Social Research with Digital Traces

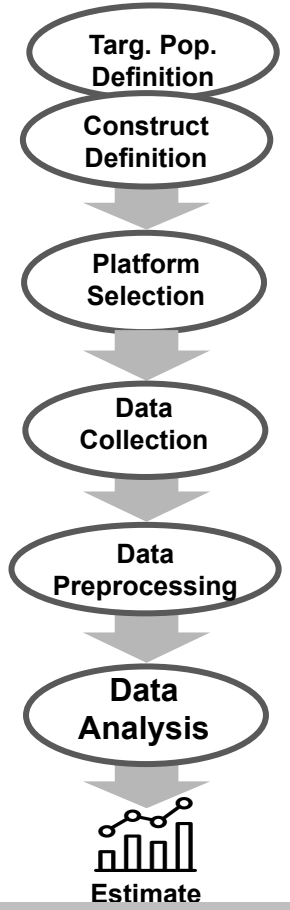
Indira Sen

University of Konstanz
Measurement and Representation Biases (MRB) in Digital Trace Data-based Studies

Agenda

- A hypothetical study that uses digital trace data
- Reading and reviewing papers
- Mini paper reading session

A typical research pipeline with digital traces



"TED-On: A Total Error Framework for Digital Traces of Human Behavior on Online Platforms" Sen et al., 2021, Public Opinion Quarterly and https://arxiv.org/pdf/1907.08228

Use case: Detecting the flu with digital traces

Google Flu (trends)

nature

Vol 457|19 February 2009|doi:10.1038/nature07634

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

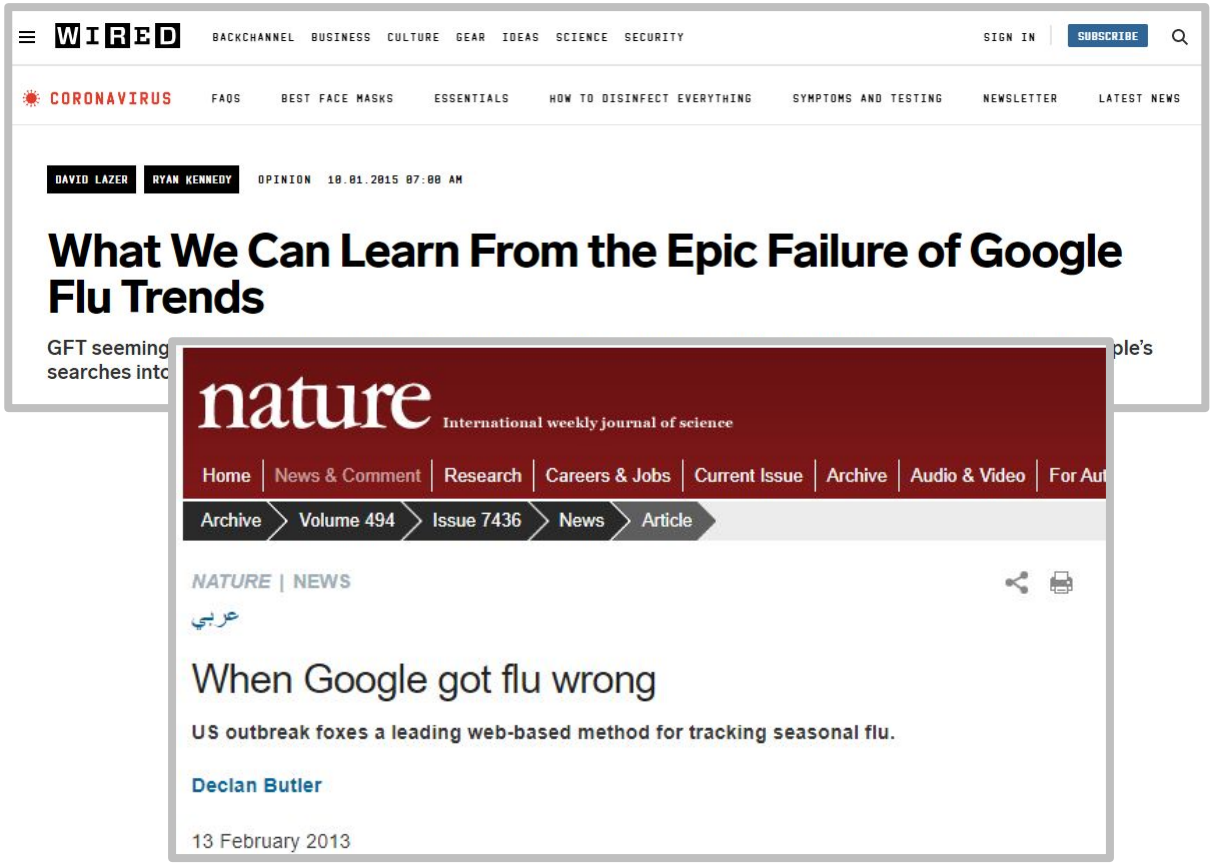
What proportion of US-Americans have the flu?

