BERKMAN KLEIN CENTER FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

Theorizing sensors for social network research

- *Momin M. Malik, PhD <momin_malik@cyber.harvard.edu>*

Data Science Postdoctoral Fellow
Berkman Klein Center for Internet & Society at Harvard University

Computational Social Science Institute, UMass Amherst
7 December 2018

Slides: https://mominmalik.com/cssi.pdf

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

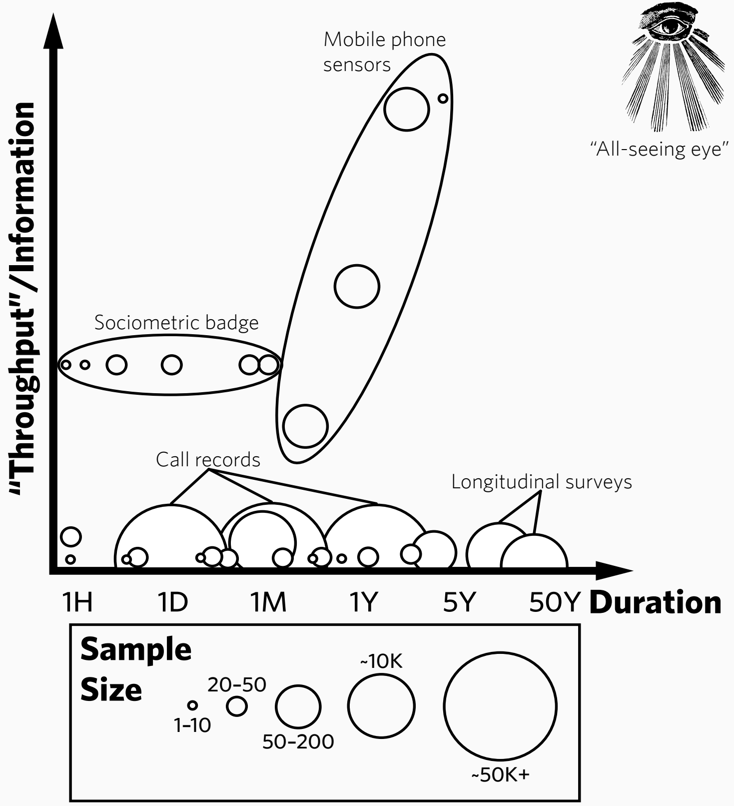
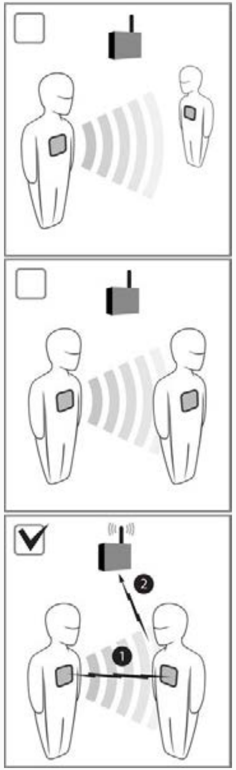


Diagram reproduced from Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland (2011). "Social fMRI: Investigating and shaping social mechanisms in the real world." *Pervasive and Mobile Computing* 7 (6), 643–659. https://doi.org/10.1016/j.pmcj.2011.09.004.

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

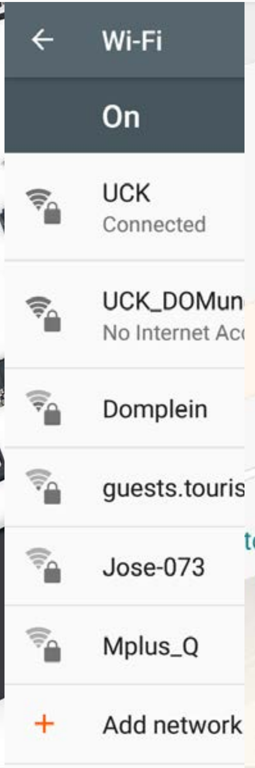
Relational sensor data

RFID[1]



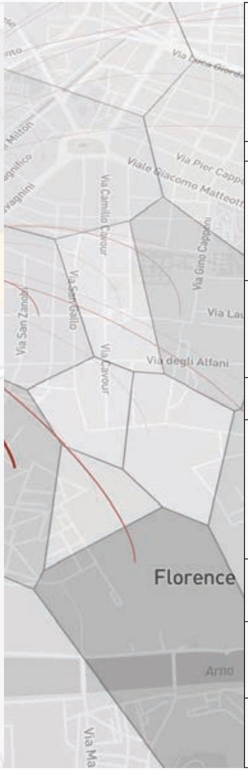
Bluetooth

WiFi

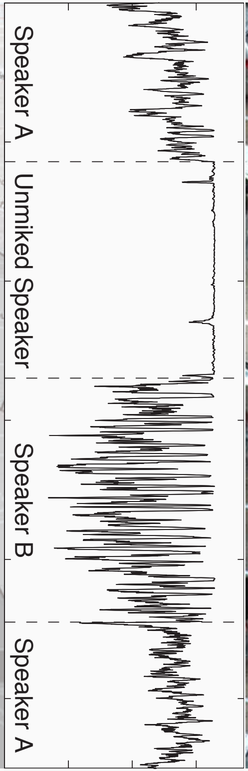


GPS

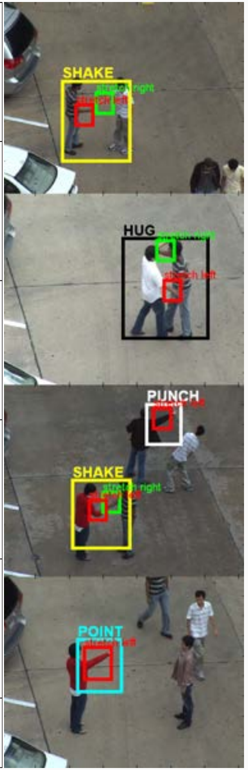
Cell towers



Audio[2]



Video[3]



Inconsistent terminology, confusion

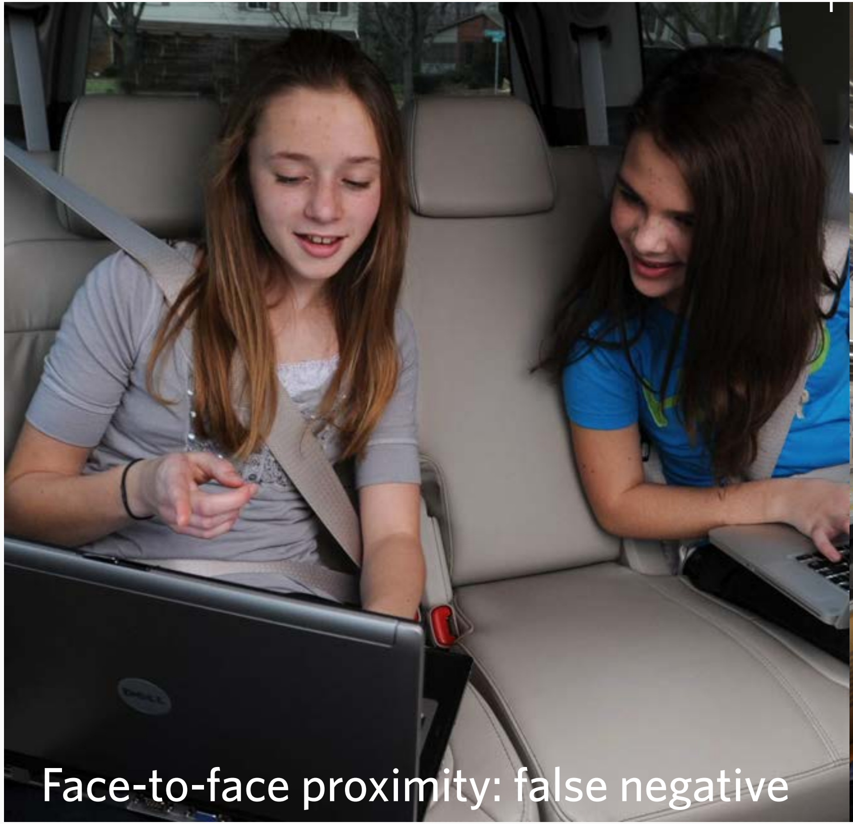
- Copenhagen Networks Study (Bluetooth):
  - "Proximity data"[1]
  - "Face-to-face interactions" [2]
  - "Close proximity interactions" [3]
  - "Face-to-face contacts" [4]
  - "Physical contacts" [5]
- SocioPatterns papers (RFID):
  - "Person-to-person interaction" [6]
  - "Face-to-face contacts" [7]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

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

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



Face-to-face proximity: false negative [1]

Face-to-face proximity: false positive [2]

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

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 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!)

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

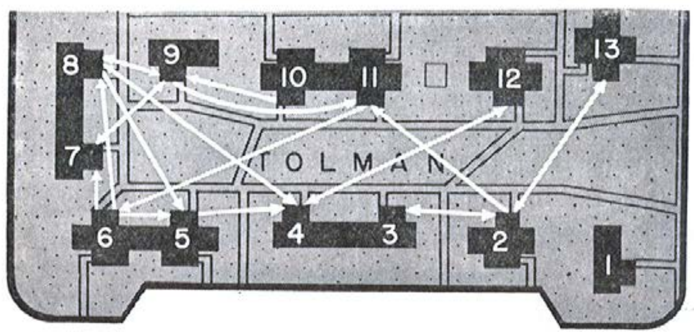


FIG. 9a. Pattern of Sociometric Connections in Tolman Court

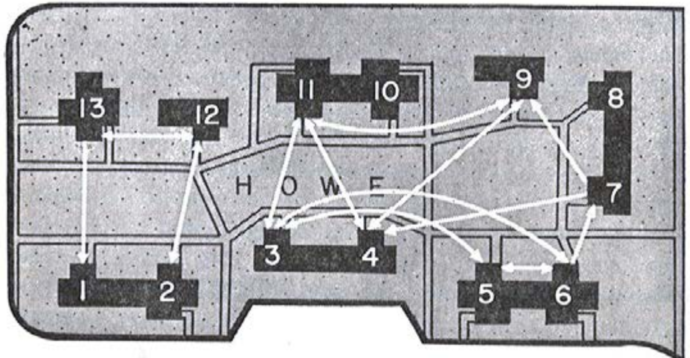


FIG. 9b. Pattern of Sociometric Connections in Howe Court

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

Key SNA move: Compare *types* of ties

Similarities			Social Relations				Interactions	Flows
Location	**Membership**	**Attribute**	**Kinship**	**Other role**	**Affective**	**Cognitive**	e.g.,	e.g.,
e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	e.g.,	Sex with	Information
Same spatial and temporal space	Same clubs	Same gender	Mother of	Friend of	Likes	Knows	Talked to	Beliefs
	Same events	Same attitude	Sibling of	Boss of	Hates	Knows about	Advice to	Personnel
	etc.	etc.		Student of	etc.	Sees as happy	Helped	Resources
				Competitor of		etc.	Harmed	etc.
							etc.	

Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca (2009). Network analysis in the social sciences. *Science* 323, 892–895. https://dx.doi.org/10.1126/science.1165821.

