Empirical Methods of Data Science

PROFESSOR MICHELLE LEVINE

WEEK 5: ETHICS & BIG DATA

2/20/19

Defining ethics in the age of big data

Does our definition of ethics need to be modified?

Problem with consent and big data

IRB approval and big data

Today

Current state of affairs

- Why and what do people share?
- Potential consequences of social media

Social Media

Who in here (show of hands) is on:

- Facebook
- Twitter
- Instagram
- Others?

Social Media Statistics

79% of internet users use Facebook

32% use Instagram

24% use Twitter

Facebook:

- 1.23 billion daily users (December 2016)
- 4.75 billion pieces of content is shared daily (May 2013)

Instagram:

- 313 million monthly active users
- Average around 6,000 tweets per second

From Pew Research & FB internal statistics, Instagram internal statistics

Twitter

First tweet on March 2006

May 2009, the billionth tweet was posted

In 2015, 302 million monthly active users

Average around 6,000 tweets per second

Social Sharing

Why do people share?

How?

When?

With whom?

Why Do We Share?

NYT Study

6 sharing personality types:

- Altruists
- Careerists
- Hipsters
- Boomerangs
- Connectors
- Selectives

5 main reasons for sharing online:

- add to lives of others
- define themselves
- connect with others
- feel fulfilled
- promote a cause

My Research

Levine, Hiatt & Shapiro

One goal: what gets shared, with whom, when, through what channels, & social consequences and rewards of sharing such content

Method:

- 21 day diary study
- Daily diary entry
- Compensated with a Kindle
- 63 participants

Questions to Consider

What exactly gets shared?

- Within longform sharing, what stories are more likely to be shared and why?
- More generally, what content do people tend to share and why?

