COMM XXX: Social Data APIs and Analysis

Ross Dahlke

Summer 2022

E-mail: rdahlke@stanford.edu
Office Hours: TBD
Office: TBD

Web: TBD
Class Hours: TBD
Class Room: TBD

Course Description

This class will combine theoretical and technical discussions and labs to give students the ability to conduct their own research with digital social data. Every week we will have discussions on the structure, content, and userbase of various social platforms and what that means for the data that is generated from them. We will also learn how to access data from these platforms via their web APIs or other databases in addition to how to analyze these data. Finally, students will be able to answer their own unique research questions. By the end of the course, students should be well positioned to conduct their own original research.

Students should have a working understanding of the statistical programming language R. This class is not an introductory computing course. We will be covering how to access web APIs and analyze the data from these sources, but students are expected to already know basics of computing such as data types, data cleaning, and basic analysis.

In addition, this course will primarily be using the `tidyverse` for programming. If you are unfamiliar with the `tidyverse` please review Hadley Wickham's R for Data Science, which can be found for free online.

This course will be primarily taught using R. Students are highly encouraged to use R. It is my opinion that R is the best choice for individuals who wish to conduct computational social science research or pursue data science in industry. However, students are able to use Python but must be aware that they will be responsible for finding the necessary resources in Python to complete their assignments and project on their own.

Course Objectives

1. Better understand the uniqueness of digital social platforms

2. Know how to access social data via web APIs

3. Learn how to analyze social data retrieved from the web

4. Conduct and present original research

Required Readings

There will not be any books that need to be bought for this course. Instead, students will have assigned readings each week. Please read through the technical documentation and examples of academic research for each week prior to the first class of the week.

Grading Policy

- **20%** of your grade will come from Platform Reviews

- **30%** of your grade will come from the weekly lab assignments

- **10%** of your grade will come from your project proposal

- **10%** of your grade will come from your project presentation

- **20%** of your grade will come from your project paper

Attendance Policy

Attendance is not mandatory for this course. However, Platform Reviews will only be accepted at the beginning of class on Mondays. The labs are intended to give students enough time to to complete most, if not all, of the lab assignment in lab. During lab you will have access to the instructor for questions, and you are encouraged to work with other students during the labs.

E-mail Policy

Students can expect an response within 24 hours during the week. Please start the subject line of your email with "COMM XXX: [your name]".

Academic Dishonesty Policy

Students are expected to abide by the Stanford Honor Code (https://communitystandards.stanford.edu/student-conduct-process/honor-code-and-fundamental-standard) .

Accessibility

Students in need of accommodation due to a disability should file a request with the Office of Accessible Education (OAE). OAE will prepare an Accommodation letter for the instructor. OAE is located at 563 Salvatierra Walk. https://oae.stanford.edu/

Class Schedule

Week 1: Introduction

- About this course
- R & RStudio
- Rmarkdown

Week 2: Twitter

Technical documentation

- `rtweet` https://github.com/ropensci/rtweet
- Twitter API https://developer.twitter.com/en/docs

Examples of research

- Barberá P, Rivero G. Understanding the Political Representativeness of Twitter Users. *Social Science Computer Review*. 2015;33(6):712-729. doi:10.1177/0894439314558836
- Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, 23(1), 76-91. doi:10.1093/pan/mpu011

Week 3: Tik tok

Technical documentation

- `tiktokr` https://github.com/benjaminguinaudeau/tiktokr
- pushshift.io Tik Tok data https://files.pushshift.io/tiktok/

Examples of research

- Zhu, C.; Xu, X.; Zhang, W.; Chen, J.; Evans, R. How Health Communication via Tik Tok Makes a Difference: A Content Analysis of Tik Tok Accounts Run by Chinese Provincial Health Committees. *Int. J. Environ. Res. Public Health* 2020, 17, 192. https://www.mdpi.com/1660-4601/17/1/192#cite
- Moran, P. Social Media: A Pandemic of Misinformation. *The American Journal of Medicine* 2020, 133, 11. https://www.amjmed.com/article/S0002-9343(20)30519-2/abstract

Week 4: Reddit

Technical documentation

- `RedditExtractoR` https://github.com/ivan-rivera/RedditExtractoR
- Reddit API https://www.reddit.com/dev/api/
- Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit Dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1), 830-839. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/7347

Examples of research

- Shen, J. D.; Rudzic, F. Detecting anxiety on Reddit. *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology* 2017. The Association for Computational Linguistics. https://www.aclweb.org/anthology/W17-3107.pdf

- Cody Buntain and Jennifer Golbeck. 2014. Identifying social roles in reddit using network structure. *In Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)*. Association for Computing Machinery, New York, NY, USA, 615–620. DOI: https://doi.org/10.1145/2567948.2579231

Week 5: YouTube

Technical documentation

- `tuber` https://github.com/soodoku/tuber
- YouTube API https://developers.google.com/youtube/v3

Examples of research

- Wattenhofer, Mirjam; Wattenhofer, R.; Zhu, Z. The YouTube Social Network. 2012. *In Sixth International AAAI Conference on Weblogs and Social MEdia (ICWSM 2012)*. https://research.google/pubs/pub37738/
- X. Cheng, C. Dale and J. Liu, "Statistics and Social Network of YouTube Videos," 2008 *16th Interntional Workshop on Quality of Service,* Enschede, 2008, pp. 229-238, doi: 10.1109/IWQOS.2008.32.

Week 6: News

Technical documentation

- `newsAPI` https://github.com/mkearney/newsAPI
- News API https://newsapi.org/

Examples of research

- Lei Guo & Yiyan Zhang (2020) Information Flow Within and Across Online Media Platforms: An Agenda-setting Analysis of Rumor Diffusion on News Websites, Weibo, and WeChat in China, *Journalism Studies*, 21:15, 2176-2195, DOI: 10.1080/1461670X.2020.1827012
- S. Mo Jang, Brooke W. Mckeever, Robert Mckeever & Joon Kyoung Kim (2019) From Social Media to Mainstream News: The Information Flow of the Vaccine-Autism Controversy in the US, Canada, and the UK, *Health Communication*, 34:1, 110-117, DOI: 10.1080/10410236.2017.1384433
- Matthew S. Weber, Peter Monge, The Flow of Digital News in a Network of Sources, Authorities, and Hubs, *Journal of Communication*, Volume 61, Issue 6, December 2011, Pages 1062–1081, https://doi.org/10.1111/j.1460-2466.2011.01596.x

Week 7: Project Meetings & Worktime

Meet with course instructor

Week 8: Presentations