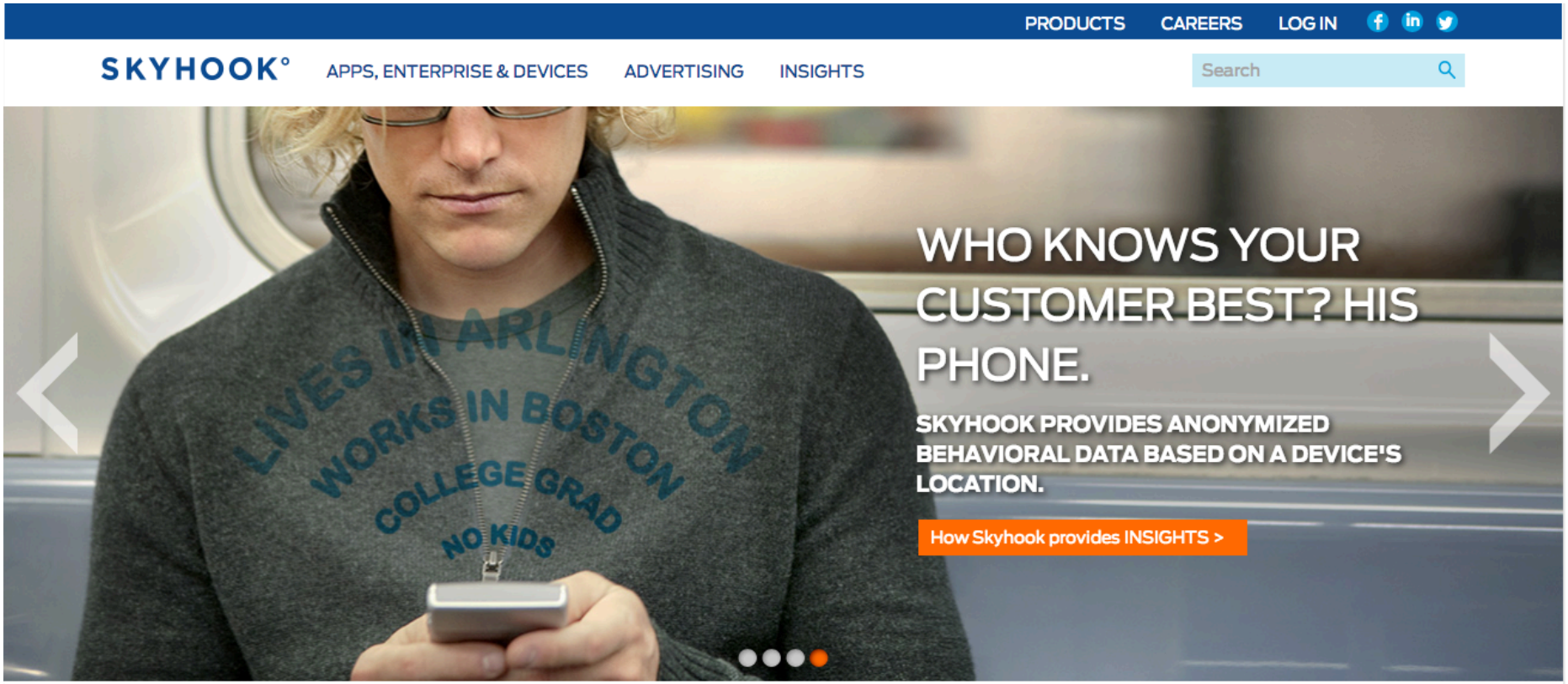


RFID tag, base station, and base station deployment. Configurable power: 1-2m or 4-5m detection.
Panisson et al., "Fingerprinting temporal networks of close-range human proximity," 2013
Barrat et al., "High resolution dynamical mappings of social interactions with active RFID," 2008

