Carnegie Mellon University
School of Computer Science

Friendship and proximity in a fraternity cohort with mobile phone sensors

http://mominmalik.com/sunbelt2018.pdf

Momin M. Malik

Sunbelt XXXVIII, 01 July 2018

Jürgen Pfeffer (Technical University of Munich)
Afsaneh Doryab (Carnegie Mellon University)
Michael Merrill (HealthRhythms)
Anind K. Dey (University of Washington)

Sunbelt
2018 Utrecht

Key points

Theory:

- RFID and Bluetooth sensors *measure* proximity, which can be a proxy for the *construct* of interaction
- But proximity is also important as a construct

Practice:

- Compare sensors to other data (e.g., survey data)
- Reduce sensor data by “feature extraction” and variable selection, done with careful cross-validation

Sensors + social network studies

Study	Sensor	Collection
Sociometric badge	Infrared	2002, 2007
Reality Mining	Bluetooth	2004
Social Evolution	Bluetooth	2008-2009
SocioPatterns	RFID	2008-2018
Lausanne	Bluetooth	2009-2010
SocialfMRI	Bluetooth	2010-2011
Copenhagen Networks Study	Bluetooth	2012-2013

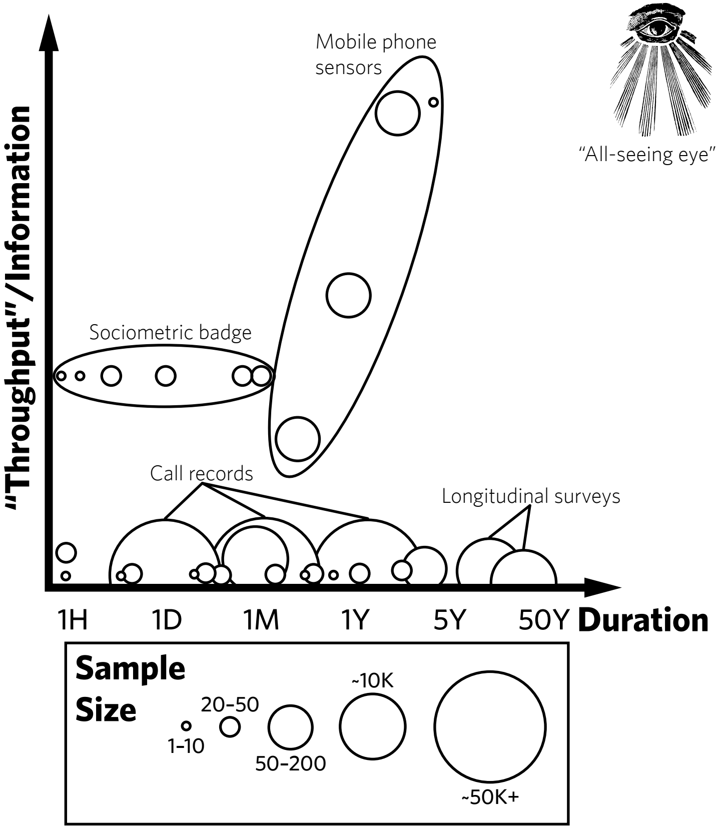
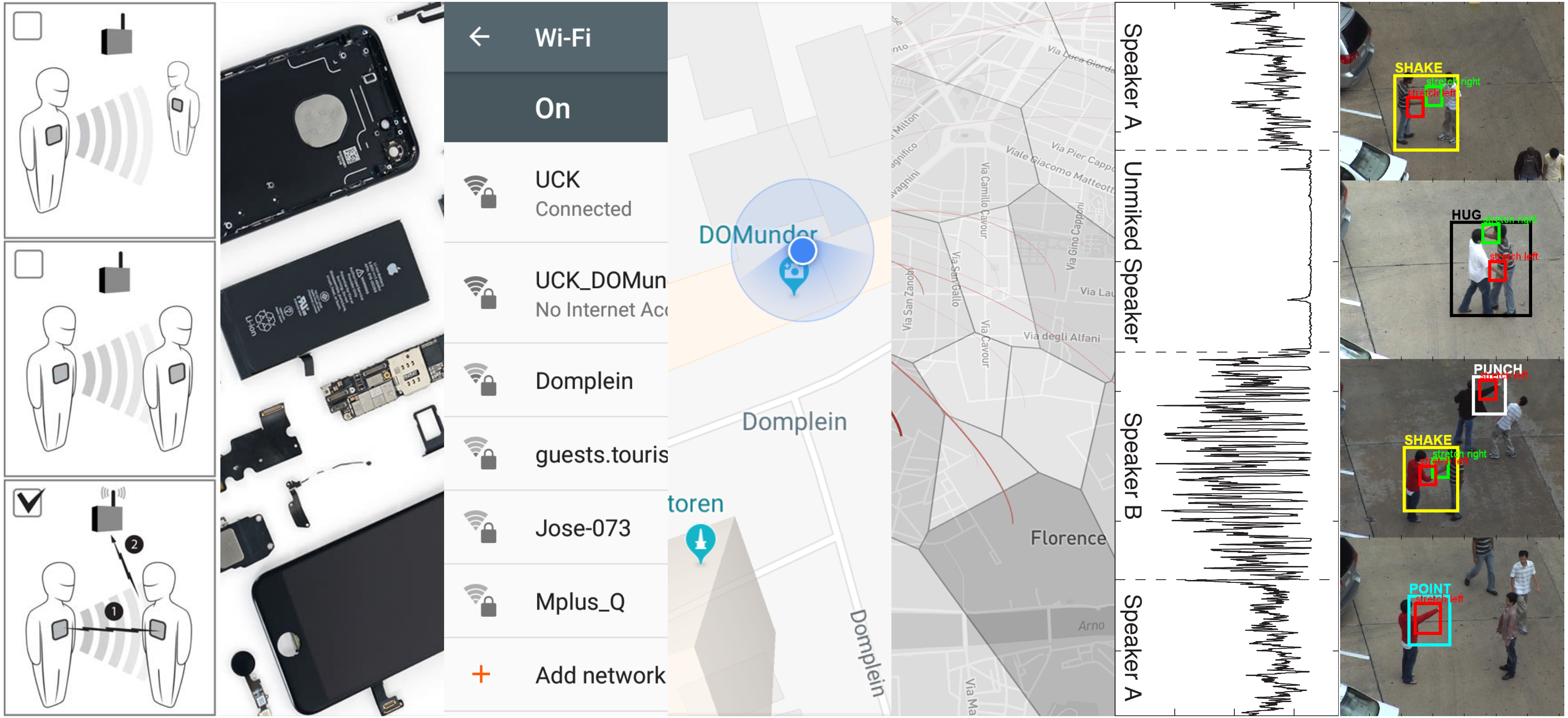