Connect what ties *represent*

- Propinquity a the relationship between a role relation and opportunity structures
- (We could further extend to behavioral interaction or interpersonal sentiments)

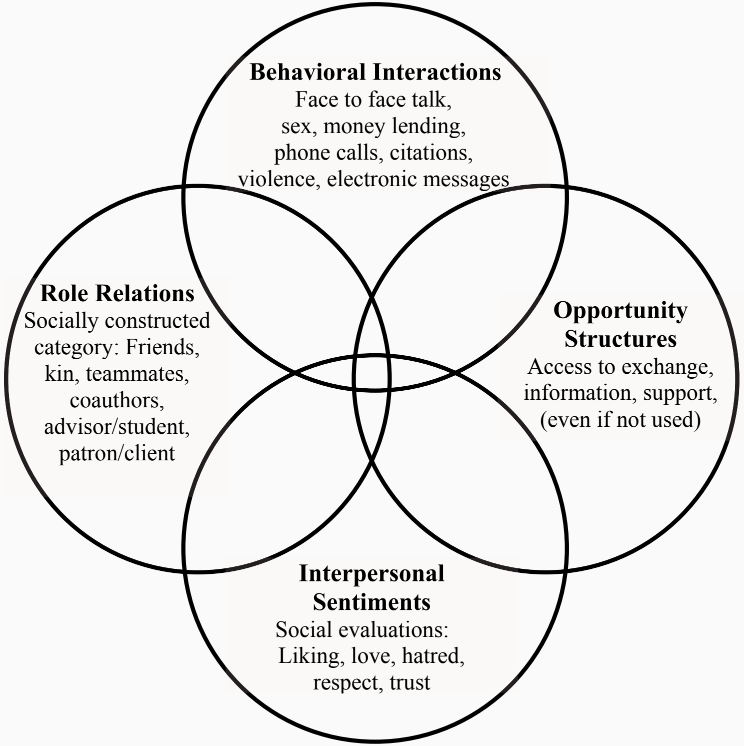


Figure 1. Four conceptualizations of social networks

James A. Kitts and Eric Quintane (2017). Rethinking networks in the era of computational social science. *Oxford Handbook of Social Networks*.

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

(Conversation: The best proxy?)

(Audio work needs updating!)

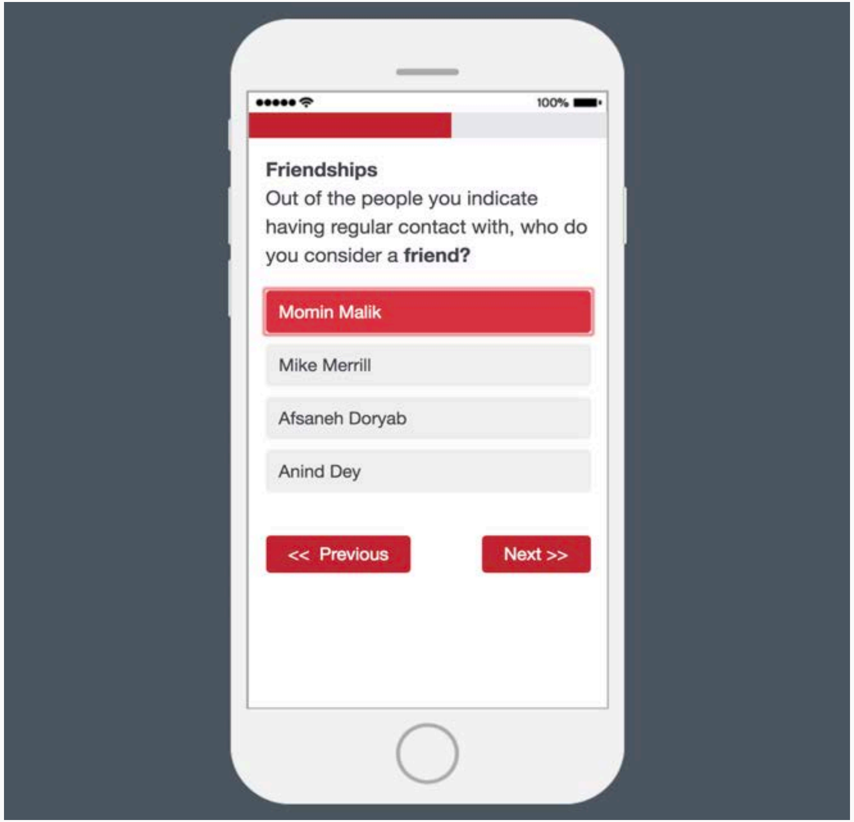
- Earliest work was pre-smartphone
- Most recent work was not audio-only and bulky
- Rich opportunities to revisit

Fig. 6. The MSB. Microphone is at top.

(a) Front: MSB is on right shoulder (b) Back: PDA is in bag. (c) PDA and data collection program.

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. https://dx.doi.org/10.1145/1889681.1889688.

