Measuring tie strength

Matthew J. Salganik

Social Network (Soc 204)
Spring 2017
Princeton University

April 5, 2017

Logistics:

- No logistics

Internet Privacy Technology and Policy: What Lies Ahead?

"Last November, the Federal Communications Commission (FCC) issued a landmark privacy rule governing how Internet service providers (ISPs) could collect and share customer data. On April 4, 2017, President Trump signed a joint resolution that repealed this rule before it could ever take effect. This panel will discuss how we arrived at this juncture and how the Internet privacy landscape may evolve in light of these developments. We will also explore the roles (and shortcomings) of both policy and technical mechanisms in protecting user privacy on the Internet."

Jennifer Rexford, Nick Feamster, Ed Felten, Arvind Naryanan, Joel Reidenberg

Wednesday, April 5, 4:30 p.m. - 5:30 p.m., 3rd floor Sherrerd Hall

`http://citp.princeton.edu/event/internet-privacy/`

Great discussion on Piazza:

Great discussion on Piazza:

- new suggestion (use likes to infer ideology)

Great discussion on Piazza:

- new suggestion (use likes to infer ideology)
- conflict of interest (my poor attempt to discuss normative issues)

Great discussion on Piazza:

- new suggestion (use likes to infer ideology)
- conflict of interest (my poor attempt to discuss normative issues)
- big gap between data and conclusions in many studies

Great discussion on Piazza:

- new suggestion (use likes to infer ideology)
- conflict of interest (my poor attempt to discuss normative issues)
- big gap between data and conclusions in many studies
- fundamental limitations so researchers should work on new problems or develop new methods

Great discussion on Piazza:

- new suggestion (use likes to infer ideology)
- conflict of interest (my poor attempt to discuss normative issues)
- big gap between data and conclusions in many studies
- fundamental limitations so researchers should work on new problems or develop new methods
- other kinds of google filter bubbles (political terms vs terms with multiple meanings)

Favorite reading:

1. Marsden, P.V. and Campbell, K.E. (1984) Measuring tie strength. Social Forces.
2. Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
3. Jones, J.J., et al. (2013). Inferring tie strength from online directed behavior. PLOS One.

Review:

- Monday: Strength of weak ties
- Wednesday: Filter bubbles
- Monday: Breaking your filter bubble
- Wednesday: Measuring weak ties

- Using models in the real world requires measurement

- Using models in the real world requires measurement
- Measurement often involves statistics

- Using models in the real world requires measurement
- Measurement often involves statistics
- Statistics is not a prerequisite for this class, but it is for doing social network research

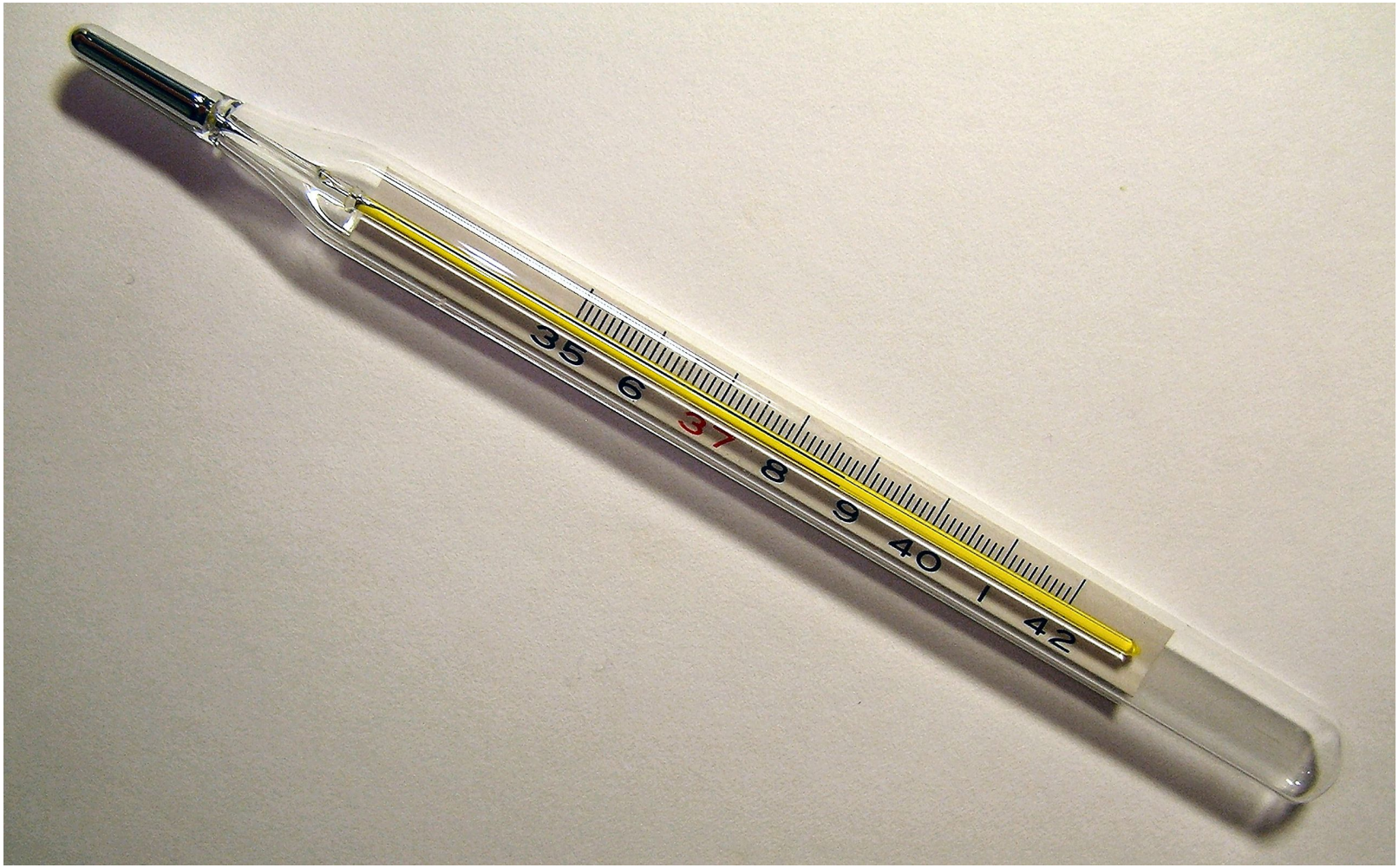
- Using models in the real world requires measurement
- Measurement often involves statistics
- Statistics is not a prerequisite for this class, but it is for doing social network research
- If you want to learn more statistics, you can take POL 345/SOC 305 in the fall (I'm teaching it)

