Welcome to CS224C, everyone!
Instructor and CAs

Diyi Yang  Kaitlyn Zhou  Wanna Qin
Course Overview

Website:
http://web.stanford.edu/class/cs224c/

Ed Discussion:
https://edstem.org/us/courses/32382/discussion/
Learning Objectives

Quantitative analysis of social phenomena
Models of network structure
Methods for text analysis
Applications to social science fields, such as political science, sociolinguistics, sociology, and economics
Additional Learning Objectives

✧ Reading and understanding contemporary research papers
✧ Presenting concise and informative summaries of research
✧ Executing computational social science research
Course Setup

(1) Lectures given by the Instructor
   NLP basics
   Statistical and casual inference

(2) