Data Science, Social Media and the Internet: A Researcher's View

Week 6

Recap Last Week

US Government Data Plans and others

The use of social media by cities and law enforcement agencies

Weekly Outline

Ethics in social media

- Challenges and Opportunities for researchers and users

Use of Social Media for medical research

Google data mining

Mind Maps!

Seek informed consent for all types of online research?

Data posted in open spaces vs. data protected with passwords

Inability to verbally acknowledge what participants consent to

Participants' Anonymity

- What if participants want to be known?
- Using email can compromise participants' confidentiality
- Online platform security
- Avoid judgement from others
- Risk of damaging reputation

http://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf

Undue Harm

Relevant support to participants

Avoid psychological harm

Participants could be more honest and open

Questions about legitimacy of researchers and organizations they are affiliated with

Risks of being identified online

Use of Social Media for Medical Research

Understand how information flows through social networks and how members influence each other

Dr. Farnan- Difficult to verify user identity and intent

- Individuals may behave differently online than they would in real life
- Follow-up is harder within the medical community

https://sciencelife.uchospitals.edu/2014/10/21/a-tempting-source-of-data-social-media-is-uncharted-ethical-territory-for-medical-research/

Researchers' View of Ethics in Social Media

a "lack of standardized practices" regarding research ethics

Social media as "deliberate public information"

- To some Twitter "entirely public", yet author may not be ok being a research subject