- Using models in the real world requires measurement
- Measurement often involves statistics
- Statistics is not a prerequisite for this class, but it is for doing social network research
- If you want to learn more statistics, you can take POL 345/SOC 305 in the fall (I'm teaching it)
- Since this is the first set of statsy papers we have read, I'm going to review them a bit more than normal, but I won't do this every time

"Strength of a tie is a (probably linear) combination of the amount of time, emotional intensity, the intimacy, and reciprocal services which characterize the tie."
- Granovetter (1973)

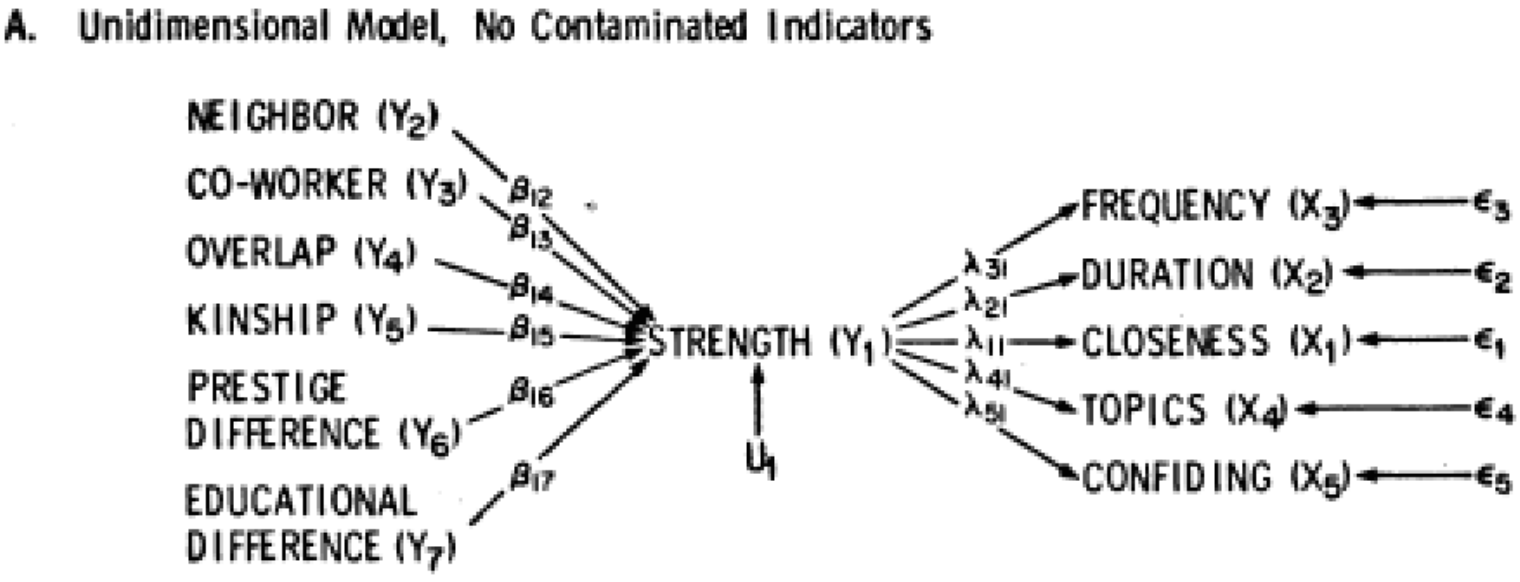
Measuring Tie Strength*

PETER V. MARSDEN, *University of North Carolina at Chapel Hill*
KAREN E. CAMPBELL, *Vanderbilt University*



https://commons.wikimedia.org/wiki/File:Clinical_thermometer_38.7.JPG

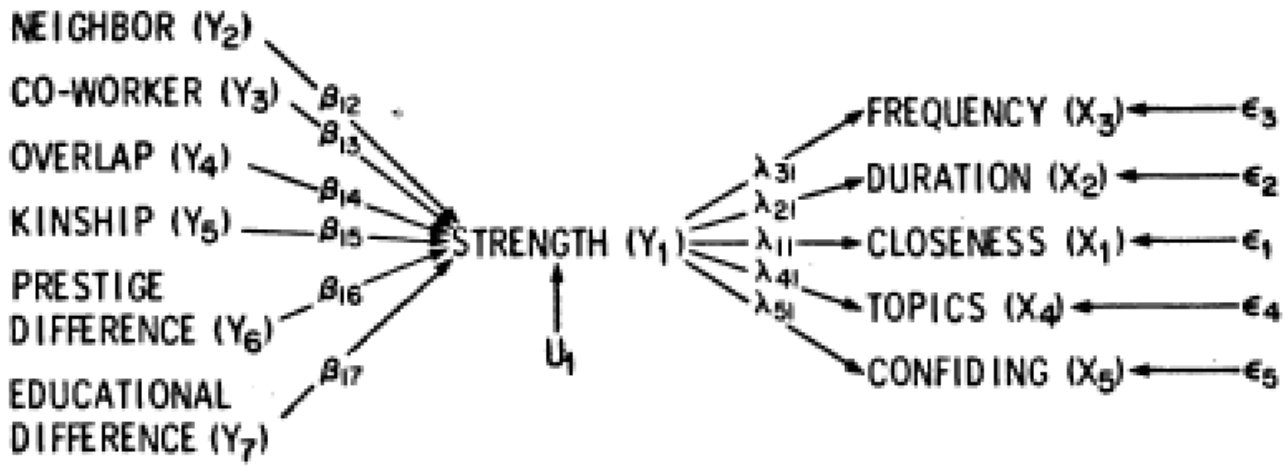
Assume that tie strength exists
look at predictors (left) and indicators (right)



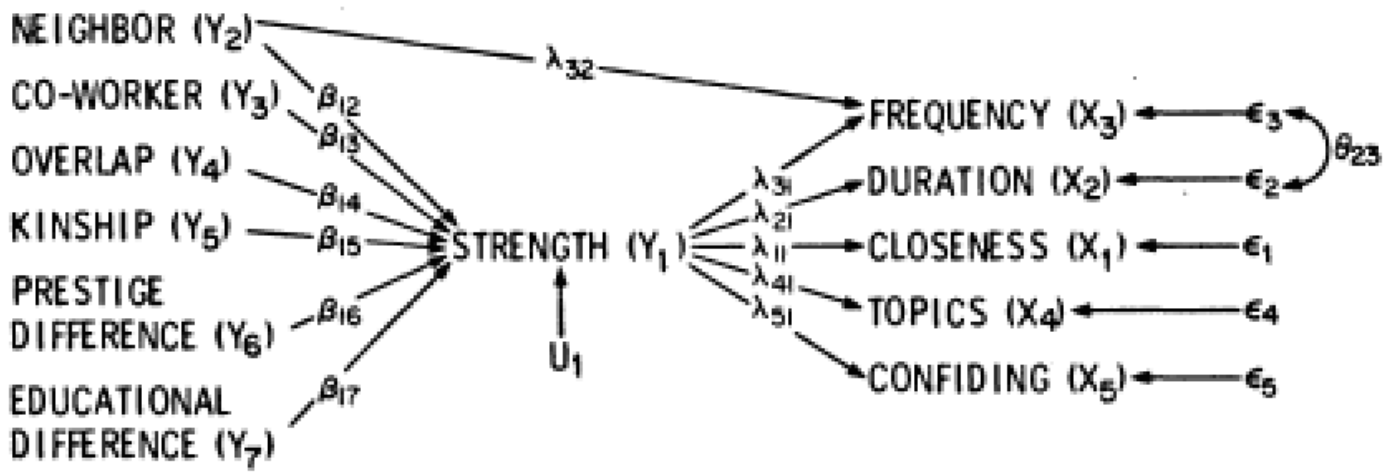
Is kinship a predictor or indicator?

