Can social media predict the closeness of your friendships?
WE’RE NOURYA AND HANNAH
These are the talking points:

01 Intro to Social ties and Social Media
02 Methods of the Paper
03 Results of the Paper
04 Discussion and Applications
Why do we care about social media? What does the literature say so far?
Guess how many social media users worldwide
Guess what percentage of Americans are active social media users
Guess how many hours are spent by the average internet user on social media?
### Some Facts About Social Media

<table>
<thead>
<tr>
<th>Facts</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Billion social media users worldwide</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>of Americans are active social media users</td>
<td>74%</td>
<td></td>
</tr>
<tr>
<td>Spent by the average internet user on social media per day</td>
<td>2.5 hrs</td>
<td></td>
</tr>
</tbody>
</table>
Given how important social media is and how much data is collected...
What can your social media say about your friendships?
How many of you guys are friends with someone on Facebook or follow someone on Instagram that you don’t actually know and have never met?
Definitions of ‘Friend’

Tie strength isn’t a binary 1 or 0…

Granovetter claimed that strengths of ties are continuous… what features are most important measures of friendship/strength of ties?
Let’s say there is a model that can accurately predict the strength of social ties down to the decimal...
Discuss the following questions:

If successful, what can be done with strength of ties for everyone on Facebook? How about TikTok, Twitter, LinkedIn...?

Who benefits from a model that predicts social ties for each of these platforms? Are they different?

Given collaboration between platforms, could there be even stronger predictors of the nature and strength of ties over time?
Facebook can predict if your relationship will last

“a couple who can't be identified through dispersion score is significantly more likely to break up -- at least, on Facebook -- within 60 days.”  (Backstrom & Kleinberg, 2014)

Dispersion: "the extent to which two people's mutual friends are not themselves well-connected."
Methods
SOME FACTS ABOUT THE DATA

26
Median age

32
people studied

62
Friends rated each
Variables Used:

**Predictive Intensity Variables**
- Wall words exchanged: 9549
- Participant-initiated wall posts: 55
- Friend-initiated wall posts: 47
- Inbox messages exchanged: 9
- Inbox thread depth: 31
- Participant's status updates: 80
- Friend's status updates: 200
- Friend's photo comments: 1352

**Intimacy Variables**
- Participant's number of friends: 729
- Friend's number of friends: 2050
- Days since last communication: 1115
- Wall intimacy words: 148
- Inbox intimacy words: 137
- Appearances together in photo: 73
- Participant's appearances in photo: 897
- Distance between hometowns (mi): 8182
- Friend's relationship status: 8%

**Duration Variable**
- Days since first communication: 1328

**Emotional Support Variables**
- Wall & inbox positive emotion words: 197
- Wall & inbox negative emotion words: 51

**Reciprocal Services Variables**
- Links exchanged by wall post: 688
- Applications in common: 18

**Social Distance Variables**
- Age difference (days): 5995
- Number of occupations difference: 8
- Educational difference (degrees): 3
- Overlapping words in religion: 2
- Political difference (scale): 4

**Structural Variables**
- Number of mutual friends: 206
- Groups in common: 12
- Norm. TF-IDF of interests and about: 73
Variables Predicted:

How strong is your relationship with this person?
- barely know them
- we are very close

How would you feel asking this friend to loan you $100 or more?
- would never ask
- very comfortable

How helpful would this person be if you were looking for a job?
- no help at all
- very helpful

How upset would you be if this person unfriended you?
- not upset at all
- very upset

If you left Facebook for another social site, how important would it be to bring this friend along?
- would not matter
- must bring them!
Choosing variables: discuss!