Geolocation often from cell towers

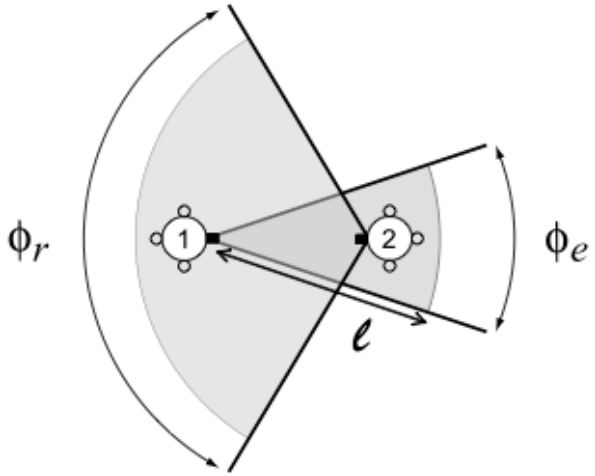
Malte Spitz, two days in August 2010

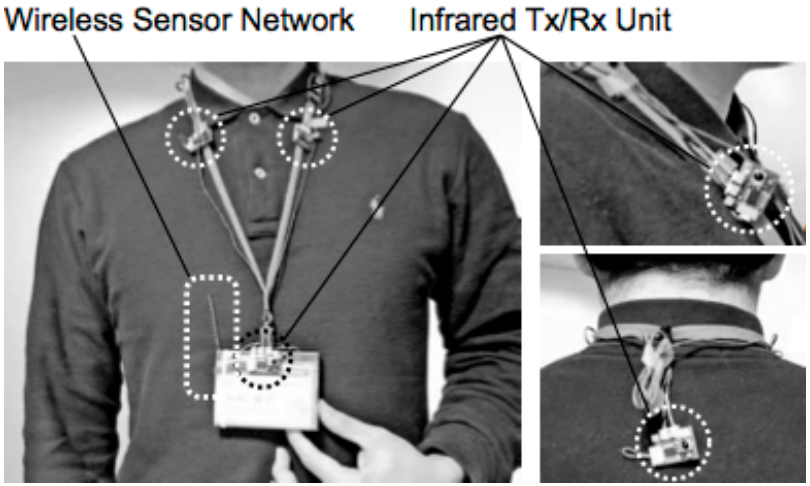
Wifi: Commercial aggregation services

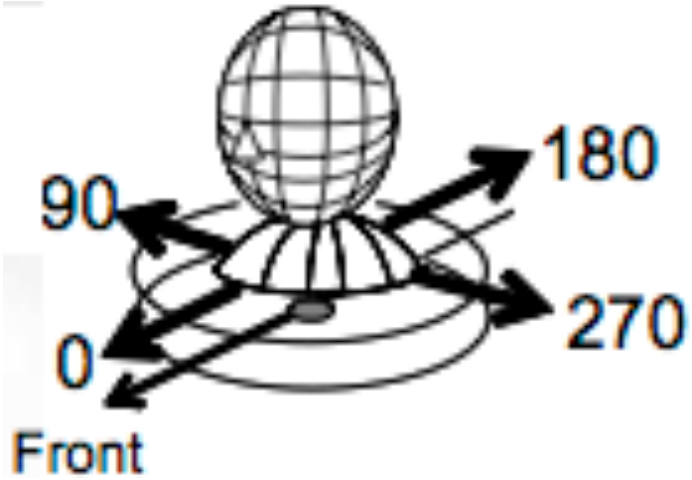


Major problem: Too narrow or too broad?

