Data and Society
Lecture 10: Data and Community Communication

4/20/18
Announcements 4/20

• Office hours today 1-2

• Check what you think your grades are (attendance, op-ed, and presentation scores) with Fran during office hours. You are responsible for being sure that these are accurate.

• No discussion article for next week.

• Next week is the last day of class. No final exam.
<table>
<thead>
<tr>
<th>Wednesday Section</th>
<th>Friday lecture</th>
<th>First Half of Class</th>
<th>Second Half of Class</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 17: NO class</td>
<td>January 19</td>
<td>L1: CLASS INTRO AND LOGISTICS</td>
<td>Presentation Model / Op-Ed Instructions</td>
<td>Op-Ed instructions</td>
</tr>
<tr>
<td>January 24: NO class</td>
<td>January 26</td>
<td>L2: BIG DATA 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>January 31: NO class</td>
<td>February 2</td>
<td>L3: BIG DATA 2 -- IoT</td>
<td>4 Presentations</td>
<td></td>
</tr>
<tr>
<td>February 7: NO class</td>
<td>February 9</td>
<td>L4: DATA AND SCIENCE</td>
<td>4 Presentations</td>
<td></td>
</tr>
<tr>
<td>February 21: 5</td>
<td>February 23</td>
<td>L5: DATA AND HEALTH / LESLIE McIntosh GUEST SPEAKER</td>
<td>4 Presentations</td>
<td></td>
</tr>
<tr>
<td>Presentations</td>
<td></td>
<td>L6: DATA STEWARDSHIP AND PRESERVATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>February 28: 5</td>
<td>March 2</td>
<td>CLASS CANCELED DUE TO SNOW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentations</td>
<td></td>
<td>L7: INFRASTRUCTURE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 21: NO class</td>
<td>March 23</td>
<td>NO CLASS / PAPER PREPARATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>April 4: NO class</td>
<td>April 6</td>
<td>L8: DATA RIGHTS, POLICY, REGULATION</td>
<td>4 Presentations</td>
<td></td>
</tr>
<tr>
<td>April 11: 4 Presentations</td>
<td>April 13</td>
<td>L9: DATA AND ETHICS</td>
<td>4 Presentations</td>
<td></td>
</tr>
<tr>
<td>April 18: 4 Presentations</td>
<td>April 20</td>
<td>L10: DATA AND COMMUNICATION</td>
<td>4 Presentations</td>
<td></td>
</tr>
<tr>
<td>April 25: NO class</td>
<td>April 27</td>
<td>L11: DATA FUTURES</td>
<td>4 Presentations</td>
<td></td>
</tr>
</tbody>
</table>
Data and Community Communication
Data and Community Communication

• Data technologies have revolutionized how we communicate to individuals, groups, organizations for public, professional and personal purposes.

• Are we communicating more? Better? To different people?

• Is our technology-driven communication producing better outcomes?
Technology and communication [Wikipedia]

Timeline of Communication tools

3000 BC: papyrus
3000 BC: communication drums, horn
2400 BC: couriers, first postal systems
490 BC: heliograph
59 BC: first newspaper
100 AD: paper
1000 AD: pens
1400 AD: first European printing presses
1500 AD: pencils
16th century: maritime flags
1790 AD: semaphore lines
1800s: typewriter
1838: telegraph
1848: telephone
1896: radio
1897: computer
19th century: Signal lamps
1927: television
1960s: computers and text editors
1969: computer networking
1983: Internet
1980s: Mobile; 1G networks
Early 1990s: Mobile; 1G networks
Late 1990s: Mobile; 2G & 2.5 networks
2000s: Mobile; 3G networks
2012: Mobile; 4G networks

Communication tools from the BC era.
Only the postal system and newspapers are still in use today.

Communication tools during the 1900 years AD.
Almost all of these tools are still in use today. They have been improved over time, and merged with later tools and technologies.

The early days of computer, networking, and internet.

Image by Alaakh (Own work) [CC BY-SA 3.0 (https://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons:
https://upload.wikimedia.org/wikipedia/commons/7/7d/Timeline_of_communication_tools%2C_2014_update..jpg
Delivering the News – Then and Now
Communication via the Internet: News and Fake news

• “Fake news is a type of yellow journalism or propaganda that consists of deliberate misinformation or hoaxes .... Fake news is written and published with the intent to mislead in order to damage an agency, entity, or person, and/or gain financially or politically ...” [Wikipedia]
News is part of a larger ecosystem

- 3 key elements
  - Different types of content that are being created and shared
  - Motivations of content creators
  - Ways content is disseminated

Information based on FirstDraft article by Claire Wardle
Social media and fake news

- Social media platforms particularly conducive to fake news:
  - Cost of entering the market and producing content extremely small → increases the **relative profitability of small-scale, short-term strategies adopted by fake news producers** and reduces the relative importance of building a long-term reputation for quality
  - Format of social media (**small amounts of information**) makes it hard to judge what is true
  - Facebook friend networks are generally **ideologically separated** → people more likely to read and share articles with others of the same point of view and less likely to receive evidence about the true state of the world that would counter their ideology

Information from https://web.stanford.edu/~gentzkow/research/fakenews.pdf
Fake and “real” news

(graph from https://web.stanford.edu/~gentzkow/research/fakenews.pdf)

Note: This figure presents the share of traffic from different sources for the top 690 US news websites and for 65 fake news websites. “Other links” means impressions that were referred from sources other than search engines and social media. “Direct browsing” means impressions that did not have a referral source. Sites are weighted by number of monthly visits. Data are from Alexa.
Facebook business leveraged for fake news

- Facebook offers advertisers a way to target ads to specific audiences, using hundreds of parameters. Service is fee-based (remember *you* are the product ...)