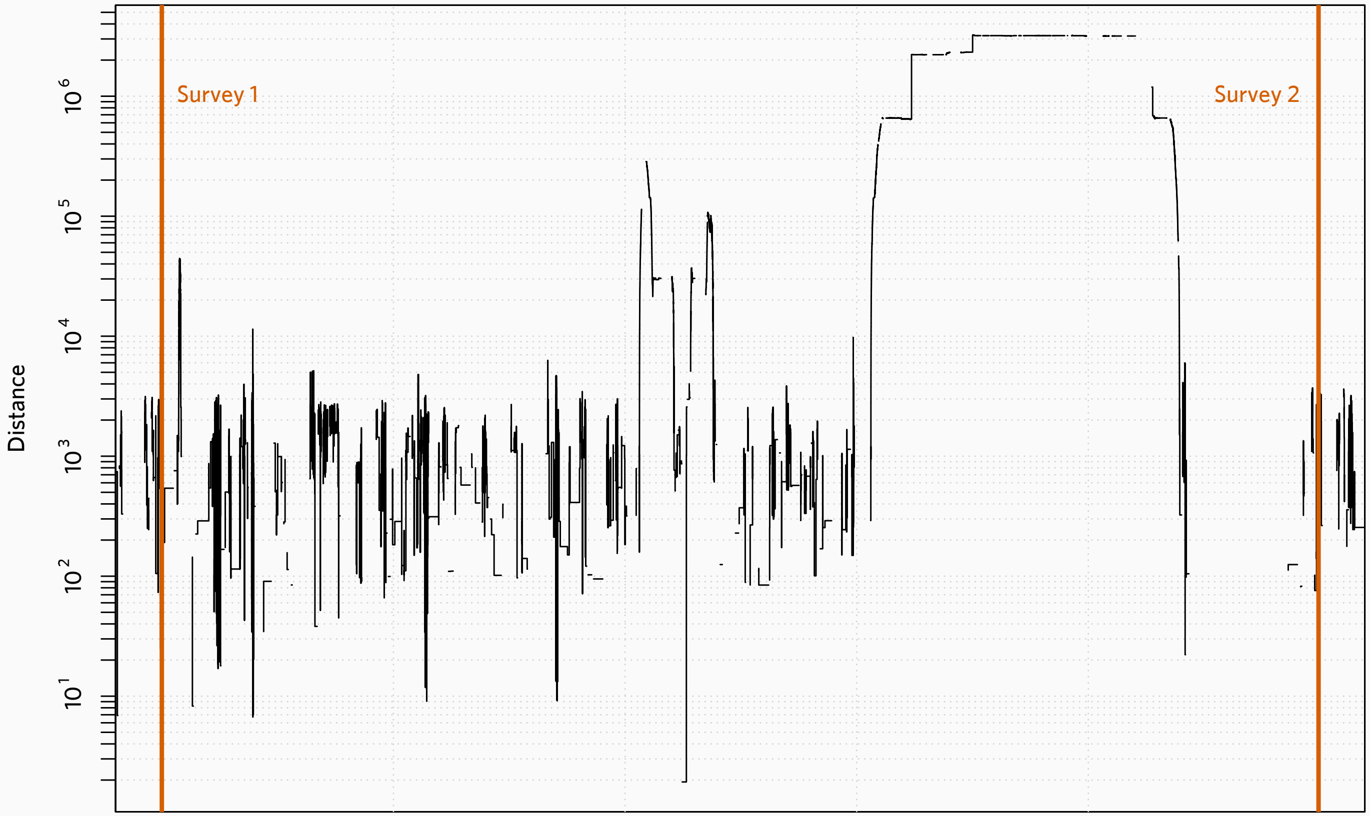
Data: Surveys + mobile phone tracking



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*

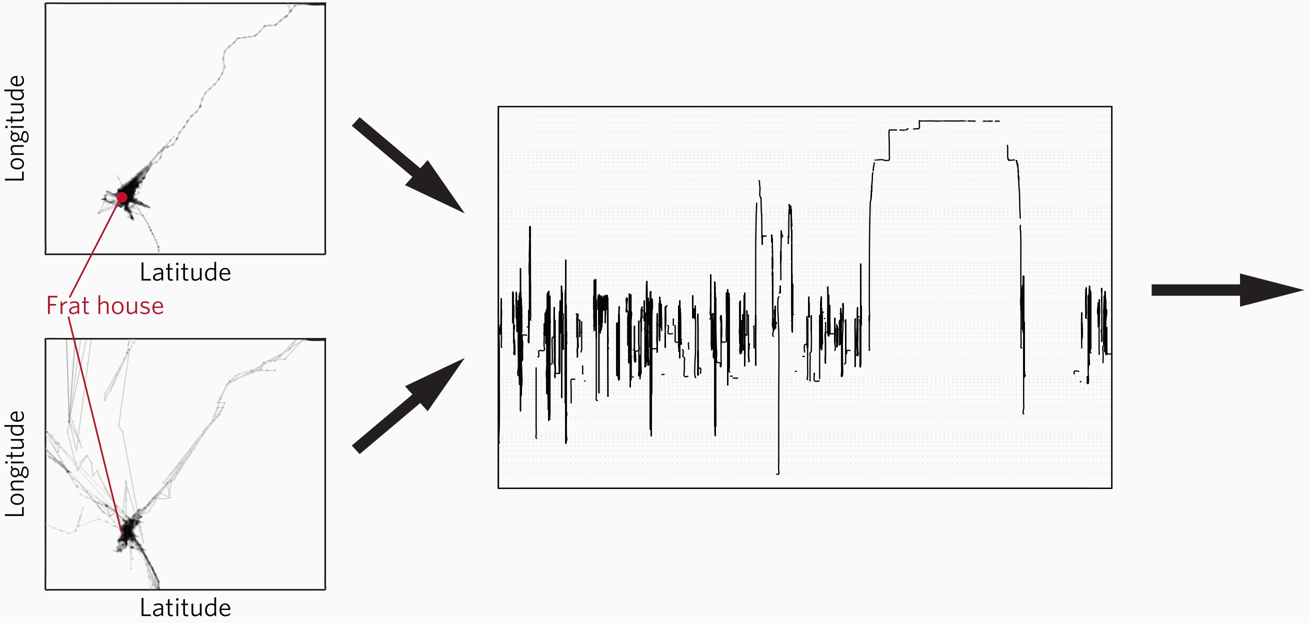


Approach: First do machine learning

- R.A. Fisher (1922): “The purpose of statistics is the reduction of data.”
- Step 1: Find out how to meaningfully characterize the association of proximity and friendship
- Step 2: Using this characterization, model co-evolution

Fisher, Ronald A. (1922). “On the mathematical foundations of theoretical statistics.” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 222, 309–368. https://dx.doi.org/10.1098/rsta.1922.0009.