What is the approval rating of A. Merkel?

Are anti-immigration sentiments on the rise?

Ginsberg, Jeremy, et al. "Detecting influenza epidemics using search engine query data." *Nature* 457.7232 (2009): 1012-1014.

The fail



"In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. Nature reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (1, 2)."

Example study

- How would a researcher study influenza prevalence in a national population using digital traces?

Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis

Construct Definition

Platform Selection

Data Preprocessing

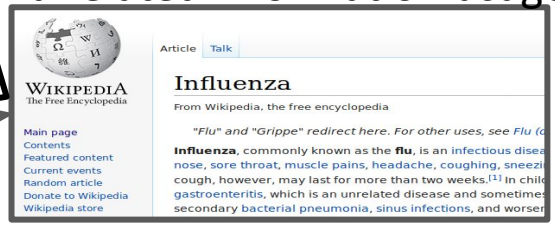
Data Analysis

SIGNALS

search queries related to flu

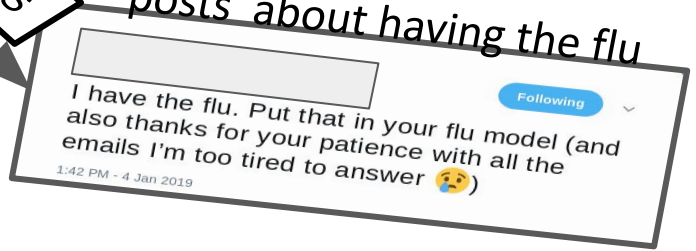
SIGNALS

flu related information usage



SIGNALS

posts about having the flu



Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis

Platform Selection

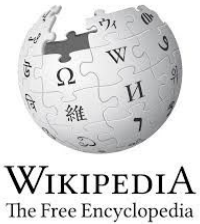
Google

Preprocessing

Data Analysis

Construct Definition

Platform Selection



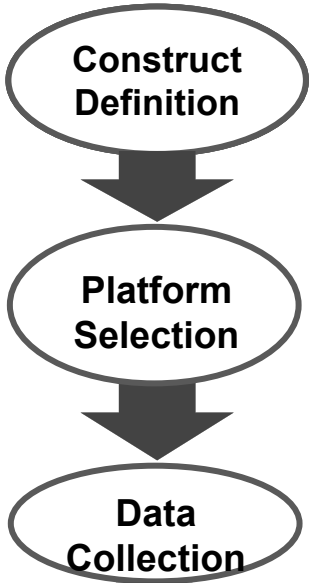
Recommendations for search queries

Only aggregate queries

Session length for readers

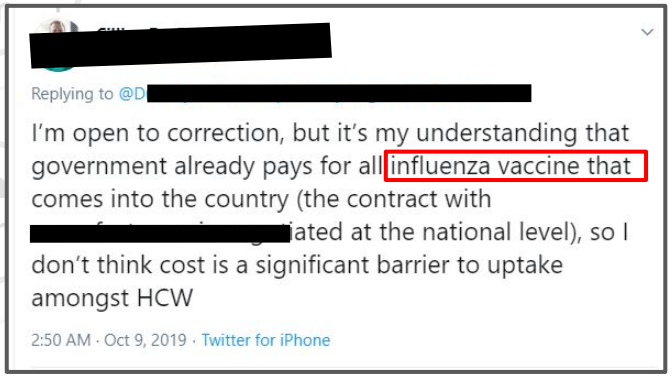
280-character limit

Trending tweets


Construct Definition
Platform Selection
Data Collection

Construct Definition

Platform Selection

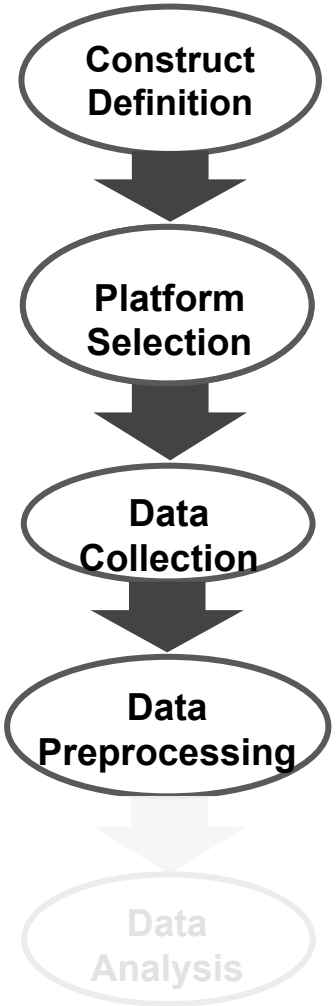


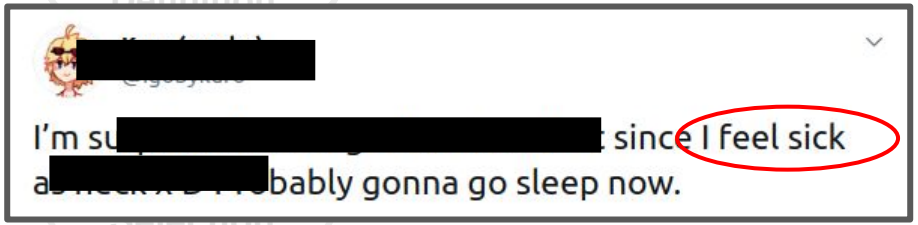
Data Collection

Data Preprocessing

Data Analysis

Tweets with keywords related to Influenza: Avian influenza, Influenza Virus B, Centers for Disease Control and Prevention, Influenza Virus C, Common Cold, **Vaccine,** Flu(the Band), Influenza


Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis



Augment signals with syntactic relations indicating the speaker being afflicted by the flu