Findings

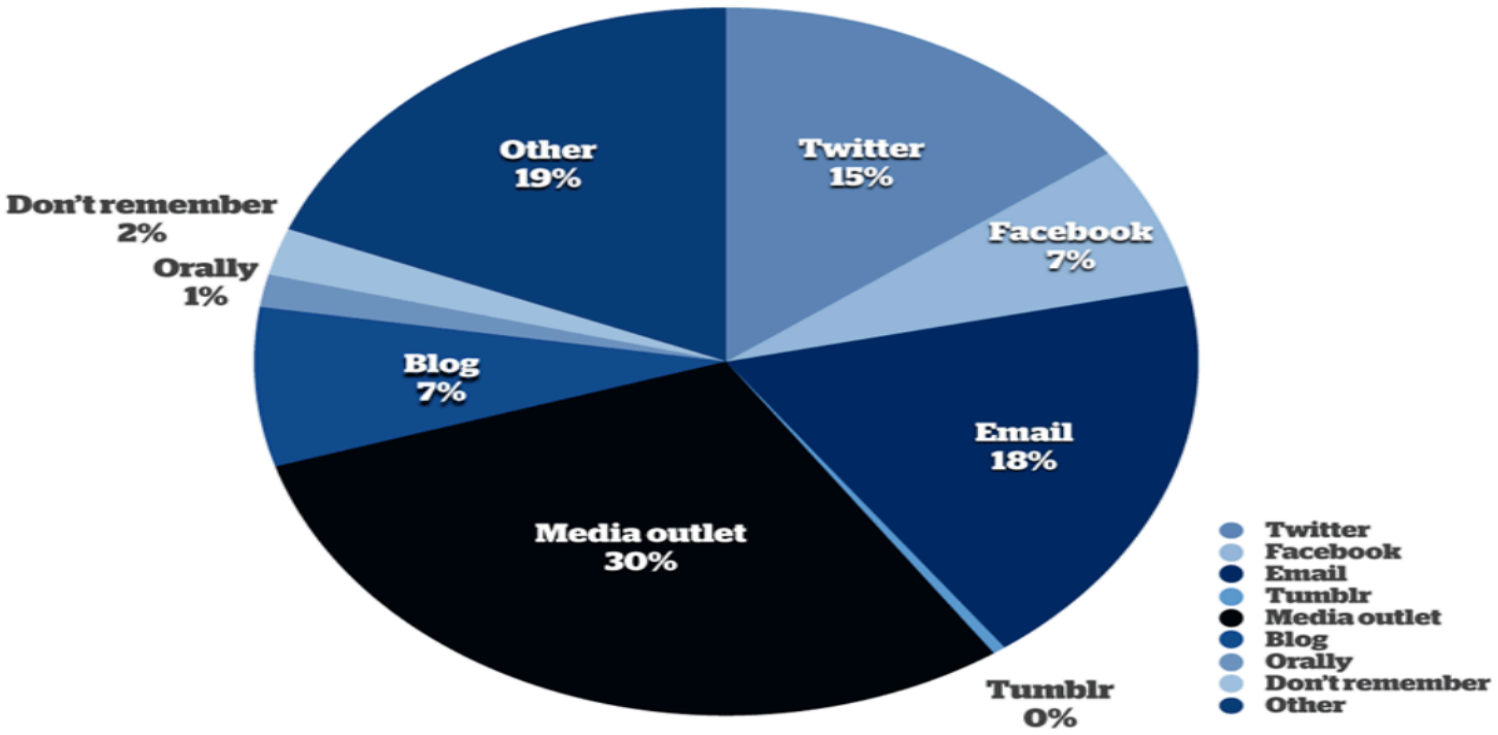
1349 stories read

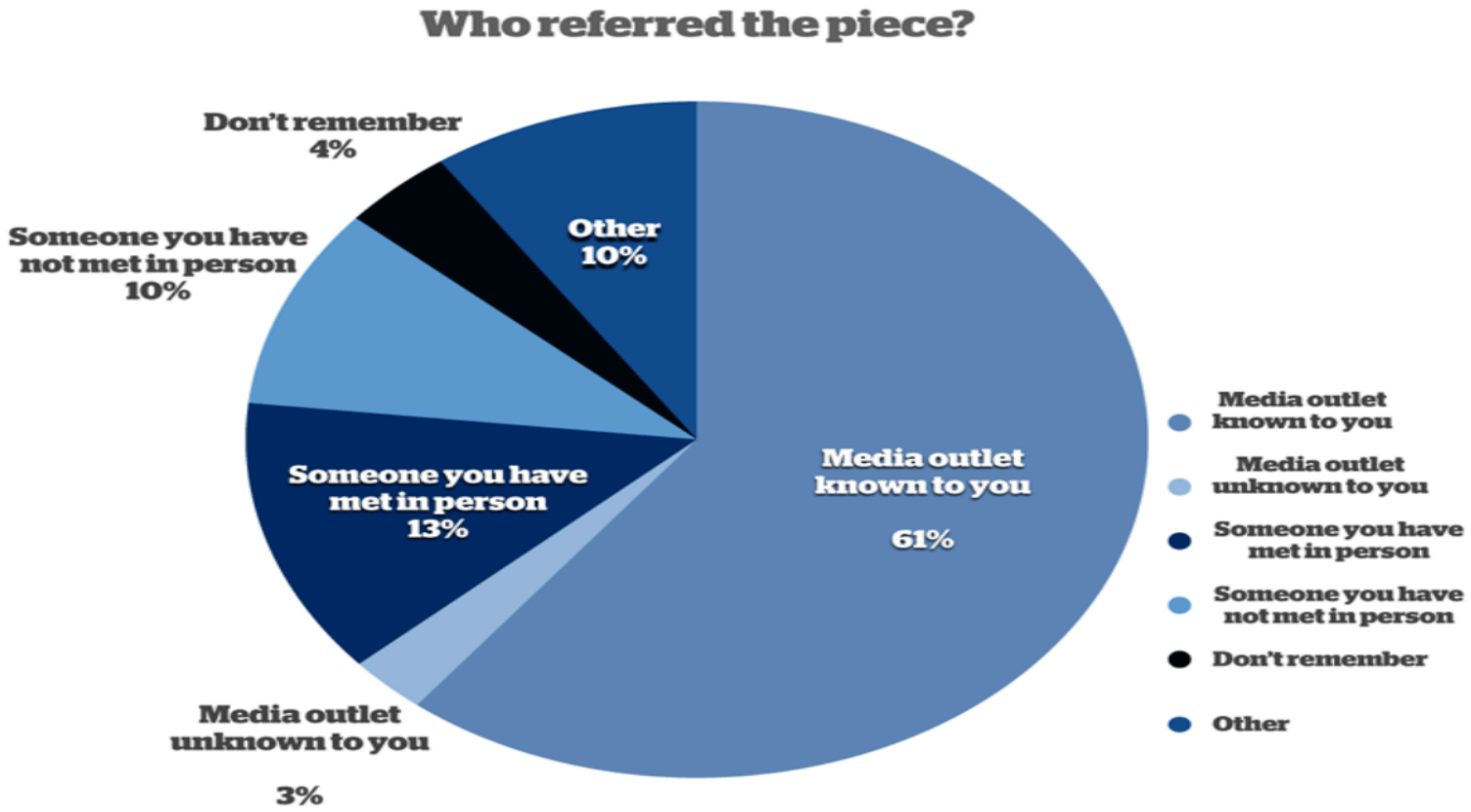