Diagram reproduced from Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland (2011). "Social fMRI: Investigating and shaping social mechanisms in the real world". *Pervasive and Mobile Computing* 7(6), 643–659. doi: [10.1016/j.pmcj.2011.09.004](https://doi.org/10.1016/j.pmcj.2011.09.004).

Relational sensor data

RFID[1] | Bluetooth | WiFi | GPS | Cell towers | Audio[2] | Video[3]



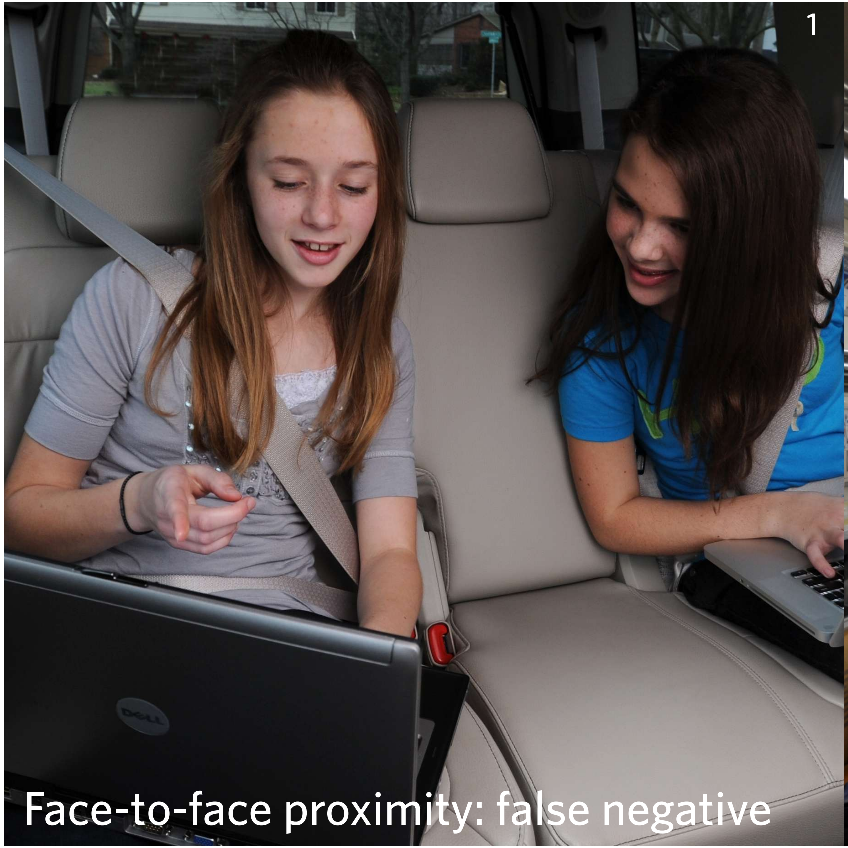
Inconsistent terminology suggests confusion