Predictors (left) and indicators (right)

A. Unidimensional Model, No Contaminated Indicators

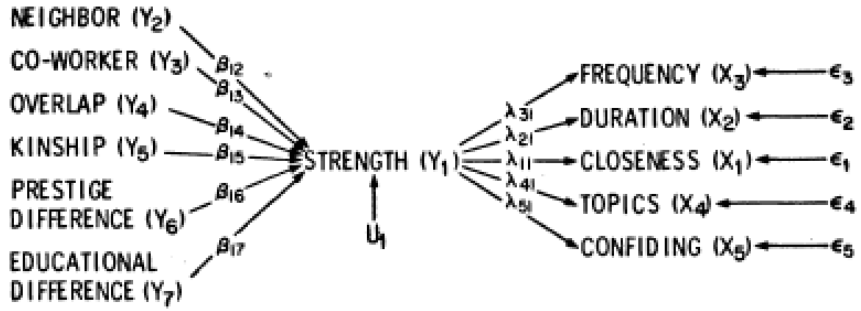


B. Unidimensional Model, with Two Types of Contamination of Indicators

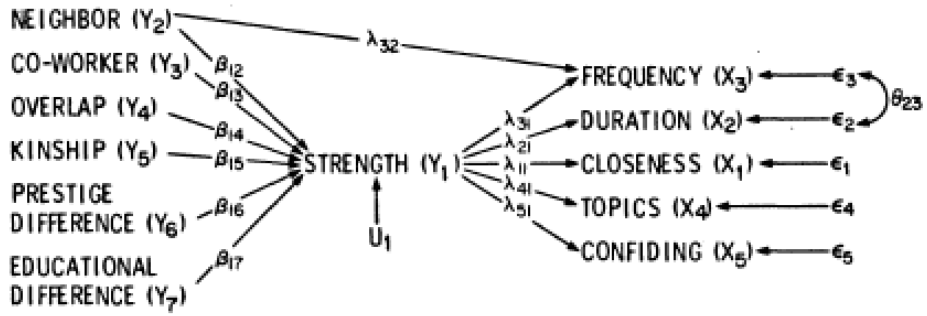


Predictors (left) and indicators (right)

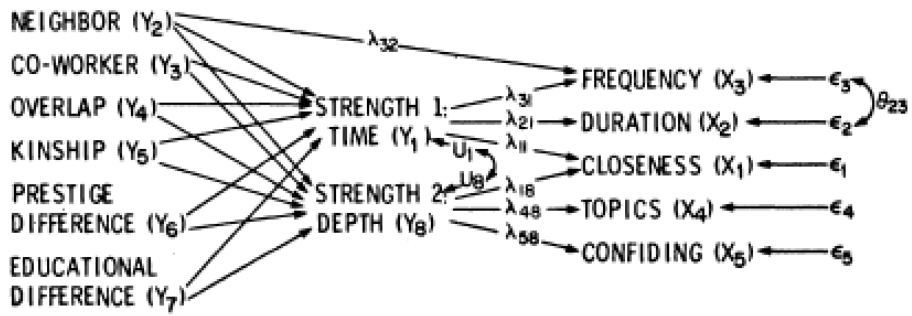
A. Unidimensional Model, No Contaminated Indicators