Construct Definition

Data Collection

Data Preprocessing

Data Analysis

Construct Definition

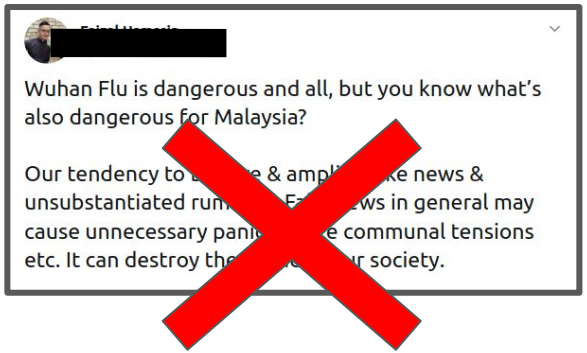
Platform Selection

Data Collection

Data Preprocessing

Data Analysis

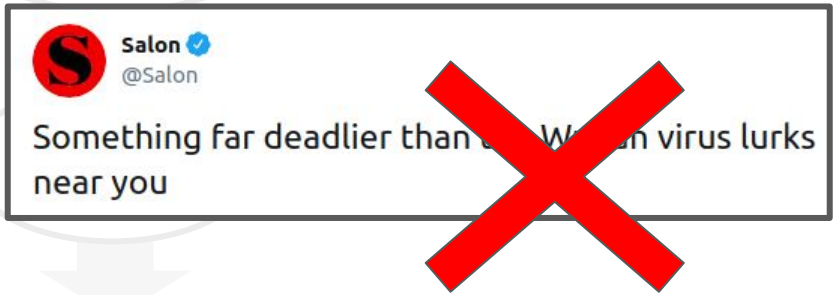
Remove non ‘first-person’ tweets



Augment entities with demographic information

Data Preprocessing