- Copenhagen Networks Study (Bluetooth):
  - "Proximity data"[1]
  - "Face-to-face interactions"[2]
  - "Close proximity interactions"[3]
  - "Face-to-face contacts"[4]
  - "Physical contacts"[5]
- SocioPatterns papers (RFID):
  - "Person-to-person interaction"[6]
  - "Face-to-face contacts"[7]
  - "Close-range interactions"[8]
  - "Face-to-face interactions"[9]
  - "Face-to-face proximity"[10]
- Audio:
  - "Face-to-face conversation"[11]

Back to basics: *Constructs.*

- *Constructs*: primitives of social science
  - A measurement might be a *proxy* for an non-observable construct (e.g., multiple choice questions and intelligence)
  - Proxies always give errors (binary construct: false negatives and false positives)
  - (Criterion-related ["predictive"] validity)
- Face-to-face interaction: neither the measure nor the construct

In-person interaction is the true construct


1
Face-to-face proximity: false negative

2
Face-to-face proximity: false positive

(*Conversation* is a separate construct)

Constructs have their own importance

- What construct do we care about?
- Depends on what we want to study/investigate.
  - Disease transmission? Directional proximity and/or physical contact.
  - Persuasion? Conversation.
  - Mimicry? Interaction.
  - Latent homophily, expressed geographically? Proximity.
  - Environmental exposure? Proximity.

Survey data, too, has its own importance

- "Objective" sensor data is not superior to survey data
  - Yes, informant inaccuracy, social desirability bias, ambiguous questions...
- But they are measuring *different things*
  - Surveys better measure the *psychological perceptions* that may ultimately be causal for behavior[1] (e.g., memorability[2])
- So, discrepancies must not be resolved in favor of the "objective" data
- Discrepancies are exactly the interesting thing to study!!
- Propinquity is an example (discrepancy is "close strangers, distant friends"[3])

Proximity is itself interesting (propinquity!)

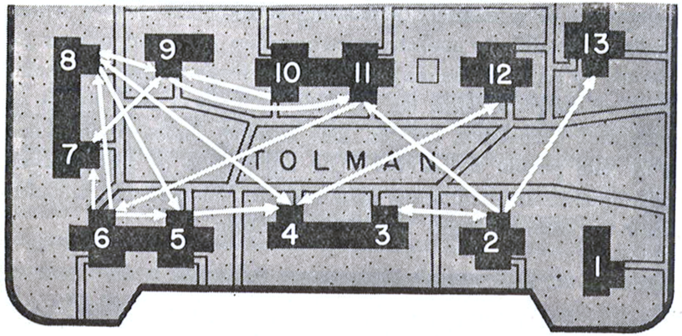


FIG. 9a. Pattern of Sociometric Connections in Tolman Court

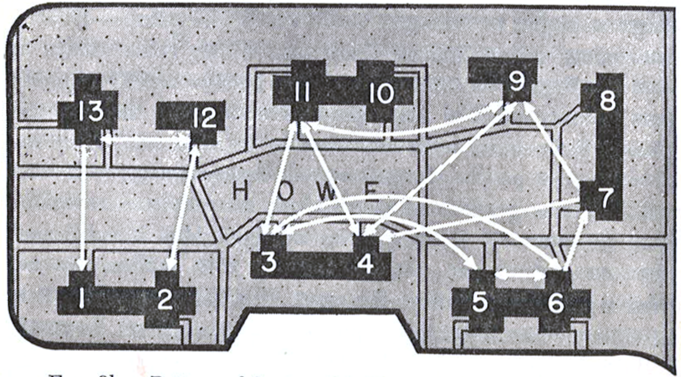
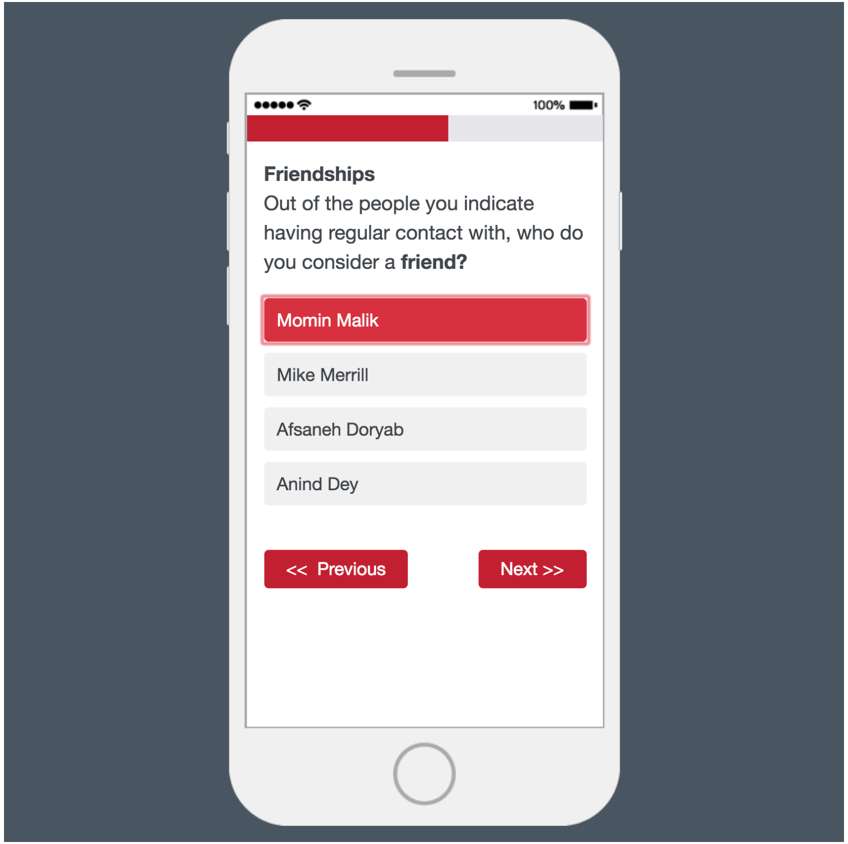