- 35% shared

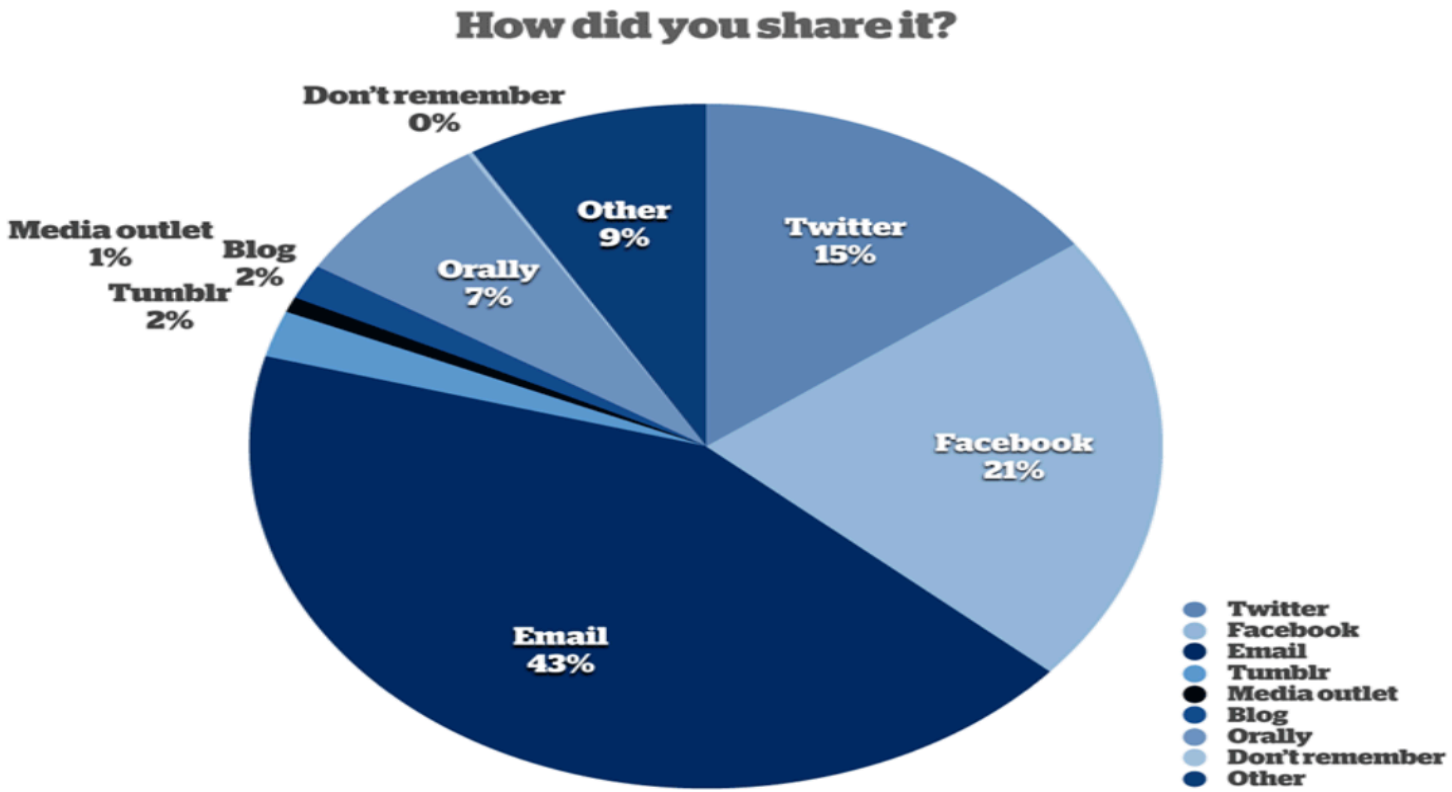
Half the stories read over the weekend; half read during the week