- The participants are all from University of Illinois community...
- Would you test the same variables to predict the strength of social ties? Or propose different variables?
- What other data might Facebook use to predict tie strength?
- To account for fatigue the study cuts the degrees of freedom in half. Is there a better way to account for fatigue?
Only done on participants in English language – does it still apply to other languages on Facebook? Also take into consideration differences between men and woman, differences in English language (AAVE, code/language switching, etc).
“It performs with surprising accuracy, modeling tie strength to 10-point resolution and correctly classifying friends as strong or weak ties more than 85% of the time”

The accuracy rate is not 100%, does that change our earlier discussion: is 85% accuracy sufficient for benefiting the platform/users?
Weight Assignment

- INTIMACY: 32.8%
  - last comm
  - num friends
  - intimacy words

- INTENSITY: 19.7%
  - wall words
  - outbound posts
  - thread depth

- DURATION: 16.5%
  - first comm

- SOCIAL DISTANCE: 13.8%
  - educational diff
  - political diff
  - occupational diff

- SERVICES: 7.9%
  - links shared
  - apps shared

- EMO. SUPPORT: 4.8%
  - inbox positive words
  - wall positive words

- STRUCTURAL: 4.5%
  - mutual strength
  - interest overlap
  - common groups

TIE STRENGTH
Adj. $R^2 = 0.534$
MAE = 0.0994
<table>
<thead>
<tr>
<th>Top 15 Predictive Variables</th>
<th>β</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days since last communication</td>
<td>-0.76</td>
<td>453</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Days since first communication</td>
<td>0.755</td>
<td>7.55</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Intimacy × Structural</td>
<td>0.4</td>
<td>12.37</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Wall words exchanged</td>
<td>0.299</td>
<td>11.51</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Mean strength of mutual friends</td>
<td>0.257</td>
<td>188.2</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Educational difference</td>
<td>-0.22</td>
<td>29.72</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Structural × Structural</td>
<td>0.195</td>
<td>12.41</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Reciprocal Serv. × Reciprocal Serv.</td>
<td>-0.19</td>
<td>14.4</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Participant-initiated wall posts</td>
<td>0.146</td>
<td>119.7</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Inbox thread depth</td>
<td>-0.14</td>
<td>1.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Participant’s number of friends</td>
<td>-0.14</td>
<td>30.34</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Inbox positive emotion words</td>
<td>0.135</td>
<td>3.64</td>
<td>0.05</td>
</tr>
<tr>
<td>Social Distance × Structural</td>
<td>0.13</td>
<td>34</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Participant’s number of apps</td>
<td>-0.12</td>
<td>2.32</td>
<td>0.12</td>
</tr>
<tr>
<td>Wall intimacy words</td>
<td>0.111</td>
<td>18.15</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Discussion & Applications
“The more messages friends exchange on a single topic, the lower their tie strength.”

– How? Discuss with partner to come up with possible reasons for this phenomenon.
How much do these results matter given that much of the accuracy rate relies on your private inbox messages which ideally are end-to-end encrypted?

- p-values: Inbox positive emotions word: 0.05, dates of communication: p < 0.001 (not encrypted) Intimacy x Structural: p < 0.001

-> Facebook didn’t add end-to-end encryption until 2016

How do we minimize personal data needed while increasing accuracy to tie prediction to expand networks and opportunities? What variables are unethical to track from the study if applied globally?
We are told that Lois knows “the actors, the writers, the doctors, the lawyers, the park lovers, the politicians, the railroad buffs, and the flea-market aficionados” and many more.

“Poverty is not deprivation. It is isolation”

Six Degrees of Lois Weisberg (Malcolm Gladwell)

Her “social power” from being part of so many different communities

-> “The Strength of Weak Ties”

We are told that Lois knows “the actors, the writers, the doctors, the lawyers, the park lovers, the politicians, the railroad buffs, and the flea-market aficionados” and many more.
Inherent Polarization from strong ties

“Strong ties are the people you really trust, people whose social circles tightly overlap with your own. Often, they are also the people most like you.”

Applied to large social networks

If Facebook decides to maximize its algorithms for strong ties, it will inherently polarize groups of people. This is because, as Granovetter discovered, people with strong ties tend to be similar to each other. As a result, social networks full of strong ties would result in less diverse communities so Facebook must work towards a balance of weak and strong ties.