FIG. 9b. Pattern of Sociometric Connections in Howe Court

Study

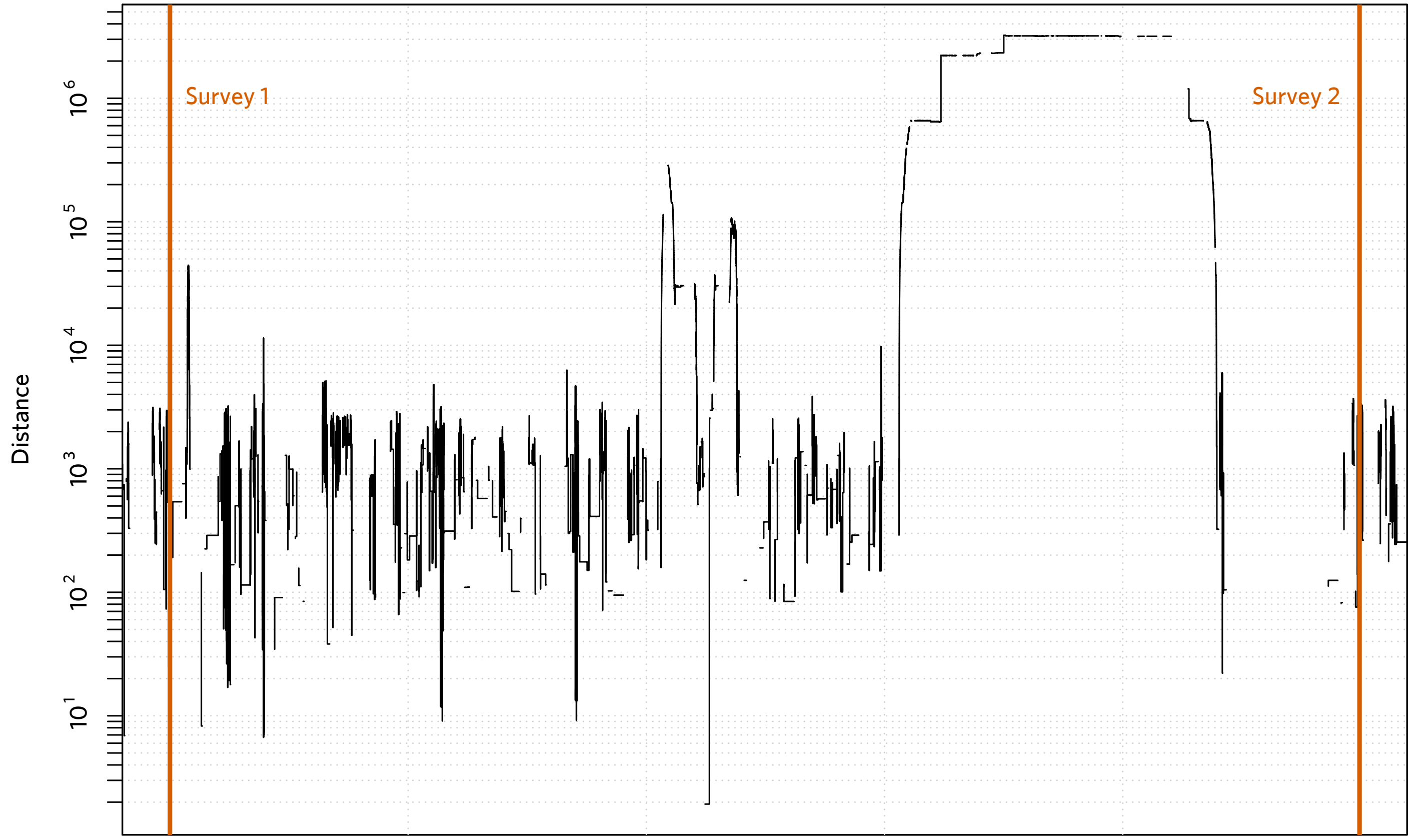
Data: Surveys + mobile phone tracking



Goal: Study propinquity

- Not proximity as proxy for interaction, but proximity itself
- Compare proximity (via “location”, WiFi) to longitudinal sociometric choice
- Look at proximity at scales larger than that of interaction
  – Small scales (proximity at <10m): underlying causal mechanism might still be interaction.