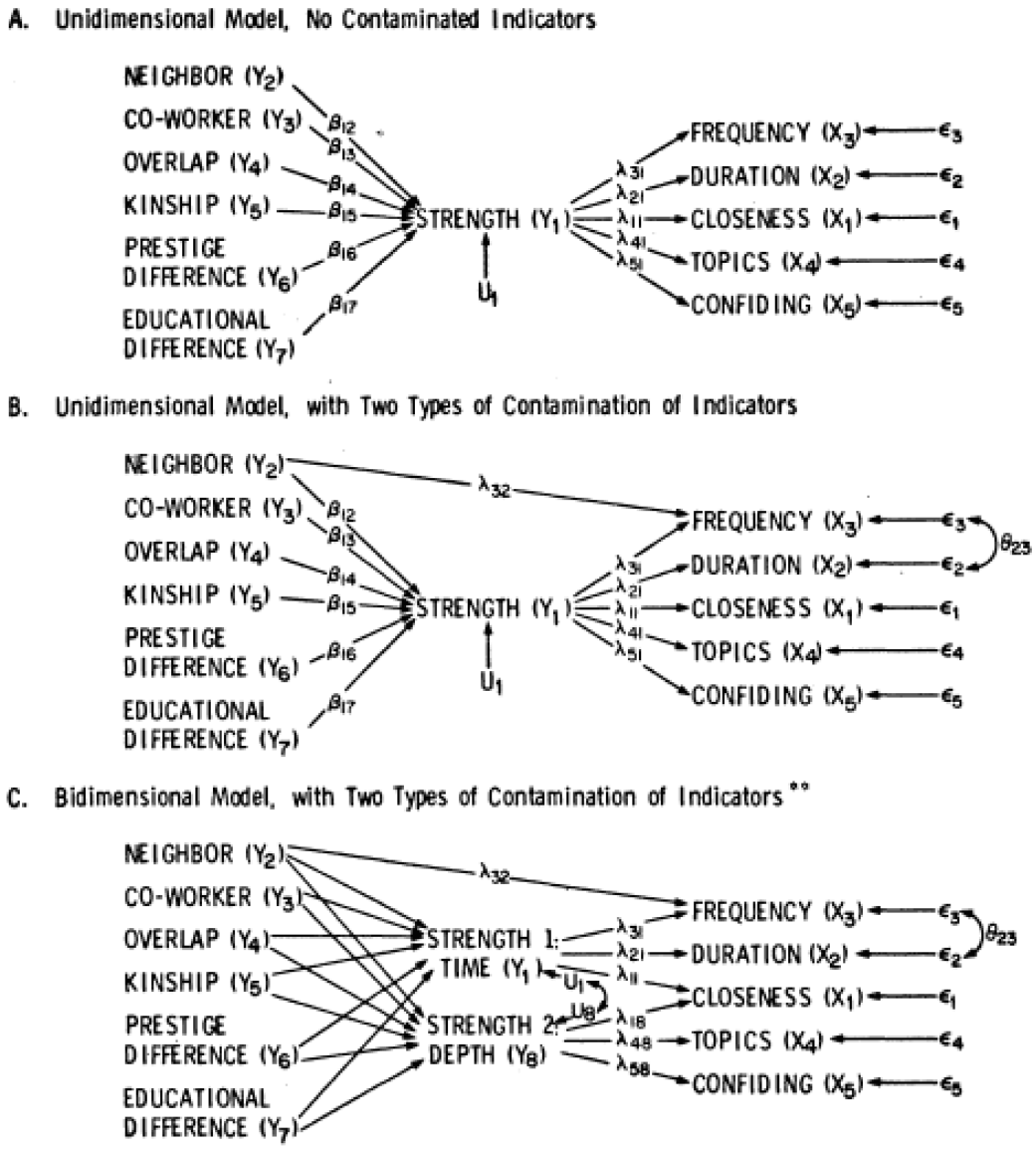
B. Unidimensional Model, with Two Types of Contamination of Indicators



C. Bidimensional Model, with Two Types of Contamination of Indicators **



Predictors (left) and indicators (right)



Example of two dimensional construct: I'm socially liberal but fiscally conservative

Three surveys that collected information on three closest friends

- Detroit, MI 1965-66
- Aurora, IL 1974-75
- Altneustadt, Germany 1971

Note: None of this data collected by Marsden and Campbell, most collected before Granovetter 1973

The results we have presented lead, as we have mentioned, to one fairly clear conclusion: a measure of "closeness," or the emotional intensity of a relationship, is on balance the best indicator of the concept of tie strength among those available to us. This measure is free of contamination by

We found that the two measures having to do with the time spent in a relationship—duration and frequency of contact—were badly contaminated by measures of foci around which ties may be organized. In particular, the use of frequency as a measure of strength will tend systematically to overestimate the strength of ties between persons who are neighbors or co-workers, while the use of duration as a measure of strength will overestimate the strength of ties between relatives. For this

Our final point concerns the possibility of using predictors themselves to form a proxy for tie strength. The results given in panel C of

- Marsden and Campbell (1984): tie strength is unobserved, try to understand it

- Marsden and Campbell (1984): tie strength is unobserved, try to understand it
- Gilbert and Karahalios (2009) & Jones et al (2013): measure tie strength in a survey then predict it using trace data

Which paper do you think is better?