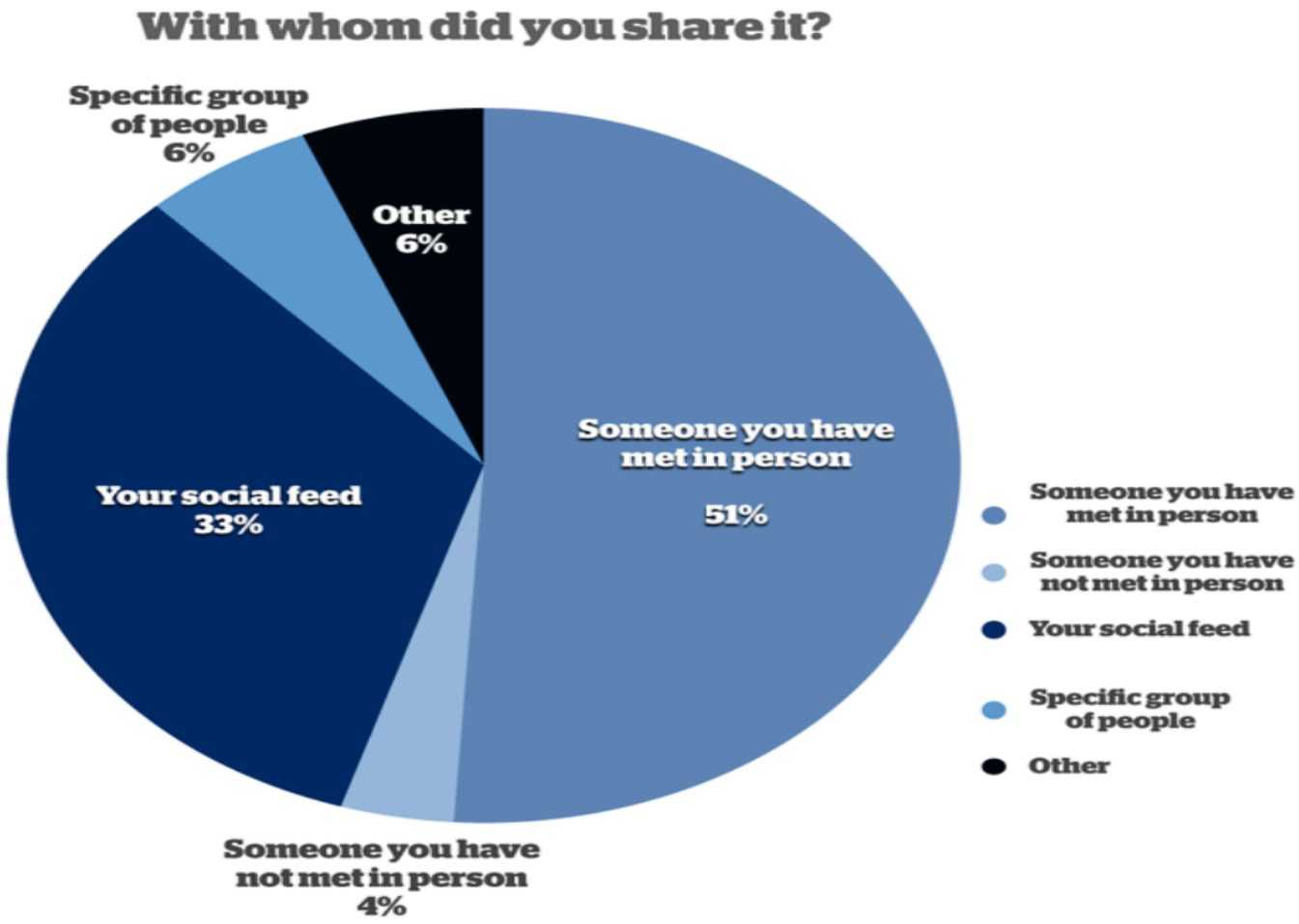
- much less likely to share a story they read over the weekend

Where did you get the piece?




Who referred the piece?
Don't remember
4%
Other
10%
Someone you have
not met in person
10%
Someone you have
met in person
13%
Media outlet
known to you
61%
Media outlet
unknown to you
3%
Media outlet known to you
Media outlet unknown to you
Someone you have met in person
Someone you have not met in person
Don't remember
Other


How did you share it?
Don't remember 0%
Other 9%
Twitter 15%
Media outlet 1%
Blog 2%
Orally 7%
Tumblr 2%
Facebook 21%
Email 43%
Twitter
Facebook
Email
Tumblr
Media outlet
Blog
Orally
Don't remember
Other


With whom did you share it?
Specific group of people 6%
Other 6%
Someone you have met in person 51%
Your social feed 33%
Someone you have not met in person 4%
Someone you have met in person
Someone you have not met in person
Your social feed
Specific group of people
Other

What can spread through social networks?

Key research by sociologist Nicholas Christakis and political scientist James Fowler

- co-authors of *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*
- Analyzes contagion with the Framingham Heart Study social network

Framingham Heart Study

Longitudinal cross-generation study

1) 1948: Original Cohort
- 5209 people from Framingham, Massachusetts
2) 1971: Offspring Cohort
- 5124 children and spouses of original cohort
3) 1994: Omni Cohort
- 508 people to increase ethnic diversity
4) 2002: Third Generation Cohort
- 4095 children of offspring cohort

The Network

FHS Network consists of:

- 12,067 people & 53,228 observed social ties
- Egos (focal individuals in the network) = Offspring Network
- Alters (people connected to the egos) come from across the network

2604 unique observed friendships

3 types:

1) Ego perceived friend
2) Alter perceived friend
3) Mutual friend

Happiness & Social Networks

Fowler & Christakis (2008)