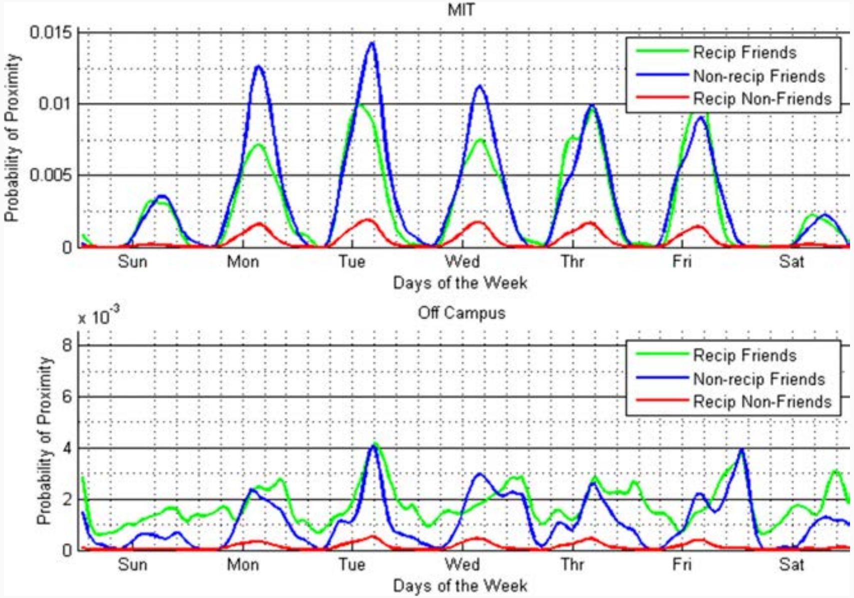
Data processing and “feature extraction”



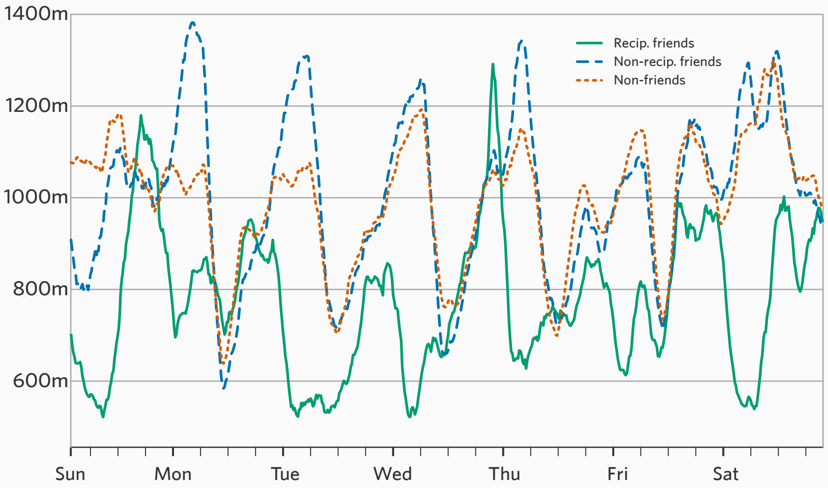
0.086	0.281	0.0793	0.079	
0.005	0.073	0.0054	0.005	
0.057	0.234	0.0547	0.054	• • •
0.007	0.086	0.0074	0.007	
0.071	0.258	0.0669	0.066	
0.024	0.154	0.0238	0.023	
	⋮			

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

"Probability of proximity" (Reality Mining[1])



Median pairwise distance (our study)