Predicting Tie Strength With Social Media

Eric Gilbert and Karrie Karahalios
University of Illinois at Urbana-Champaign
[egilber2, kkarahal]@cs.uiuc.edu

Inferring Tie Strength from Online Directed Behavior

Jason J. Jones[1,2]*, Jaime E. Settle[2], Robert M. Bond[2], Christopher J. Fariss[2], Cameron Marlow[3], James H. Fowler[1,2]

1 Medical Genetics Division, University of California, San Diego, La Jolla, California, United States of America, **2** Political Science Department, University of California, San Diego, La Jolla, California, United States of America, **3** Data Science, Facebook, Inc., Menlo Park, California, United States of America

Predicting Tie Strength With Social Media

Eric Gilbert and Karrie Karahalios
University of Illinois at Urbana-Champaign
[egilber2, kkarahal]@cs.uiuc.edu

- 35 participants brought into a lab
- only has access to data participants have (no "behind-the-scenes" data)
- about 60 randomly chosen friends per participant

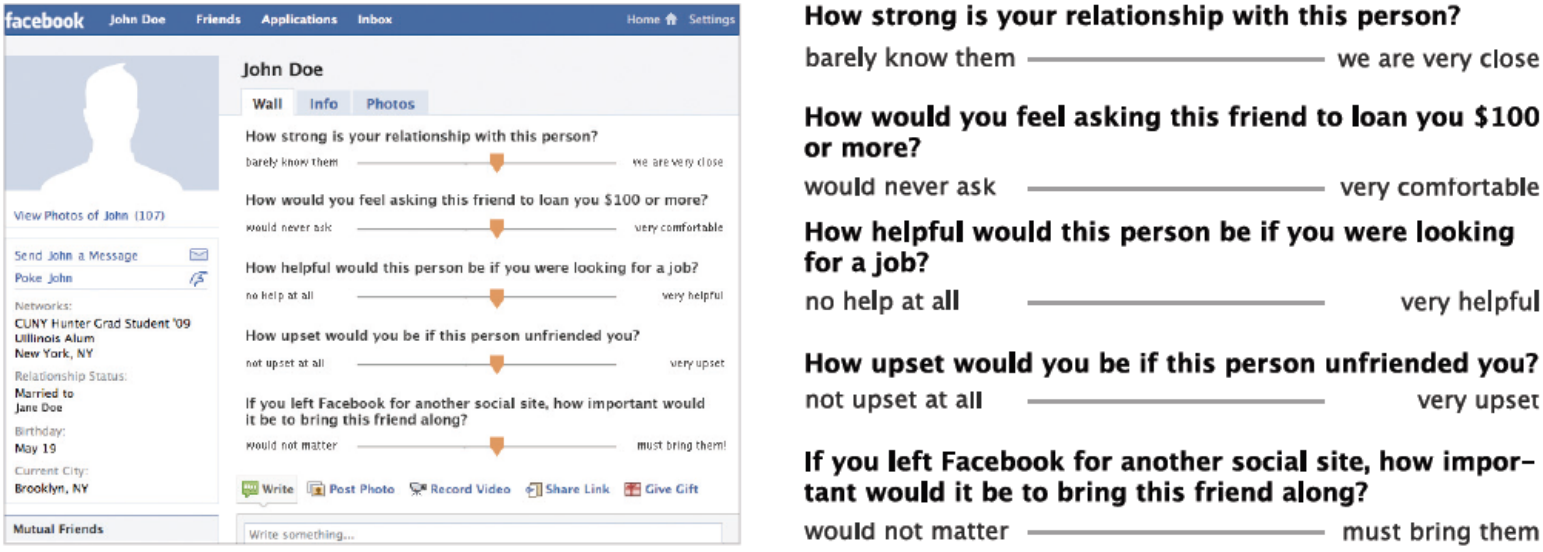


Figure 1. The questions used to assess tie strength, embedded into a friend's profile as participants experienced them. An automated script guided participants through a random subset of their Facebook friends. As participants answered each question by dragging a slider, the script collected data describing the friendship. The questions reflect a diversity of views on tie strength.

Focused on first question

74 variables $\rightarrow$ statistical model $\rightarrow$ estimated tie strength

74 variables $\rightarrow$ statistical model $\rightarrow$ estimated tie strength
Examples:

- Wall words exchanged, inbox messages (similar to "intensity" in Granovetter (1973))
- Number of mutual friends (not in Granovetter's definition, causal ordering unclear)