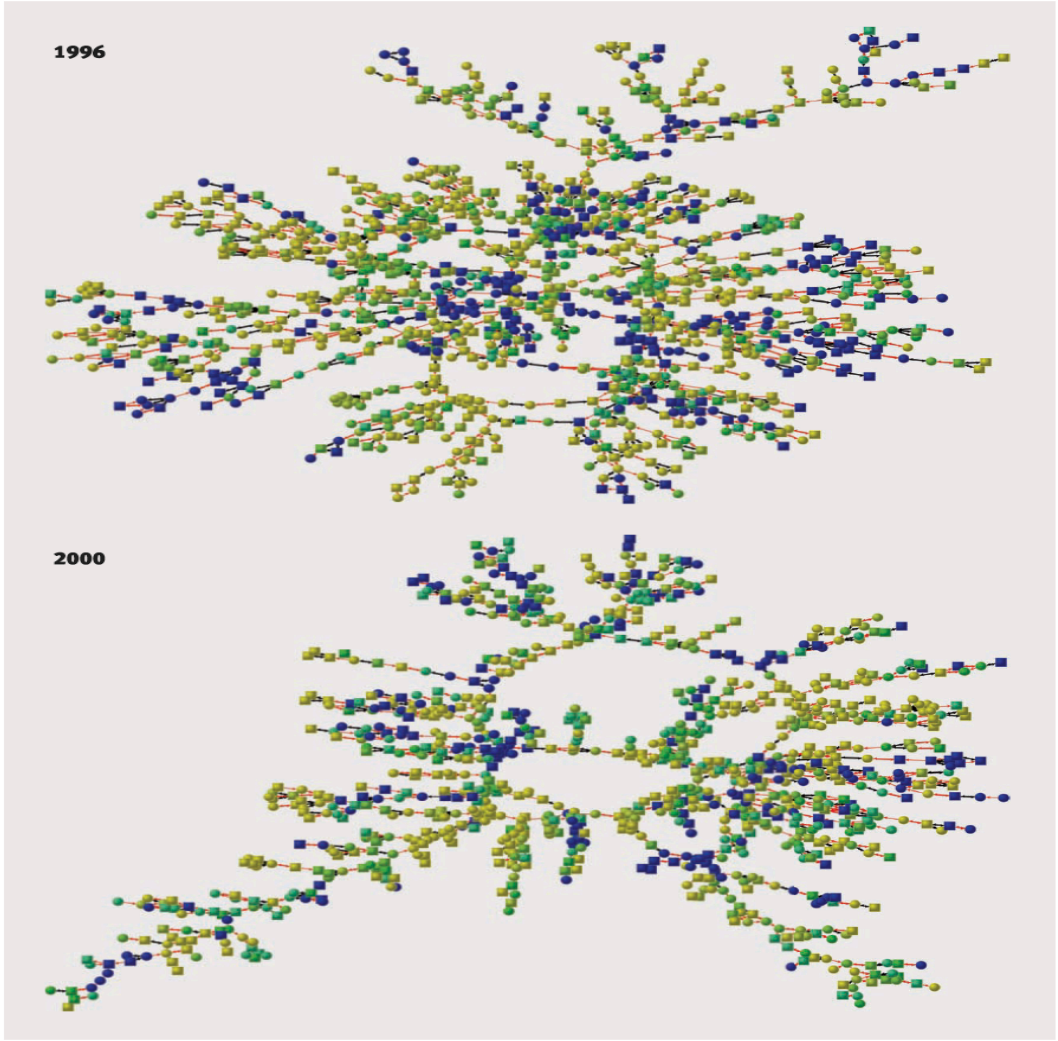
Can happiness spread across networks?