Remove bots or organizations

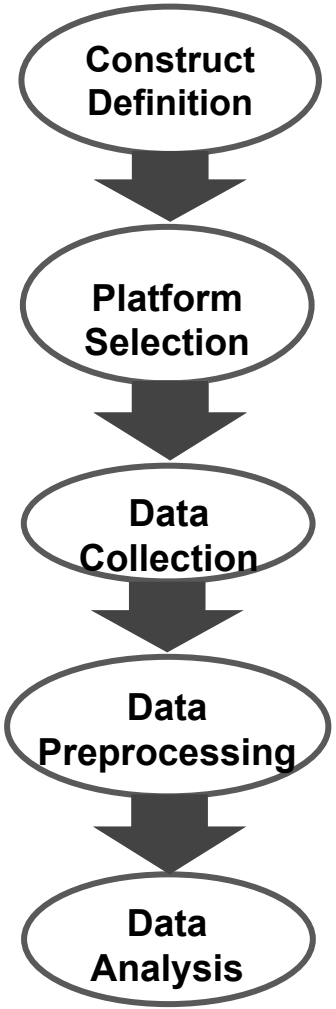


Construct Definition

Data Collection

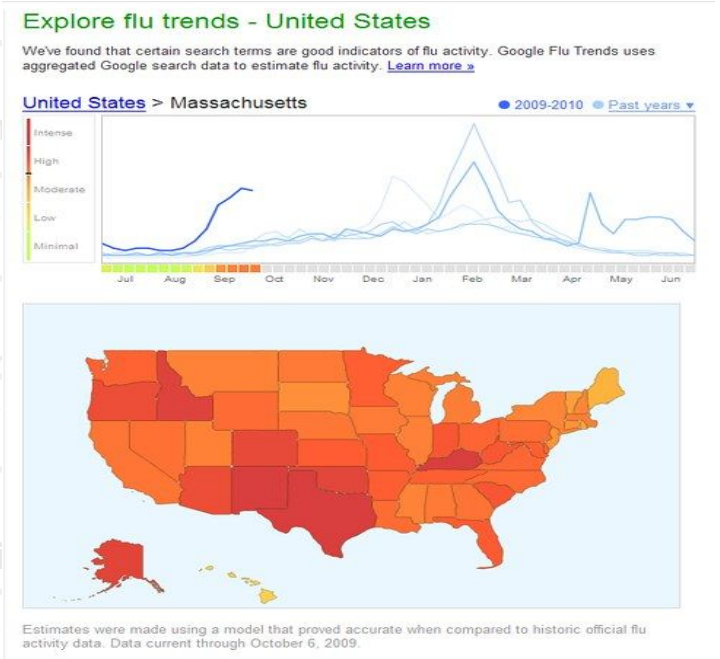
Data Preprocessing

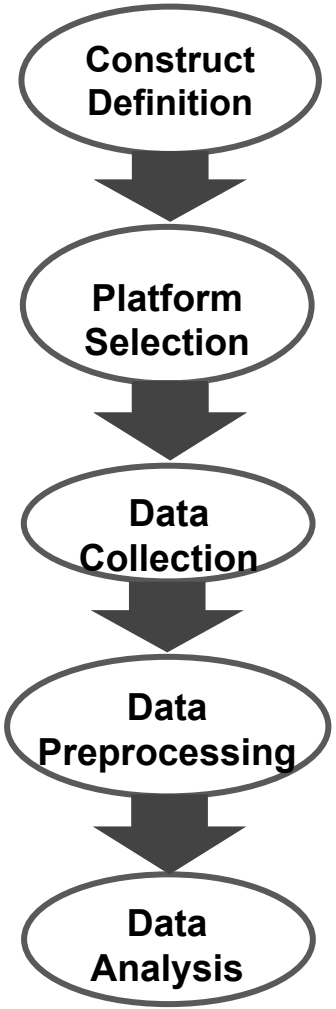
Data Analysis


Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis

Develop machine learning model to measure rates of flu based on tweeting activity