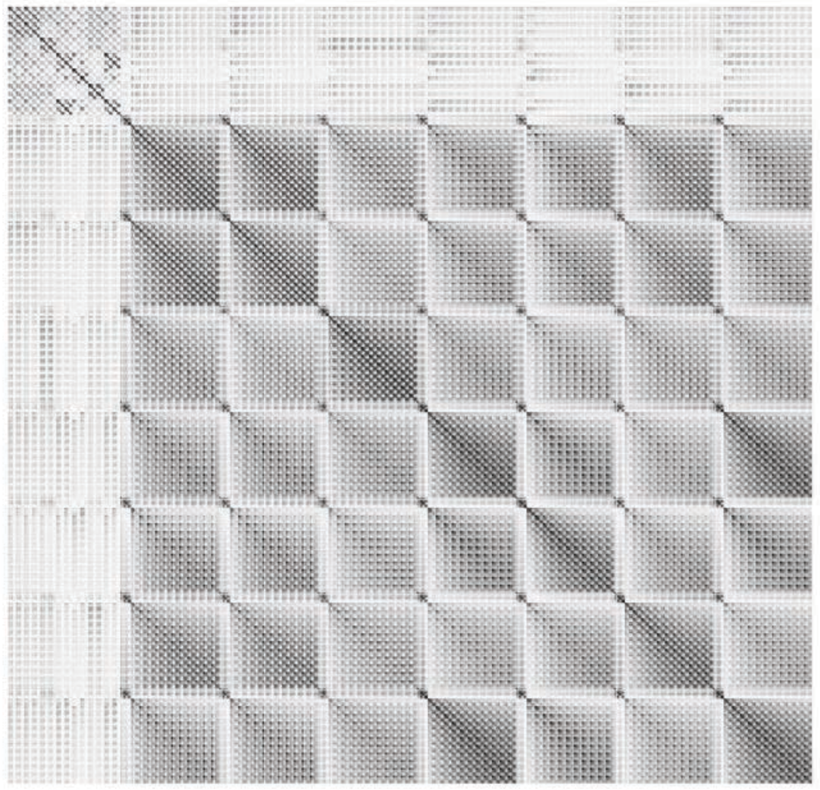


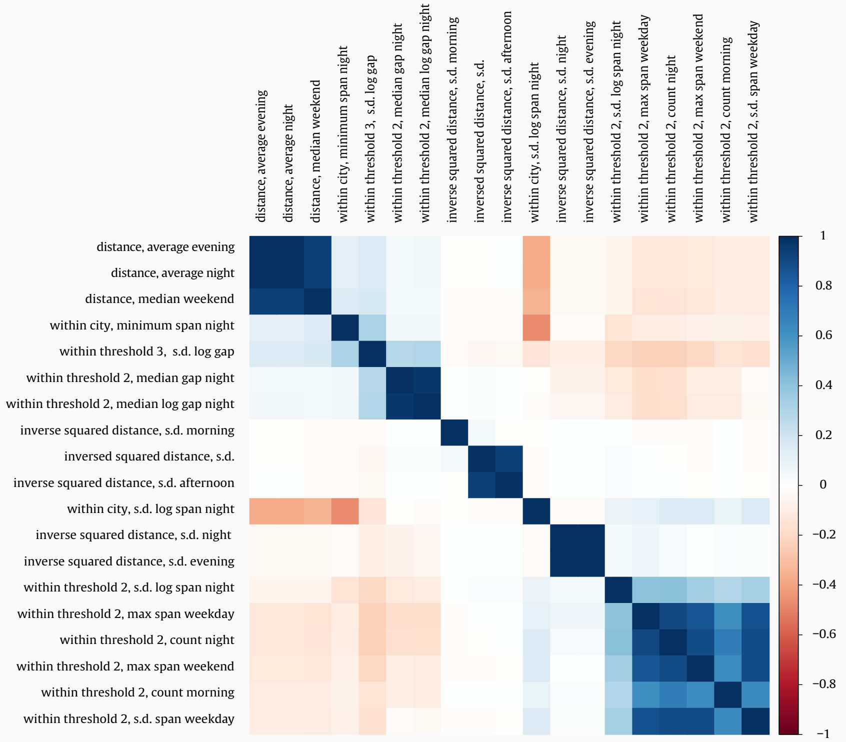
We found what looked like a compelling pattern as well, but it proved ineffective for prediction when tested with cross-validation. Why? Aggregate trends obscure between-dyad and week-to-week variance.

Test the performance via *cross-validation*

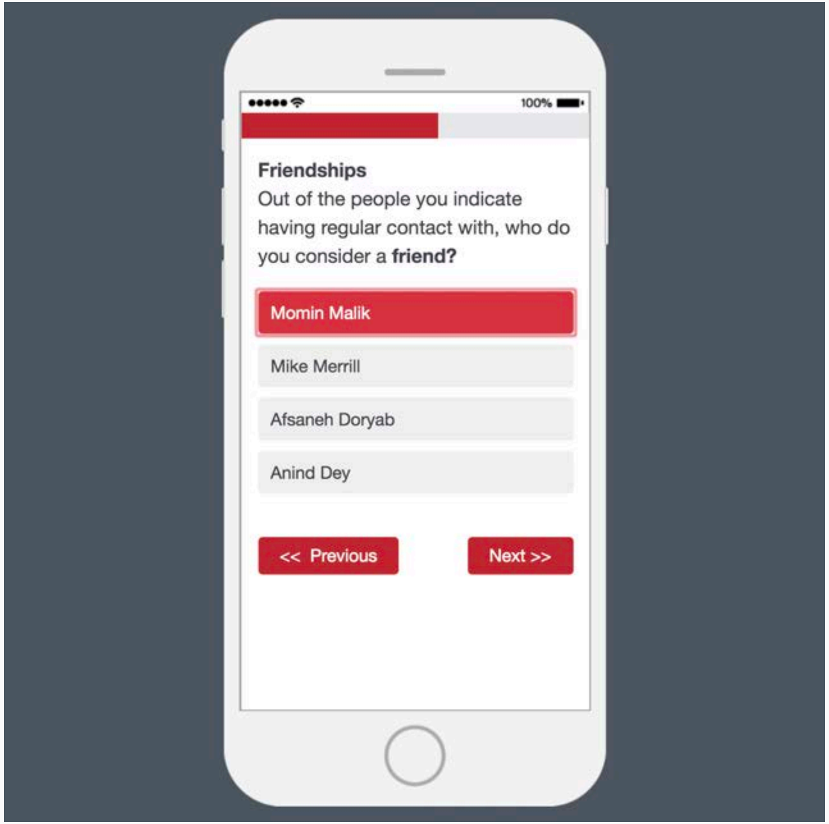
- Split data into “training” and “test”
- Fit model on training, evaluate on test
- Done correctly, simulates out-of-sample data, thereby directly establishing external validity
- But dependencies (e.g. time, networks) can complicate cross-validation
- We use multiple cross-validation schema to control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