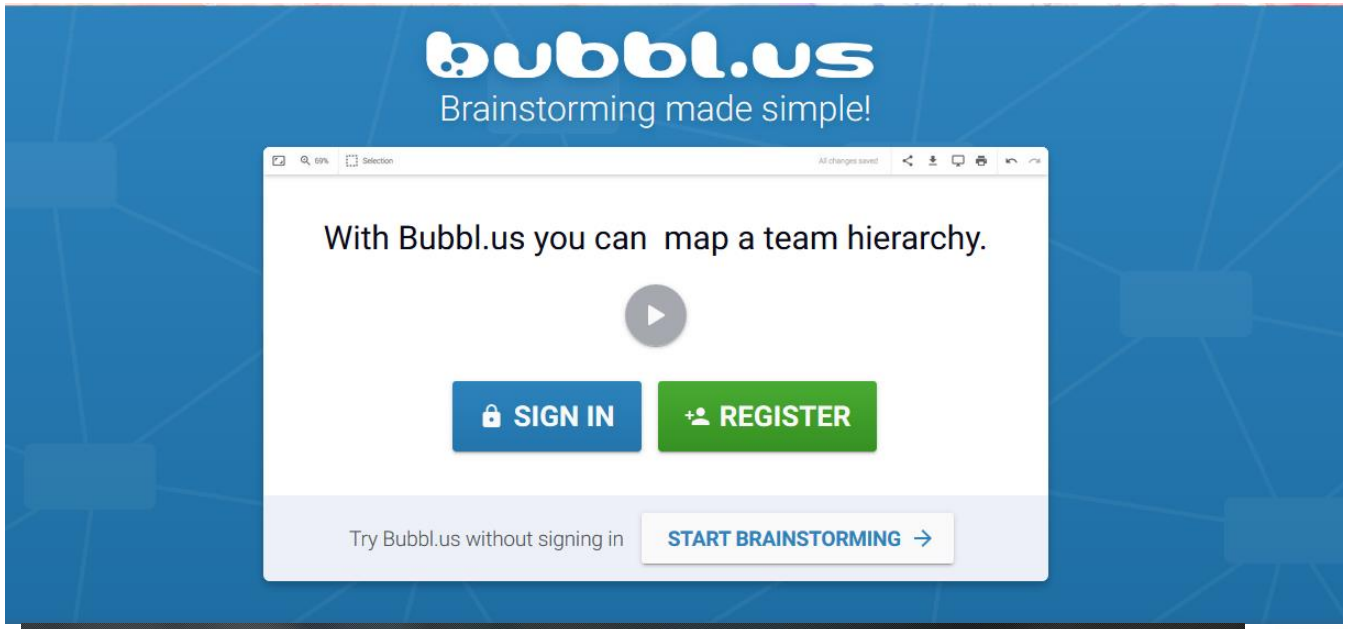
Anonymity is difficult to achieve

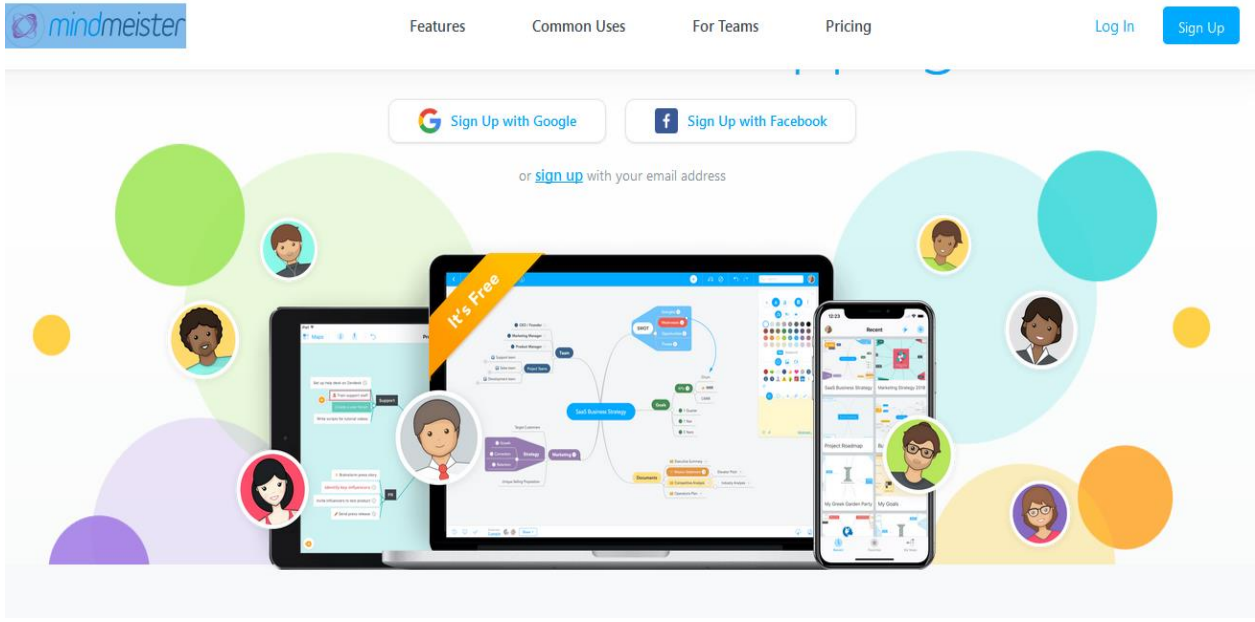
- Researchers "quote[ed] tweets that had been retweeted beyond a certain threshold"
- Some user groups considered particularly vulnerable, i.e. religious groups

Flexible approach to ethical issues

https://katrinweller.files.wordpress.com/2012/08/hiddendataethics_wellerkinder-kurlanda_ir15-preprint.pdf

Mind Maps!





https://www.analyticsvidhya.com/blog/2016/06/12-free-mind-mapping-tools-data-scientist-enhance-structured-thinking/

Community Commentary on Social Credit System

Social Credit System Commentary

Common Good from Social Media Data Assignment

This week, create either a presentation or an APA-formatted paper and turn it in to the week 6 dropbox.

Imagine you are a data science researcher working for a non-profit. Your first task is to use social media data to contribute to the common good. Choose a specific common good topic (e.g. mental health, income inequality, human rights, worker's rights, a specific healthcare topic, socioeconomic injustices, etc.) and choose a social media data source for mining the data. Collect some data from your source, and at least create a wordcloud or barchart of word counts showing the top words from the data (if using text). The files "Reddit_Scraping_Starter.ipynb" and "Reddit_Sentiment_Analysis_Instructions.docx" show how to create a wordcloud from a subreddit. You could also collect data from X, Facebook, or other social media sites instead. **At a minumum create the wordcloud and include it in your presentation or paper. To go above and beyond, apply other NLP and text analytics techniques, such as topic modeling and sentiment analysis.** Note that many people consider wordclouds uninformative and bad practice, so you should strive to create a bar chart of top word counts or other visualizations instead.

As part of your paper or presentation, also create and include a **mindmap.** The mindmap should include aspects of the common good with social media project. For example, it could include the common good issue in the center with social media sites as the next layer, and ethics/privacy issues beyond that. The next layer out could be examples of these ethics/privacy issues. You can also use the mindmaps in the examples as inspiration.

In your work, include an overall introduction, an introduction of your data set, the purpose (what problem are you trying to solve?), ethics and privacy issues, the status of the common good issue you selected, and a summary. See the weekly reading list for links to videos and mindmap resources.

Social Media Common Good Project Rubric

Social Media Common Good Project Rubric

Criteria	Points
Overall quality	2
Introduction and conclusion/summary	2
Description of common good issue, mindmap, and wordcloud	2
Ethics and privacy issues included and described	2
Status of common good Issue	2
Total Points	**10**