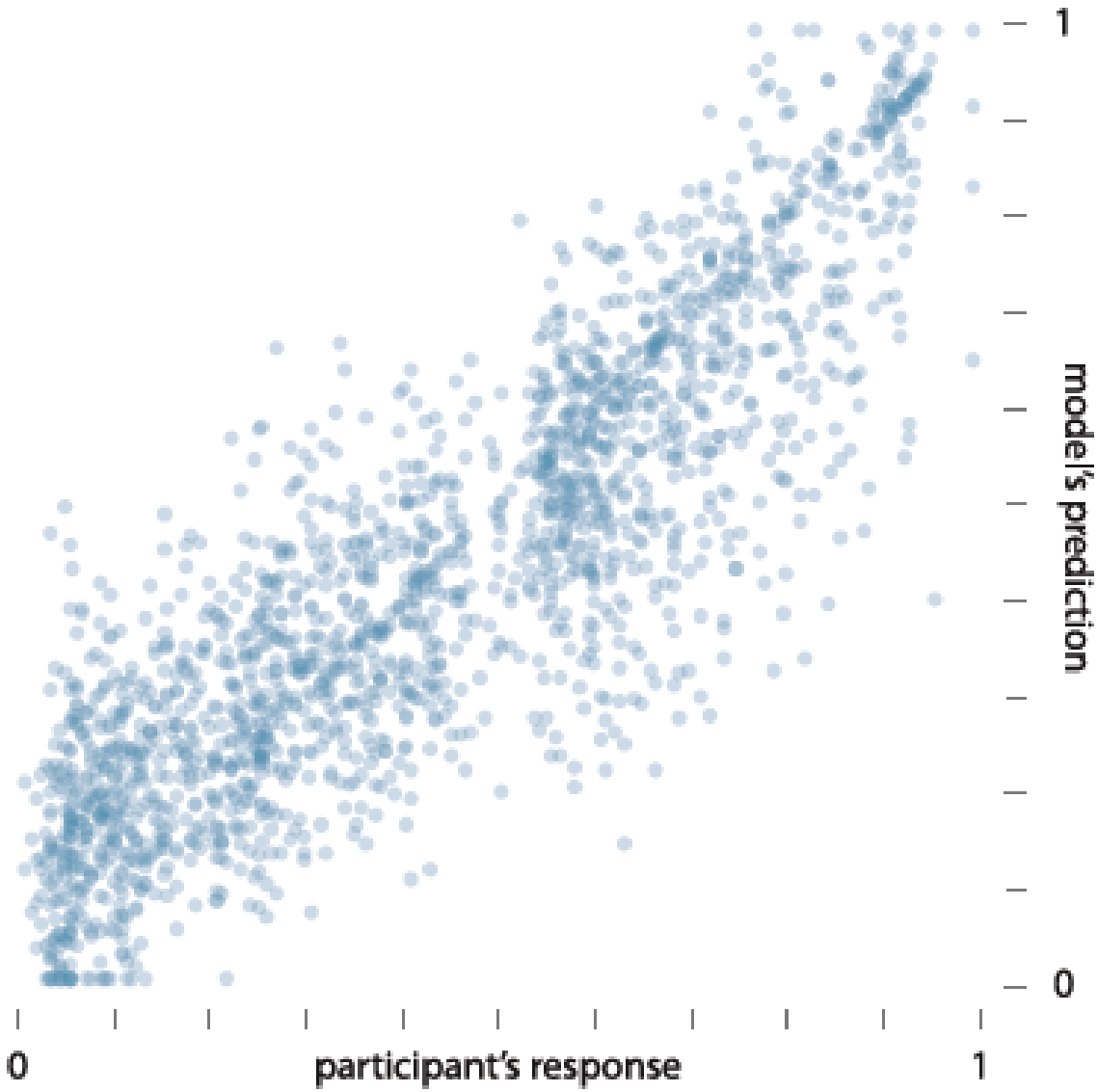


Figure 4. The model's performance across all ties in our dataset. There is a strong correlation, yet the model shows a slight bias toward underestimation, represented as the larger cloud in the bottom-right of the figure. The gap in the center results from participants' inclination to move the slider from its starting point, if only slightly.

This friend is very special. He and I attended the same high school, we interacted a lot over 3 years and we are very very close. We trust each other. My friend are I are still interacting in ways other than Facebook such as IM, emails, phones. Unfortunately, that friend and I rarely interact through Facebook so I guess your predictor doesn't have enough information to be accurate.

rating: 0.96; prediction: 0.47

Off channel behavior

Inferring Tie Strength from Online Directed Behavior

Jason J. Jones[1,2]*, Jaime E. Settle[2], Robert M. Bond[2], Christopher J. Fariss[2], Cameron Marlow[3], James H. Fowler[1,2]

1 Medical Genetics Division, University of California, San Diego, La Jolla, California, United States of America, **2** Political Science Department, University of California, San Diego, La Jolla, California, United States of America, **3** Data Science, Facebook, Inc., Menlo Park, California, United States of America

- 1,565 people recruited with Facebook ads, half responses were useful
- 1 closest friend and 1 random friend

Table 1. The names and descriptions of features used to estimate tie strength.