Use demographic data (say, location) to correct for coverage errors using post-stratification




Construct Definition
Platform Selection
Data Collection
Data Preprocessing
Data Analysis

How to read and review papers?

1. Keshav, Srinivasan. "How to read a paper." ACM SIGCOMM Computer Communication Review 37.3 (2007): 83-84.
2. Pain, Elisabeth “How to review a paper”

The typical “anatomy” of a paper:

- Title and authors
- Abstract/summary
- Introduction
- Materials and Methods
- Results
- Discussion
- Acknowledgements
- References
- Figures/Tables

“How to read a scientific paper” by Kelly Hogan

Reviewing papers

Two purposes:

1. Quality control: publish the paper or not?
2. Constructive criticism: how to improve the paper?

Aim: be as efficient as possible with the first, to leave most time for the second.

How to [read, present, review] a research paper by Perdita Stevens:
https://www.cl.cam.ac.uk/~ey204/teaching/ACS/R212_2015_2016/aid/stevens.pdf

Why learn to review papers in this course?

Obvious purpose: because if you stay in academia you will, eventually, have to write reviews.

More importantly here, though:

1. because reviewing a paper is one way to ensure you've really read it – trying to write about a paper, in whatever way, often reveals what one hasn't understood;

2. because understanding how the reviewing process works should help you to do better research and research writing.

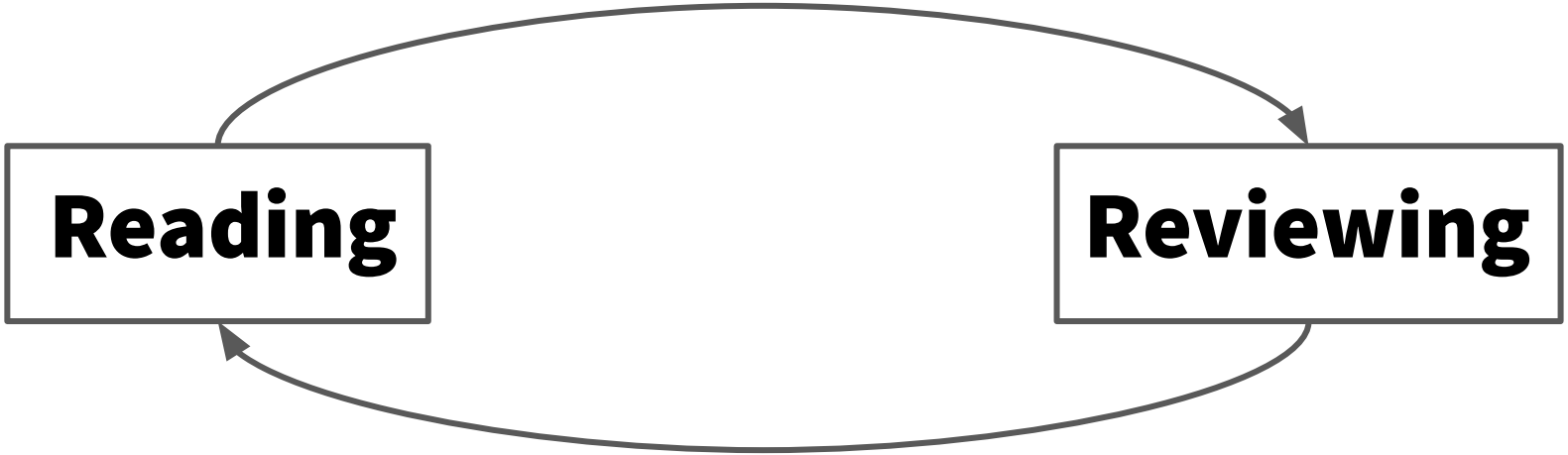
How to [read, present, review] a research paper by Perdita Stevens:
https://www.cl.cam.ac.uk/~ey204/teaching/ACS/R212_2015_2016/aid/stevens.pdf