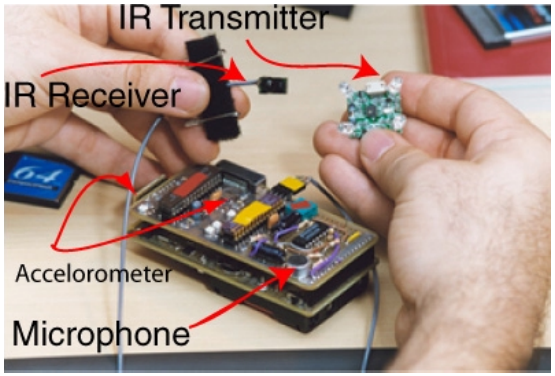
- Example solution, from Nakamura et al., 2007, A method for estimating position and orientation with a topological approach using multiple infrared tags

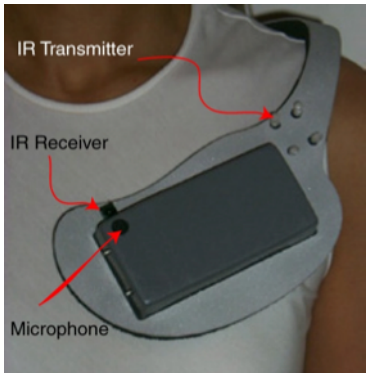


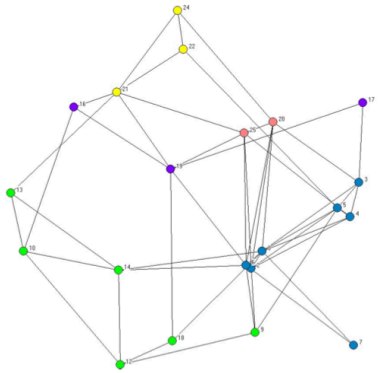




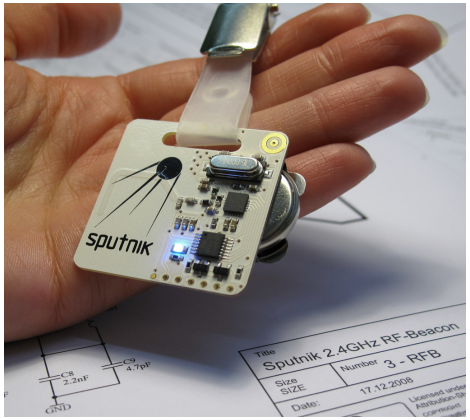
Past 10 years: Rise of social network 'sensing'

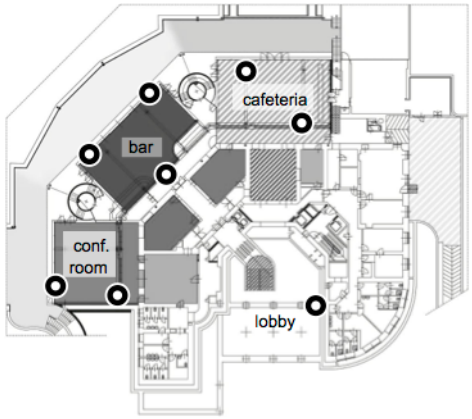


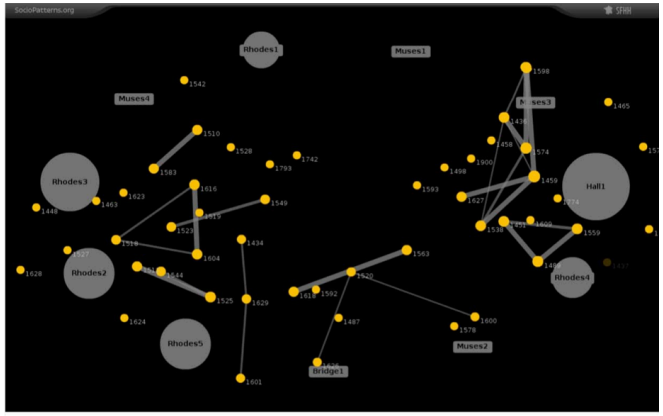




Choudhury & Pentland, (2003) Modeling face-to-face communication using the sociometer.







Cattuto et al., Dynamics of person-to-person interactions from distributed RFID sensor networks, 2010
Barrat et al., High resolution dynamical mappings of social interactions with active RFID, 2008