  – Large scales (proximity >20m): will capture other mechanisms, e.g. latent homophily, common environmental exposure, etc.

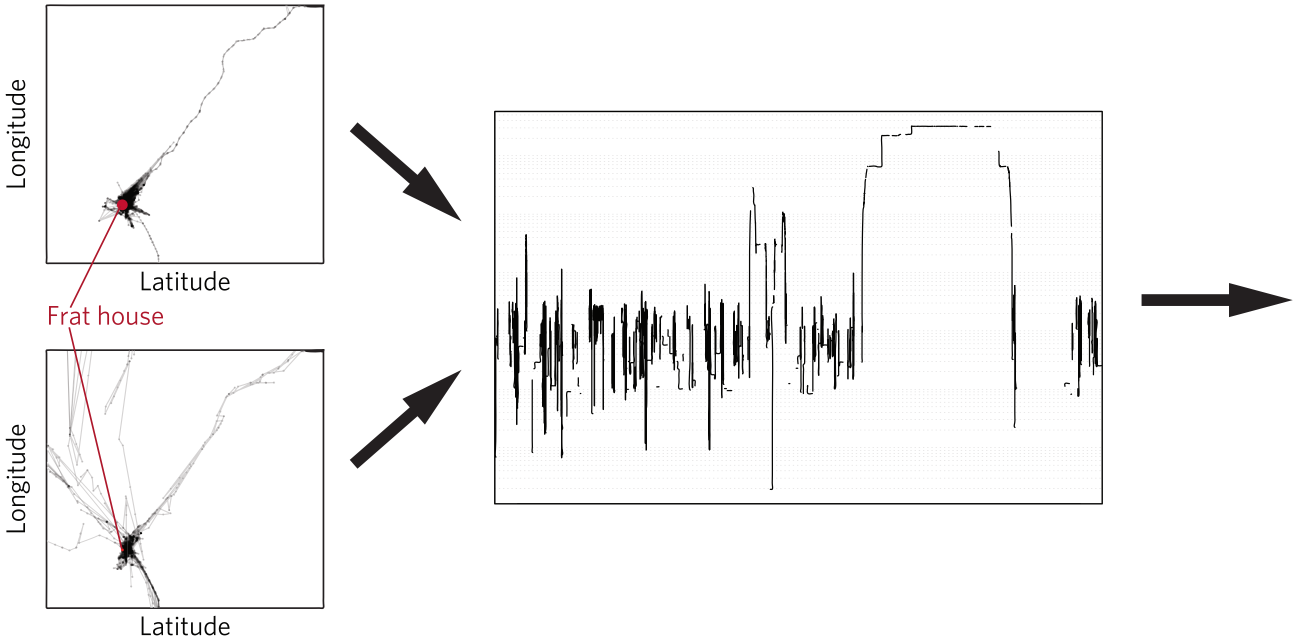
Core problem: *Different resolutions*



Approach: First do machine learning

- Step 1: Find out how to meaningfully characterize the association of proximity and friendship
- Step 2: Using this characterization, model co-evolution