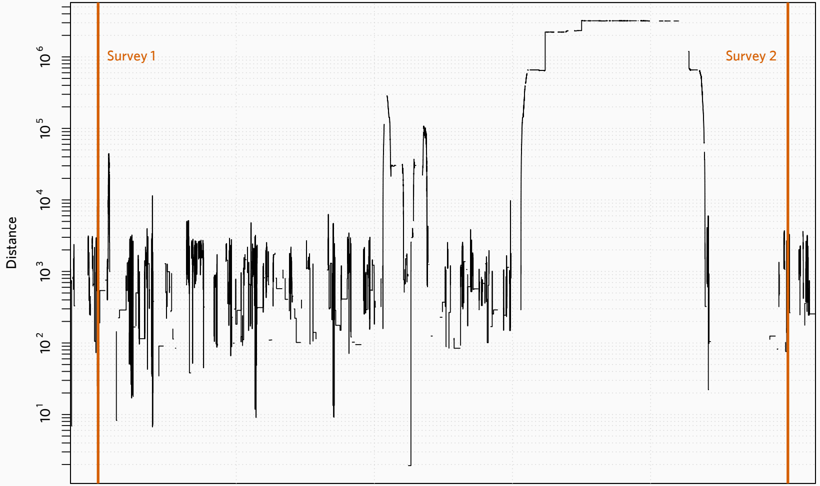
Result: ~30% association. Can get with 2.5K features… or 19, after feature selection.





Form of network surveys: Deliberate



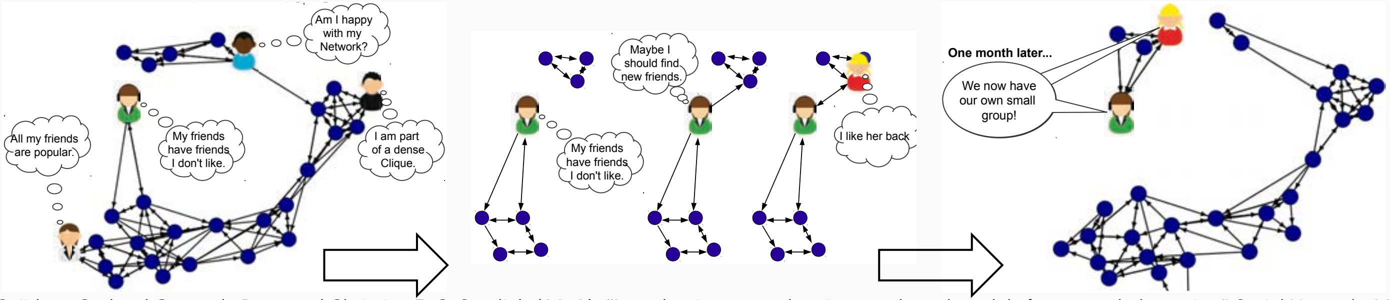


BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY

Surveys based on SAOM studies

Stochastic Actor-Oriented Models (SAOMs) are the only class of models that can handle the co-evolution of network structure and behavior (they require longitudinal data)

- Combines exponential random graph models, choice models, and agent-based simulation... statistically, a doozy
- Increasing work on generalizing SAOMs, with implementations



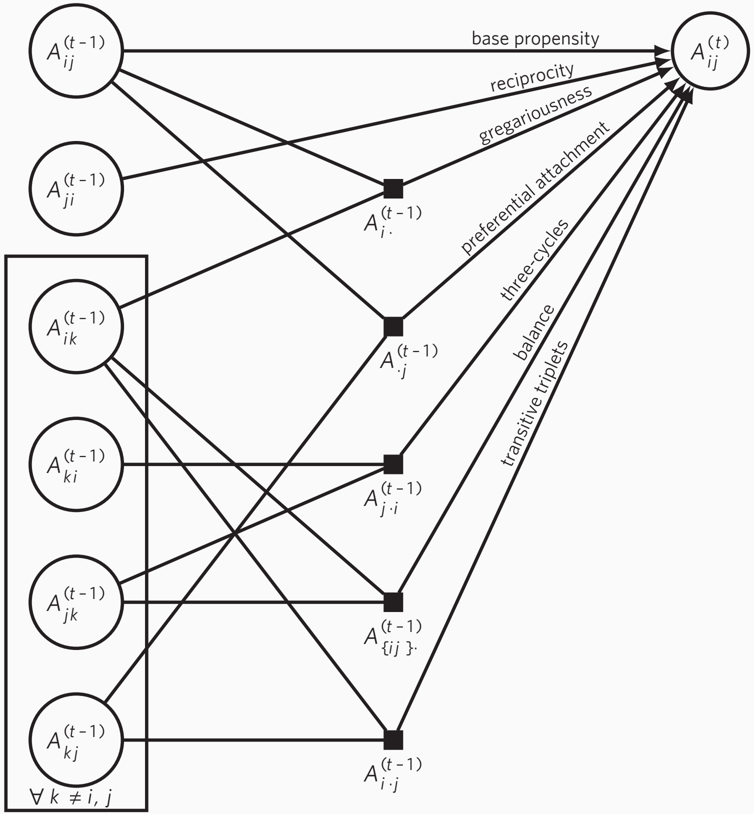
Tom A. B. Snijders, Gerhard G. van de Bunt, and Christian E. G. Steglich (2010). "Introduction to stochastic actor-based models for network dynamics." *Social Networks* 32 (1), 44–60. https://dx.doi.org/10.1016/j.socnet.2009.02.004.

Christian E. G. Steglich, Tom A. B. Snijders, and Michael Pearson (2010). "Dynamics networks and behavior: Separating selection from influence." *Sociological Methodology* 40 (1), 329–393. https://dx.doi.org/10.1111/j.1467-9531.2010.01225.x.