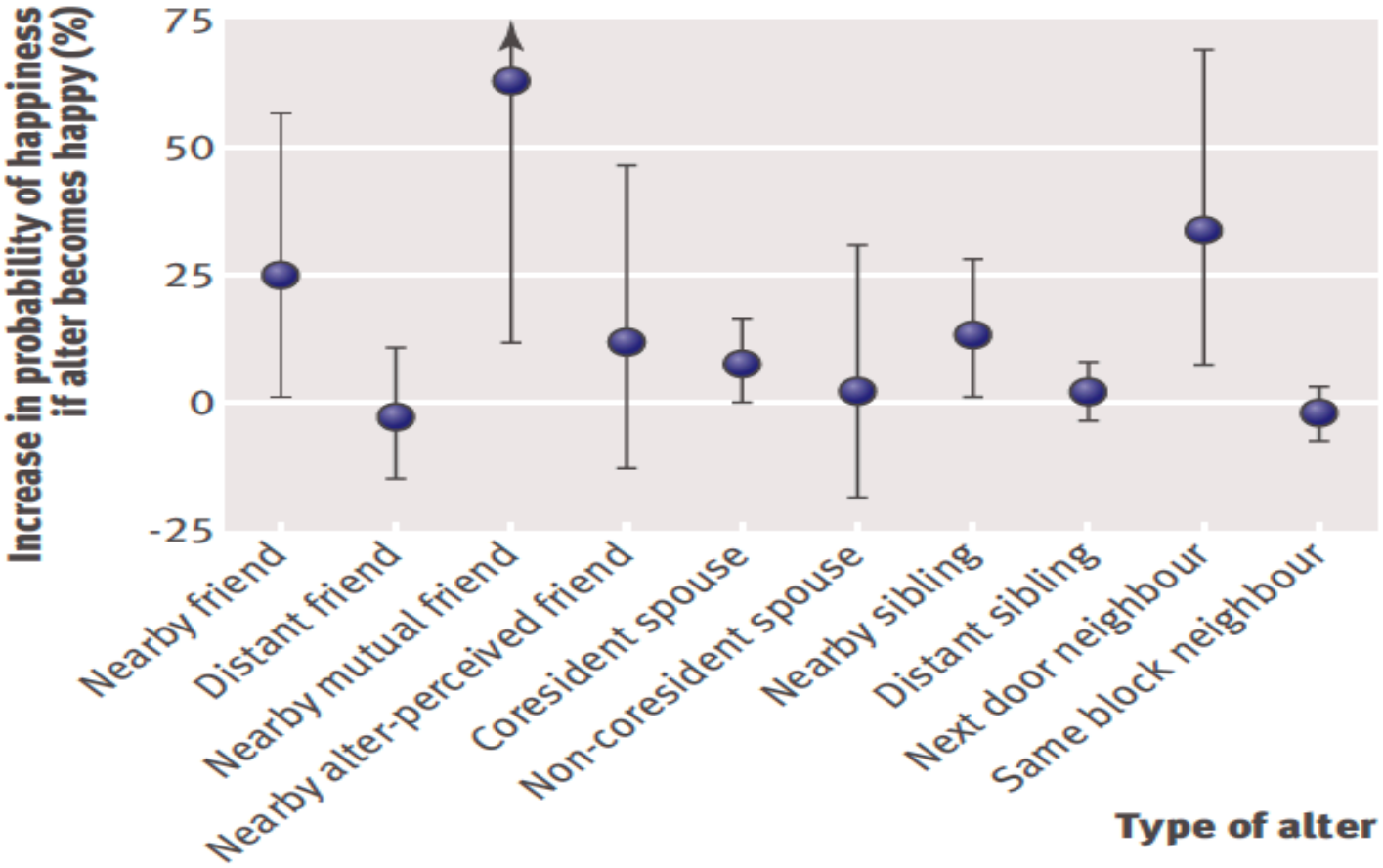
- If yes, is it only across direct relationships or also indirect ones?
- Any geographic or temporal limitations?

Defined happiness as positive emotions

- Perfect score on 4 items from Center For Epidemiological Studies Depression Scale (CES-D)

Looked at happiness and changes in happiness over time


1996
2000


Increase in probability of happiness if alter becomes happy (%)
75
50
25
0
-25
Nearby friend
Distant friend
Nearby mutual friend
Nearby alter-perceived friend
Coresident spouse
Non-coresident spouse
Nearby sibling
Distant sibling
Next door neighbour
Same block neighbour
Type of alter

Depression & Social Networks

Rosenquist, Fowler & Christakis (2011)

Can depression spread across networks?

- Measured depression from 20 items from Center For Epidemiological Studies Depression Scale (CES-D)

Similar findings:

- Depression can travel across networks
- Influenced by degree of separation
- Now on peripheral of network

Any change could be because of:

- **Induction:** emotion/state of one person induces it in another connected person
- **Homophily:** people choose to be connected to similar people
- **Confounding:** a common external situation that elicits a particular emotion/state

Emotional Contagion Across FB

Coviello, Sohn, Kramer, Marlow, Franceschetti, Christakis & Fowler (2014)

Are the same effects found across online social networks?