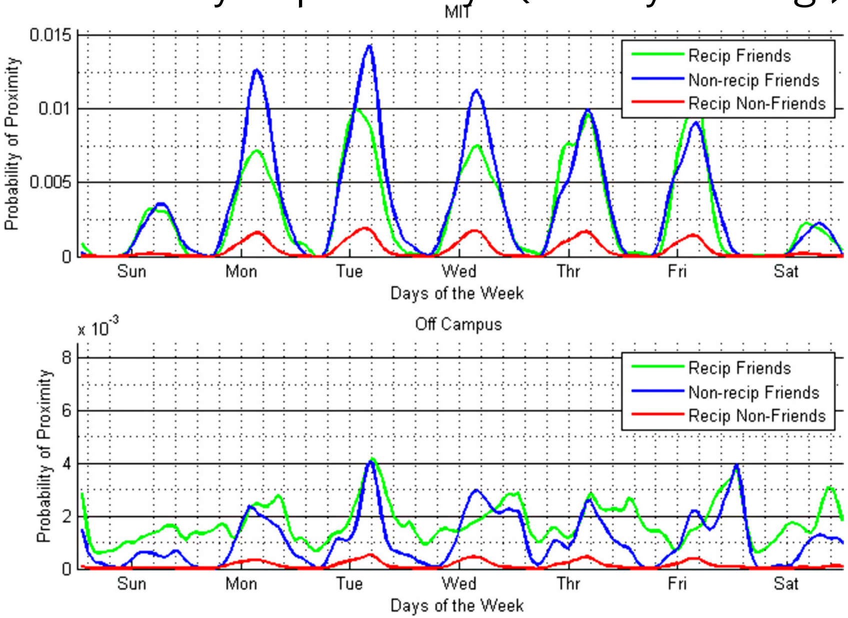
Data processing and “feature extraction”



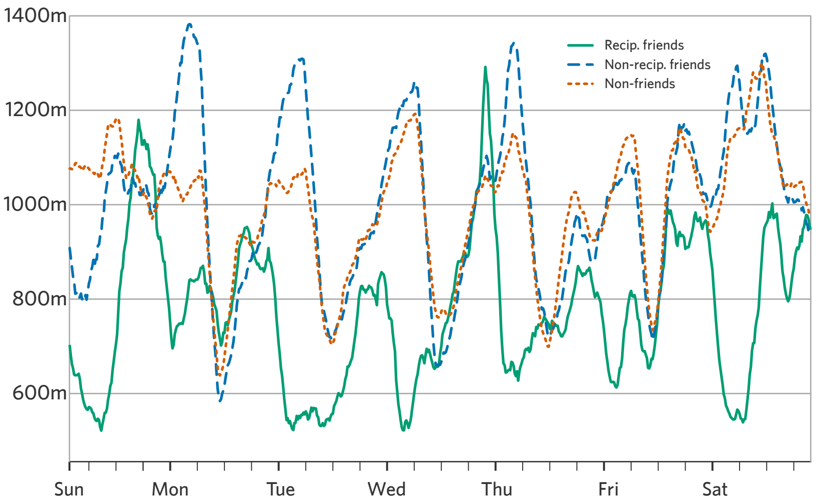
0.086	0.281	0.0793	0.079	
0.005	0.073	0.0054	0.005	
0.057	0.234	0.0547	0.054	…
0.007	0.086	0.0074	0.007	
0.071	0.258	0.0669	0.066	
0.024	0.154	0.0238	0.023	
	⋮			

Caution: Aggregates can mislead. Better test of an association is its predictive performance

“Probability of proximity” (Reality Mining[1])



Median pairwise distance (our study)