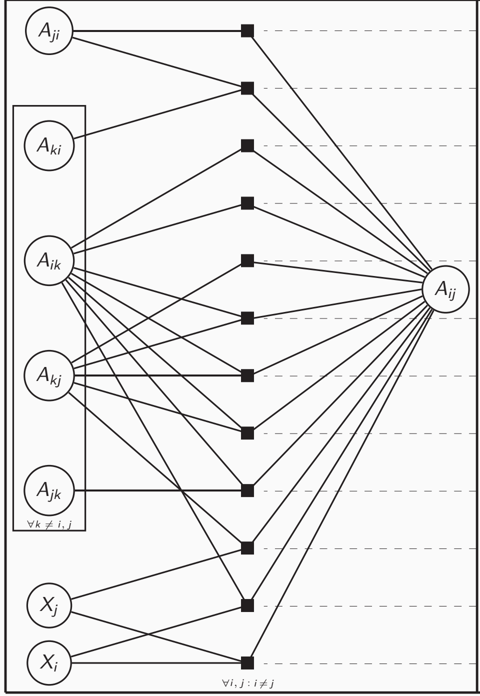
Christoph Stadtfeld and Zsófi Boda (2016). Introduction to SIENA – Part 1. SIENA Workshop, Sunbelt 2016.

(Aside: SAOMs as a graphical model)

- SAOMs can relate to machine learning in another way: probabilistic graphical models
- So far, poor connections between graphical models and network models
- I am hoping this unification will help do inference



(Proper factor graph for ERGMs)

Factor graph	Parameter name	Network Motif	Parameterization	Matrix notation
	mutual dyads		$\sum_{i<j} A_{ij}A_{ji}$	$\frac{1}{2}\operatorname{tr}\left(\mathbf{AA}^T\right)$
	in two stars		$\sum_{(i,j,k)} A_{ji}A_{ki}$	$\operatorname{sum}\left(\mathbf{AA}^T\right)-\operatorname{tr}\left(\mathbf{AA}^T\right)$
	out two stars		$\sum_{(i,j,k)} A_{ij}A_{ik}$	$\operatorname{sum}\left(\mathbf{A}^T\mathbf{A}\right)-\operatorname{tr}\left(\mathbf{A}^T\mathbf{A}\right)$
	geom. weighted out degrees		$\sum_i \exp\left\{-\alpha\sum_k A_{ik}\right\}$	$\operatorname{sum}\left(\exp\{-\alpha\operatorname{rowsum}(\mathbf{A})\}\right)$
	geom. weighted in degrees		$\sum_j \exp\left\{-\alpha\sum_k A_{kj}\right\}$	$\operatorname{sum}\left(\exp\{-\alpha\operatorname{colsum}(\mathbf{A})\}\right)$
	alternating transitive k triplets		$\lambda\sum_{i,j} A_{ij}\left\{1-\left(1-\frac{1}{\lambda}\right)^{\sum_{k\neq i,j} A_{ik}A_{kj}}\right\}$	$\lambda\operatorname{sum}\left(\mathbf{A}\odot\left(1-\left(1-\frac{1}{\lambda}\right)^{\mathbf{AA}-\operatorname{diag}(\mathbf{AA})}\right)\right)$
	alternating indep. two paths		$\lambda\sum_{i,j}\left\{1-\left(1-\frac{1}{\lambda}\right)^{\sum_{k\neq i,j} A_{ik}A_{kj}}\right\}$	$\lambda\operatorname{sum}\left(1-\left(1-\frac{1}{\lambda}\right)^{\mathbf{AA}-\operatorname{diag}(\mathbf{AA})}\right)$
	two paths (mixed two stars)		$\sum_{(i,k,j)} A_{ik}A_{kj}$	$\operatorname{sum}(\mathbf{AA})-\operatorname{tr}(\mathbf{AA})$
	transitive triads		$\sum_{(i,j,k)} A_{ij}A_{jk}A_{ik}$	$\operatorname{tr}\left(\mathbf{AAA}^T\right)$
	activity effect		$\sum_i X_i\sum_j A_{ij}$	$\operatorname{sum}\left(\mathbf{X}\odot\operatorname{rowsum}(\mathbf{A})\right)$
	popularity effect		$\sum_j X_j\sum_i A_{ij}$	$\operatorname{sum}\left(\mathbf{X}\odot\operatorname{colsum}(\mathbf{A})\right)$
	similarity effect		$\sum_{i,j} A_{ij}\left(1-\frac{\lvert X_i-X_j\rvert}{\max_{k,l}\lvert X_k-X_l\rvert}\right)$	$\operatorname{sum}(\mathbf{A}\odot\mathbf{S})$

Tom A. B. Snijders, Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock, 2006, "New specifications for Exponential Random Graph Models." *Sociological Methodology* 36, 99–153.

Ethics: Companies as foil

- Companies are already using digital trace data—I want to know what they can and can't do
- Debunk what they can't do, regulate what they can do
- My study was with a non-vulnerable population. If it wasn't, I would be far more cautious
- Who is left out is important. See Frances Cherry's (1995) critique of Festinger et al. (1950): they ignored women!

theguardian

Your search terms | Search

Stop complaining about the Facebook study. It's a golden age for research

We should *insist* that Facebook do experiments on the decisions it's already making for us. Anything else would be unethical

Duncan J Watts
theguardian.com, Monday 7 July 2014 07.45 EDT

The editor of the journal that published the Facebook study now calls it 'an important and emerging area of social science research that needs to be approached with sensitivity.' Photograph: Jeff Chiu / AP

Frances Cherry (1995). "One man's social psychology is another woman's social history." In *The stubborn particulars of social psychology: Essays on the research process*, pp. 68–83. London: Routledge.

Ethics of audio collection?

Summary: How we *should* use sensors

- If using Bluetooth, RFID proxies for interaction, do more testing against human-coded benchmarks
- But *proximity* (a connection of role relations and opportunity structures) 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
- Future: use conversation add in behavioral interaction?

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@cyber.harvard.edu>
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 9

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 20

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.