How to review papers

Many different ways, this is what I do:

1. Read the abstract
2. Skim the introduction, results, and conclusions
3. Then read the entire paper more carefully
4. Make notes as you go: comments and questions
5. Then write the first review draft:
    a. Summary of the paper; focus on it's contributions
    b. Strengths [does it address an important gap? Who does it 'help'?]
    c. Weaknesses [not limitations, but e.g., does the paper *do* what it says it *does*?]

Bullet points are all you need for your first draft!


Reading
Reviewing

Let’s read and review a paper together

"Narco" Emotions: Affect and Desensitization in Social Media during the Mexican Drug War

Munmun De Choudhury[1] **Andrés Monroy-Hernández**[1] **Gloria Mark**[2]
[1] Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
{munmund, andresmh}@microsoft.com

[2] Department of Informatics
University of California, Irvine
Irvine, CA 92697 USA
gmark@uci.edu

ABSTRACT
Social media platforms have emerged as prominent information sharing ecosystems in the context of a variety of recent crises, ranging from mass emergencies, to wars and political conflicts. We study affective responses in social media and how they might indicate desensitization to violence experienced in communities embroiled in an armed conflict. Specifically, we examine three established affect measures: negative affect, activation, and dominance as observed on Twitter in relation to a number of statistics on protracted violence in four major cities afflicted by the Mexican Drug War. During a two year period (Aug 2010-Dec 2012), while violence was on the rise in these regions, our findings show a decline in negative emotional expression as well as a rise in emotional arousal and dominance in Twitter posts: aspects known to be psychological markers of desensitization. We discuss the implications of our work for behavioral health, facilitating rehabilitation efforts in communities enmeshed in an acute and persistent urban

as it can lead to cognitive performance decline, attentional impairment [20], and is a stressor of the onset of PTSD (post-traumatic stress disorder), an anxiety disorder associated with harmful physiological outcomes [30].

The Mexican Drug War is an example of the type of armed conflict that has exposed people to persistent acts of violence. Since the war started in, many Mexican cities have seen a rapid increase of shootings and homicides that, on occasions, affect innocent civilians. Furthermore, the conflict has triggered an increase of criminal activities such as extortions, and kidnappings affecting the general population [24]. This generalized violence in some Mexican cities, coupled with constrained information reporting on news media, have contributed to the emergence of citizen alert networks using platforms like Twitter and Facebook to inform and collectively grieve, critique, and express frustration about the violence in the streets [25].

Previous research in crisis informatics has demonstrated the role of social media as a lens to understand how society

What are we going to do now?

1. Read the paper based on the strategy we saw (the abstract in detail, skimming the introduction, conclusion, and optionally, the results)

2. Read the introduction and conclusion in more detail

3. Skim the data, methods, and results if time permits

4. Write at least 1-2 points under strengths and weaknesses

Next Time

We will read: Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. "Social data: Biases, methodological pitfalls, and ethical boundaries." Frontiers in big data 2 (2019): 13.
https://www.frontiersin.org/articles/10.3389/fdata.2019.00013/full

We will start the discussions based on the different roles:
https://colinraffel.com/blog/role-playing-seminar.html

Acknowledgements and More Detailed Resources on Reading and Reviewing Papers

"How to read a scientific paper" by Kelly Hogan

A four-step method based on: Ann McNeal, School of Natural Science, Hampshire College, Amherst MA:
http://hampshire.edu/~apmNS/design/RESOURCES/HOW_READ.html

How to [read, present, review] a research paper by Perdita Stevens:
https://www.cl.cam.ac.uk/~ey204/teaching/ACS/R212_2015_2016/aid/stevens.pdf