Method

Instrumental variables regression

Instrument = rainfall

Data collected across 1180 days (January 2009 - March 2012)

Focused on Facebook status updates/posts of English-speakers in 100 most popular US cities

Aggregated people by city and day

Used LIWC to determine positive vs negative emotion

- Measured emotional state of a user in 2 ways:
  - Positive rate: fraction of posts with positive emotion
  - Negative rate: fraction of posts with negative emotion

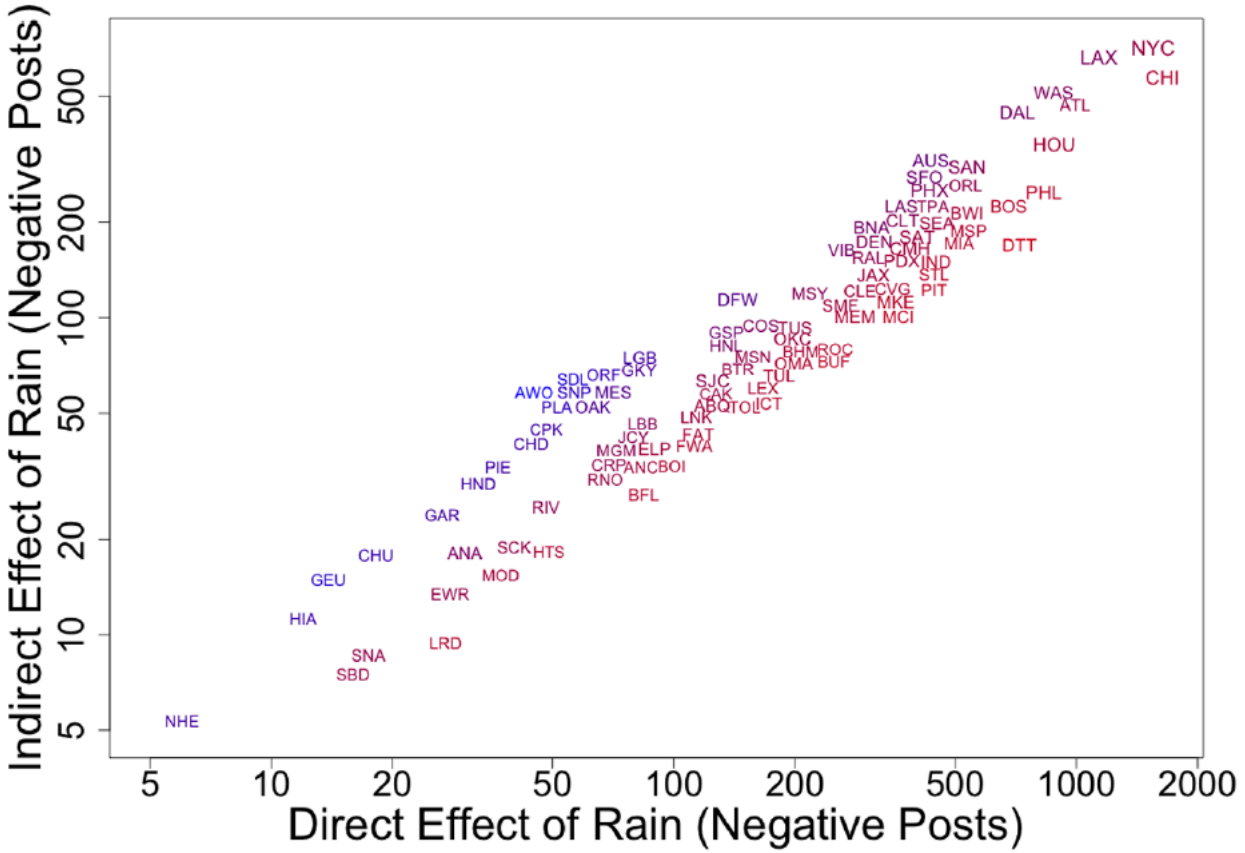
Results

Rainfall directly affects emotionality of status messages

- of people in the city with rain
- AND of friends in other cities not experiencing rain

For every 1 person directly effected, 1-2 other people will be indirectly effected

What Happens When it Rains in NYC?



Results

Rainfall directly affects emotionality of status messages

- of people in the city with rain
- AND of friends in other cities not experiencing rain

For every 1 person directly effected, 1-2 other people will be indirectly effected

- **Positive emotional posts → 1.75**
- **Negative emotional posts → 1.29**

Emotional Contagion Across FB
(with experimental manipulation)

Kramer, Guillory & Hancock (2014)

Does a decrease in exposure to emotional posts (positive; negative) affect the emotionality of people's own posts?