Feature	Description
Comments	The number of comments the survey-taker made on an object (a post, photo, link, etc.) owned by the friend.
Messages	The number of private messages (similar to email, but sent through facebook.com) the survey taker sent to the friend. Messages are visible only to the sender and recipient.
Wall Posts	The number of posts the survey-taker made on the friend's Facebook.com wall. Wall posts are visible to the friend's friends.
Likes	The number of times the survey-taker liked an object owned by the friend.
Photo Tags	The number of times the survey-taker tagged the friend in a photo posted on facebook.com.
Same Photo	The number of photos in which both the survey-taker and the friend were tagged (regardless of who performed the tagging).
Pokes	The number of times the survey-taker "poked" the friend through facebook.com.
Family Edges	The number of times the survey-taker requested the friend add to their profile a familial relationship to the survey-taker.
Event Invites	The number of times the survey-taker invited the friend to an event through facebook.com.
Same Gender	A binary variable that is 0 if the survey-taker and friend are of different genders or 1 if the same gender.
Same Employer	A binary variable that is 1 if the survey-taker and friend include the same employer in their facebook.com employment history and 0 otherwise.
Same School	A binary variable that is 1 if the survey-taker and friend ever attended the same school according to their facebook.com academic history and 0 otherwise.
Group Invites	The number of times the survey-taker invited the friend to join an online interest group.
Age Difference	The absolute value of the age difference between the survey-taker and friend.

doi:10.1371/journal.pone.0052168.t001

Note: no connection to theoretical constructs

Table 2. Descriptive statistics of the feature variables for Closest Friend dyads and Not Closest Friend dyads.

	Closest Friends		Not Closest Friends		
Feature	**Mean**	**SEM**	**Mean**	**SEM**	**r**
Comments	37.51	2.16	1.99	0.27	0.66
Messages	27.37	3.62	0.64	0.14	0.60
Wall Posts	7.01	0.53	0.32	0.04	0.60
Likes	22.25	1.74	1.83	0.30	0.59
Photo Tags	11.42	1.16	0.29	0.05	0.57
Same Photo	5.89	0.66	0.15	0.03	0.50
Pokes	10.05	1.80	0.14	0.05	0.31
Family Edges	0.21	0.02	0.02	0.005	0.25
Event Invites	0.56	0.05	0.19	0.02	0.21
Same Gender	0.69	0.02	0.55	0.02	0.15
Age Difference	4.96	0.39	8.08	0.51	-0.14
Same Employer	0.04	0.001	0.01	0.002	0.11
Same School	0.33	0.02	0.28	0.02	0.05
Group Invites	0.07	0.02	0.04	0.01	0.04

The final column contains the Spearman rank correlation of each feature with the Closest Friend target variable. Interaction counts represent the total interactions in the six months prior to an online survey that asked users to identify their closest friends in real life.

doi:10.1371/journal.pone.0052168.t002

Table 4. Confusion matrix for the complete additive model.

	Closest	*Not Closest*
Predicted Closest	582	47
Predicted Not Closest	207	742

Columns represent ground truth and rows represent predicted values.
doi:10.1371/journal.pone.0052168.t004

Simple model of with one predictor—interactions—does well. In this context, probably not contaminated by living in the same neighborhood.

Predicting Tie Strength With Social Media

Eric Gilbert and Karrie Karahalios
University of Illinois at Urbana-Champaign
[egilber2, kkarahal]@cs.uiuc.edu

Inferring Tie Strength from Online Directed Behavior

Jason J. Jones[1,2]*, Jaime E. Settle[2], Robert M. Bond[2], Christopher J. Fariss[2], Cameron Marlow[3], James H. Fowler[1,2]

1 Medical Genetics Division, University of California, San Diego, La Jolla, California, United States of America, 2 Political Science Department, University of California, San Diego, La Jolla, California, United States of America, 3 Data Science, Facebook, Inc., Menlo Park, California, United States of America

Key points:

- measurement is important and usually requires statistics

Key points:

- measurement is important and usually requires statistics
- analog age and digital age approaches have different emphasis

Key points:

- measurement is important and usually requires statistics
- analog age and digital age approaches have different emphasis
- it is possible to infer tie strength from digital traces (it turns out that you can infer other things from digital traces too)

http://bit.ly/socnet204

http://bit.ly/socnet204

- Fountain, H. (2006). The lonely American just got a bit lonelier. The New York Times.
- McPherson, M., Smith-Lovin, L., and Brashears, M.E. (2006). Social isolation in America: Changes in core discussion networks over two decades. American Sociological Review.