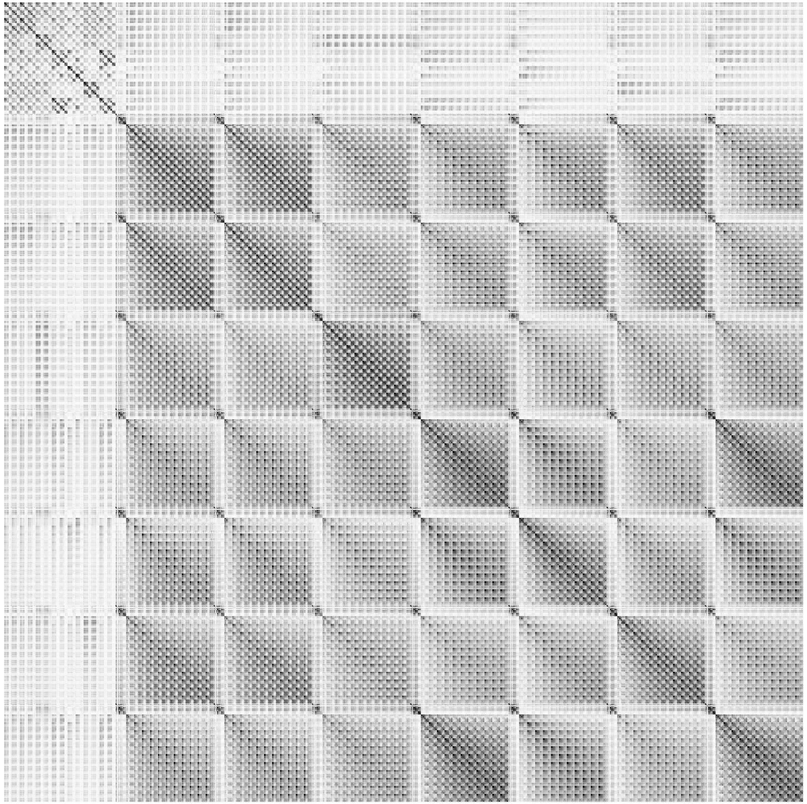


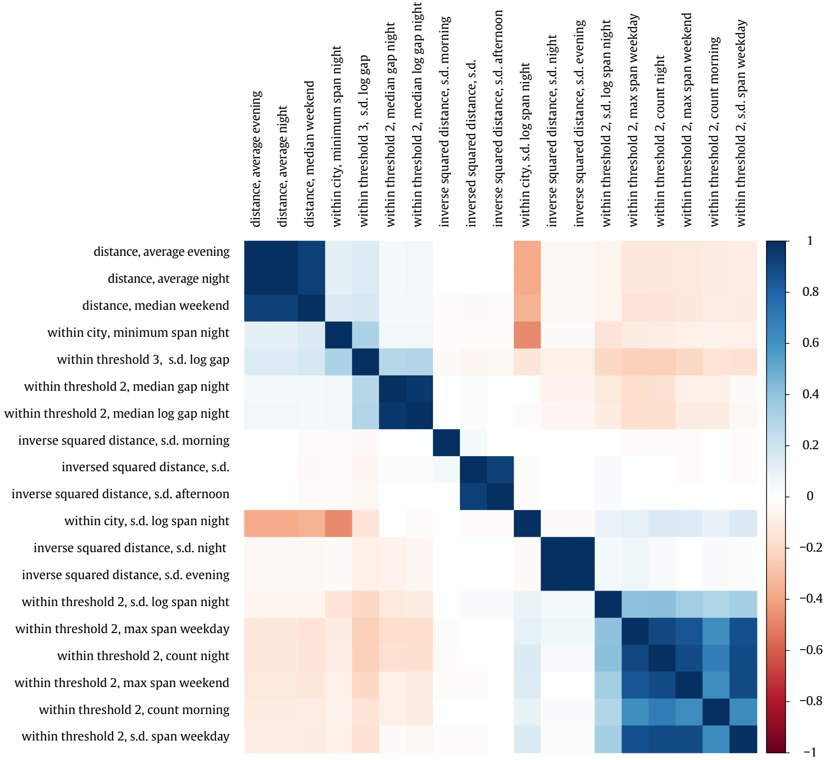
We found what looked like a compelling pattern as well, but it proved ineffective for prediction when tested with cross-validation. Why? Aggregate trends obscure between-dyad and week-to-week variance.

Test the performance via *cross-validation*

- Split data into “training” and “test”
- Fit model on training, evaluate on test
- Done correctly, simulates out-of-sample data, thereby directly establishing external validity
- But dependencies (e.g. time, networks) can complicate cross-validation
- We use multiple cross-validation schema to control for this (details in forthcoming work)

Result: ~30% association. Can get with 2.5K features… or 19, after feature selection.





Summary: How we *should* use sensors

- If using Bluetooth, RFID proxies for interaction, do more testing against human-coded benchmarks
- But *proximity* is also inherently interesting
- Compare proximity other forms of data (e.g., friendship for propinquity/influence vs. exposure)
- Comparing sensor data and survey data, e.g. via SAOMs, is a good framework
- Reduce/summarize rich signals through feature extraction + selection, not naïve aggregation

Carnegie Mellon University
School of Computer Science

Thank you!

Theory:

- RFID and Bluetooth sensors *measure* proximity, which can be a proxy for the *construct* of interaction
- But proximity is also important as a construct

Practice:

- Compare sensors to other data (e.g., survey data)
- Reduce sensor data by "feature extraction" and variable selection, done with careful cross-validation

Contact: Momin Malik <momin.malik@gmail.com>
Work with Jürgen Pfeffer, Afsaneh Doryab. Michael Merrill, and Anind K. Dey

Thanks also to Yuvraj Agarwal and Nynke Niezink.

Endnotes/references (1 of 2)

Slide 4

1. Ciro Cattuto, Wouter van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani (2010). "Dynamics of person-to-person interactions from distributed RFID sensor networks". *PLOS ONE* 5(7), e11596. doi: 10.1371/journal.pone.0011596.
2. Danny Wyatt, Tanzeem Choudhury, Jeff Bilmes, and James A. Kitts (2011). "Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science". *ACM Transactions on Intelligent System Technologies* 2(1), 7:1–7:41. doi: 10.1145/1889681.1889688.
3. M. S. Ryoo and J. K. Aggarwal (2009). "Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities". *Proceedings of the IEEE 12th International Conference on Computer Vision* (ICCV), 1593–1600.