Hypotheses:

- If emotional contagion exists then
  - people exposed to fewer positive posts should post less positive posts
  - people exposed to fewer negative posts should post less negative posts
- If cross-emotional contagion exists then
  - people exposed to fewer positive posts should post more negative posts
  - people exposed to fewer negative posts should post more positive posts

Method

689,003 randomly selected Facebook users

- Had to post at least one status update during the data collection week (January 11-18, 2012)

Manipulated news feeds (2 conditions):

- Reduced exposure to friends' positive posts OR
- Reduced exposure to friends' negative posts

Dependent variables:

- % of positive words
- % of negative words

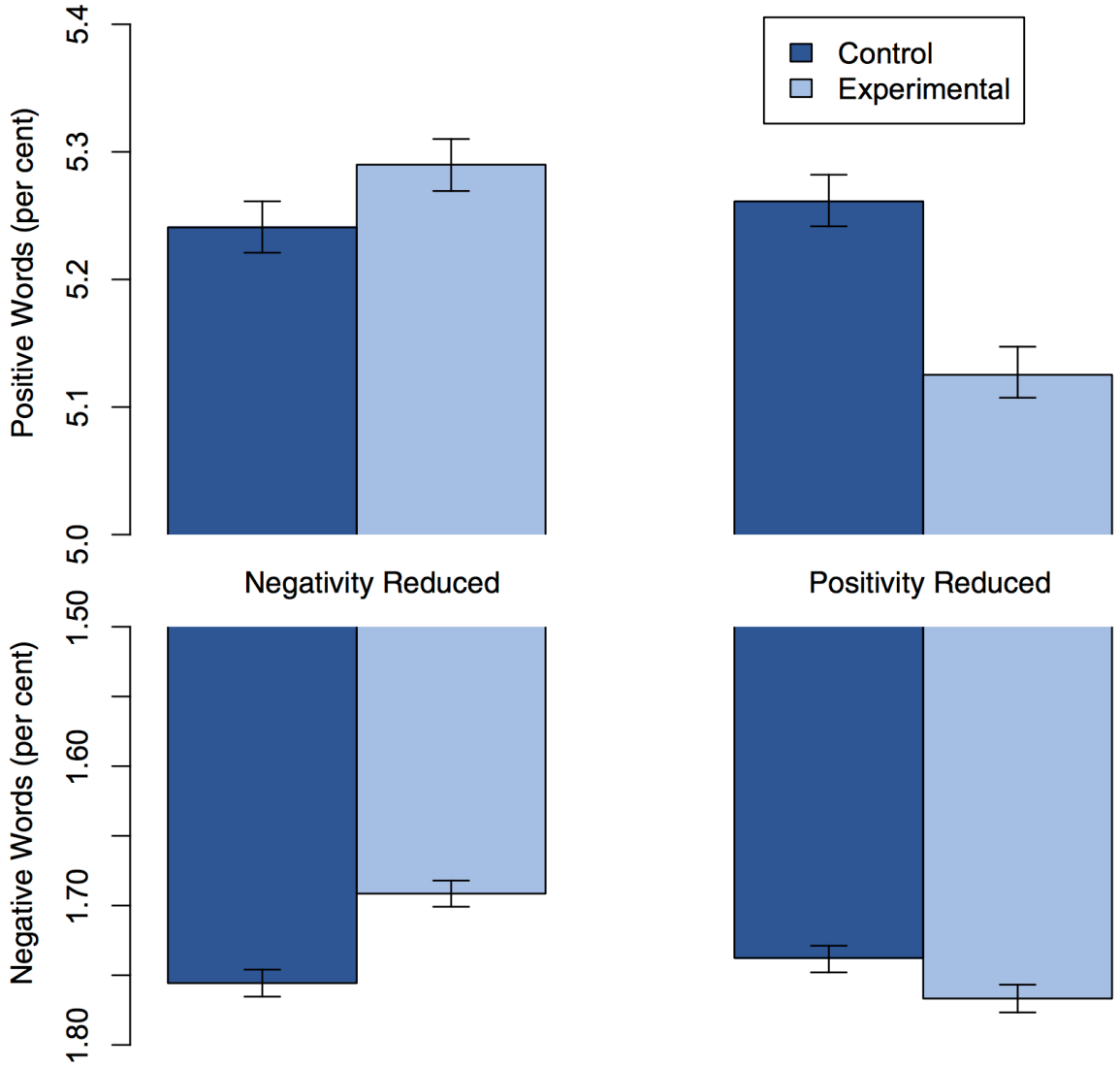
How were posts categorized?

- LIWC
- Negative posts: contains at least one negative word
- Positive posts: contains at least one positive word

Over 3 million posts

122 million words

- 4 million positive words (3.6%)
- 1.8 million negative words (1.6%)


Control
Experimental
Positive Words (per cent)
5.0
5.1
5.2
5.3
5.4
Negativity Reduced
Positivity Reduced
Negative Words (per cent)
1.50
1.60
1.70
1.80

”Overhearing” Emotions

Findings in contrast to previous findings in psycholinguistic research

- E.g., Schober & Clark (1989)

FB study on emotional contagion

If you were the Cornell IRB board, would you have approved the study? Why or why not?

Do you consider this research ethical?

Why might see this research as ethical vs unethical?