- Facebook business model:
  - Advertisers set up a FB page and ad account.
  - Advertisers create a post and pay FB to boost this post as a “sponsored” post.
  - Posts can be tailored to specific audiences.
  - The more an advertiser pays, the bigger and more targeted an audience it can reach.
  - Ad accounts can be tied to multiple pages, but there does not have to be a visible link between them.
How Fake news can happen

- https://www.wired.com/video/2017/02/heres-how-fake-news-works-and-how-the-internet-can-stop-it/ (3 minutes)

Facebook and Cambridge Analytica

• Cambridge Analytica gained access to private information (e.g. identities, friends, likes) on more than 50M Facebook users.
  – Data gathered from FB users that responded to a personality survey (~270K) and their friends. Survey developed by researcher [Aleksandr Kogan] who worked with Cambridge Analytica

• Firm offered tools that could identify the personalities of American voters and attempt to influence their behavior

• Facebook routinely works with researchers but prohibits data to be sold or transferred to data brokers or “monetization-related service”

• Cambridge analytica acquired the data and claims that it has been deleted.

• Facebook said it removed Kogan’s app when it learned that the data had been turned over to Cambridge Analytica and is doing a digital forensics investigation “to determine the accuracy of the claims that the Facebook data ... still exists”

• Incident has spiked prevalent trust issues about how Facebook treats their users.
  – Stock has gone down
  – Developing investigations from Massachusetts, U.S. Congress and U.K.
Work in progress: What Facebook has tried to do about fake news

• **Detection:**
  – Customers can report news as false, FB also monitors sharing links to myth-busting sites such as Snopes

• **Verification:**
  – Attempt to technically classify misinformation
  – Engage third-party verification

• **Awareness**
  – Labels on stories that have been flagged as false by third parties or communities (discontinued) ...

• Improve *quality / quantity of related articles* under links in News Feed

• **Ad policies** that discourage ad farms

• **Crowdsourcing trustworthiness:** surveys that ask users if they are familiar with a news source and if so, whether they trust the source.

• **Academic partnership:** Working with non-profits on research and analysis to better identify and reduce fake news.
Fake news: individual responsibility

• What can we do?
  – Don’t passively accept information without double-checking
  – Don’t share a post, image or video before you’ve verified it
  – Readers now have the responsibility for independently checking what we see online.

Infographic by IFLA - [http://www.ifla.org/publications/node/11174](http://www.ifla.org/publications/node/11174), CC BY 4.0, [https://commons.wikimedia.org/w/index.php?curid=57084301](https://commons.wikimedia.org/w/index.php?curid=57084301); information from FirstDraft article
Personal relationships: On-line dating

- On-line dating increasingly prevalent
  - 15% of Americans have used on-line dating or mobile apps
  - 41% of Americans know someone who uses online dating; 29% know someone who has met a spouse or long-term partner via online dating

- 5% of Americans who are in a marriage or committed relationship say they met their significant other online.

From http://www.pewinternet.org/2016/02/11/15-percent-of-american-adults-have-used-online-dating-sites-or-mobile-dating-apps/
Paths to relationships

- Personal contacts in "real world" settings
- Family arrangement
- Matchmakers
- Personal ads
- On-line sites and apps
Increasing prevalence of couples who met online

Fig. 1. Percentage of Americans who met their partners online as a function of year met. Data are from Wave 1 of the national data set, How Couples Meet and Stay Together (collected in 2009), for 2,535 heterosexual respondents who were married or in a romantic relationship (cohabiting, dating). The percentages for same-sex couples were generally higher, particularly in the most recent period (61%). Adapted from Table 7 in Rosenfeld and Thomas (2010) with permission from the authors.

From http://journals.sagepub.com/stoken/rbtfl/cK9EB6/4zQ0AM/full
On-line dating vs. conventional off-line dating

• 3 differences:
  
  – **Access** – exposure to large pool of romantic partners
  
  – **Communication** – opportunity to utilize computer-mediated communication before meeting
  
  – **Matching** – use of an algorithm to select potential partners

Matching algorithms for on-line dating sites

• Compatibility matching offered by on-line sites can be in-depth
• Matching algorithms differ in
  – Use of data *self-collected from users*
  – Other data about users
  – What *variables considered* and how they are weighted
  – Method to determine how desirable is the individual as a relationship partner
  – Site *goal* (e.g. long-term compatibility, initial attraction)
Matching algorithms largely proprietary (competitive advantage)