Slide 5

1. Vedran Sekara and Sune Lehmann (2014). "The strength of friendship ties in proximity sensor data". *PLOS ONE* 9(7), 1–8. doi: 10.1371/journal.pone.0100915.
2. Arkadiusz Stopczynski, Vedran Sekara, Piotr Sapiezynski, Andrea Cuttone, Mette My Madsen, Jakob Eg Larsen, and Sune Lehmann (2014). "Measuring large-scale social networks with high resolution". *PLOS ONE* 9(4), 1–24. doi: 10.1371/journal.pone.0095978.
3. Stopczynski, Arkadiusz, Piotr Sapiezynski, Alex Pentland, and Sune Lehmann (2015). "Temporal fidelity in dynamic social networks". *The European Physical Journal B* 88(249). doi: 10.1140/epjb/e2015- 60549-7.
4. Anders Mollgaard, Ingo Zettler, Jesper Dammeyer, Mogens H. Jensen, Sune Lehmann, and Joachim Mathiesen (2016). "Measure of node similarity in multilayer networks". *PLOS ONE* 11(6), 1–10. doi: 10.1371/journal.pone.0157436.
5. Enys Mones, Arkadiusz Stopczynski, and Sune Lehmann (2017). "Contact activity and dynamics of the social core". *EPJ Data Science* 6(1). doi: 10.1140/epjds/s13688-017-0103-y.
6. Ciro Cattuto, Wouter van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani (2010). "Dynamics of person-to-person interactions from distributed RFID sensor networks". *PLOS ONE* 5(7), e11596. doi: 10.1371/journal.pone.0011596.
7. Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Caterina Rizzo, Alberto Eugenio Tozzi, and Wouter van den Broeck (2012). "Wearable sensor networks for measuring face-to-face contact patterns in healthcare settings". *Revised Selected Papers from the Third International Conference on Electronic Healthcare* (eHealth 2010), p192–195. doi: 10.1007/978-3-642-23635-8_24.

Endnotes/references (2 of 2)

8. Ciro Cattuto, Marco Quaggiotto, André Panisson, and Alex Averbuch (2013). "Time-varying social networks in a graph database: A Neo4J use case". *Proceedings of the First International Workshop on Graph Data Management Experiences and Systems* (GRADES '13), 11:1–11:6. doi: 10.1145/2484425.2484442.
9. Alain Barrat, Ciro Cattuto, Vittoria Colizza, Francesco Gesualdo, Lorenzo Isella, Elisabetta Pandolfi, Jean- François Pinton, Lucilla Ravà, Caterina Rizzo, Mariateresa Romano, Juliette Stehlé, Alberto Eugenio Tozzi, and Wouter van den Broeck (2013). "Empirical temporal networks of face-to-face human interactions". *The European Physical Journal Special Topics* 222(6), 1295–1309. doi: 10.1140/epjst/ e2013-01927-7.
10. Alain Barrat, Ciro Cattuto, Alberto Eugenio Tozzi, Philippe Vanhems, and Nicolas Voirin (2014). "Measuring contact patterns with wearable sensors: Methods, data characteristics and applications to data- driven simulations of infectious diseases". *Ilinical Microbiology and Infection* 20(1), 10–16. doi: 10.1111/1469-0691.12472.
11. Wyatt et al., op. cit.

Slide 7

1. Ford Motor Company (2009, December 22). "In-car connection." https://flic.kr/p/7pHMeE
2. Doug Ball (flickr user Caribb) (2009, August 29). "Osaka subway car." https://flic.kr/p/76wuD6

Slide 10

1. David Krackhardt (1987). "Cognitive social structures". *Social Networks* 9(2), 109–134. doi: 10. 1016/0378-8733(87)90009-8.
2. Bibb Latané, James H. Liu, Andrzej Nowak, Michael Bonevento, and Long Zheng (1995). "Distance matters: Physical space and social impact". *Personality and Social Psychology Bulletin* 21(8), 795–805. doi: 10.1177/0146167295218002.
3. jimi adams [sic] (2010). "Distant friends, close strangers? Inferring friendships from behavior". *Proceedings of the National Academy of Sciences* 107(9), E29–E30. doi: 10.1073/pnas.0911195107.

Slide 11

1. Leon Festinger, Kurt W. Back, and Stanley Schachter (1950). *Social pressure in informal groups: A study of human factors in housing*. Stanford University Press.

Slide 18

1. Nathan Eagle, Alex Pentland, and David Lazer (2009). "Inferring friendship network structure by using mobile phone data". *Proceedings of the National Academy of Sciences* 106(36), 15274–15278. doi: 10.1073/pnas.0900282106.