- **E-Harmony General approach**
  - Algorithms developed based on
    - interviews conducted with married couples to see what variables might be relevant to predicting success in long-term relationships.
    - Psychometric tests performed on large sample of couples
  - Current algorithm based on survey with 13 sections and 300 items.
    - Survey captures 29 dimensions that allegedly predict long-term relationship success.
    - Traits are distinguished as core traits (unlikely to change in adulthood) and vital attributes (which may change based on learning and experience)
  - Behavioral data on site (how long spent on site, how long to respond to an email, how people contacted respond, etc.) used by the company to predict how users will respond to proposed matches
  - E-Harmony’s matchmaking software gathers **600 data points** for each user
Similarity and Complementarity Matching

• **PerfectMatch approach**
  
  – Clients traits analysed as whether they are similar or complementary
  
  – PerfectMatch seeks to leverage both similarity and complementary traits, claiming that couples are more compatible when they are similar on romantic impulsivity, personal energy, outlook, and predictability and when they are different on flexibility, decision-making style, emotionality, and self-nurturing.
Anthropological perspective

• **Chemistry approach**
  – Developed by Anthropologist Fisher
  – Clients typed according to 4 personality types (Explorer, Builder, Negotiator, Director), 3 of which are linked to 2 sex hormones (testosterone and estrogen) and 2 neurotransmitters (dopamine and serotonin).
  – Survey contains around 60 items (including a question about the length of one’s index finger compared to the ring finger, allegedly indicating testosterone level)
  – Users are shown visual representations and asked for interpretation as well as answering questions
### 3 approaches to matching

#### Table 2. The Variables Measured at the Three Major Scientific Matching Sites and Matching Principles Used

<table>
<thead>
<tr>
<th>Matching site</th>
<th>Domain</th>
<th>Construct</th>
<th>Variables assessed</th>
<th>Matching principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>eHarmony</td>
<td>Core traits: “defining aspects of who you are that remain largely unchanged throughout your adult life”</td>
<td>Emotional temperament</td>
<td>Self-concept, emotional status, energy, passion</td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Social style</td>
<td>Character, kindness, dominance, sociability, autonomy, adaptability</td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cognitive mode</td>
<td>Intellect, curiosity, humor, artistic passion</td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td>Vital Attributes: “based on learning experience, and are more likely to change based on life events and decisions you make as an adult”</td>
<td>Physicality</td>
<td>Energy, passion, vitality and security, industry, appearance</td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relationship skills</td>
<td>Communication style, emotion management, conflict resolution</td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td>Values and beliefs</td>
<td></td>
<td>Spirituality, family goals, traditionalism, ambition, altruism</td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td>Key experiences</td>
<td></td>
<td>Family background, family status, education</td>
<td>Similarity</td>
</tr>
<tr>
<td>PerfectMatch</td>
<td></td>
<td></td>
<td>Romantic impulsivity, Personal energy, Outlook, Predictability</td>
<td>Similarity or complementarity</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Sex hormones</td>
<td></td>
<td>Director personality type</td>
<td>Complementarity</td>
</tr>
<tr>
<td></td>
<td>Neurotransmitters</td>
<td>Testosterone</td>
<td>Negotiator personality type</td>
<td>Complementarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Estrogen</td>
<td>Explorer personality type</td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dopamine</td>
<td>Builder personality type</td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serotonin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*From www.eharmony.com/why/dimensions

From [http://journals.sagepub.com/stoken/rbtfl/cK9EB6/4zQ0AM/full](http://journals.sagepub.com/stoken/rbtfl/cK9EB6/4zQ0AM/full)
Does it work?

• **Positives:**
  
  – 80% of Americans who have used online dating agree that online dating is a good way to meet people.
  
  – 62% agree that online dating allows people to find a better match, because they can get to know a lot more people.
  
  – 61% agree that online dating is easier and more efficient than other ways of meeting people.

• **Negatives:**
  
  – 45% of online dating users agree that online dating is more dangerous than other ways of meeting people.
  
  – 31% agree that online dating keeps people from settling down, because they always have options for people to date.
  
  – 16% agree with the statement “people who use online dating sites are desperate.”
Data-mining into a relationship

“How I hacked on-line dating”, Amy Webb
(17:23 min)

https://www.ted.com/talks/amy_webb_how_i_hacked_online_dating/up-next
Lecture 10 Sources

- Fake news: Firstdraft, https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79
- Facebook and Cambridge Analytica: https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html
- Online Dating: A critical analysis from the perspective of psychological science, Sage Journals, http://journals.sagepub.com/stoken/rbtfl/cK9EB6/4zQ0AM/full
Break
Discussion article for April 20

Presentation articles for April 27


- “The trouble with quitting Facebook is that we like Facebook,” NY Times, https://fivethirtyeight.com/features/the-trouble-with-leaving-facebook-is-that-we-like-facebook/ [Tae P]
Presentation articles for Today


- “With this DNA dating app, you swab, then swipe for love”, Wired, https://www.wired.com/story/with-this-dna-dating-app-you-swab-then-swipe-for-love/ [John L]

- “A combination of personality traits might make you more addicted to social networks,” ScienceDaily, https://www.sciencedaily.com/releases/2018/03/180312084911.htm [Jie C]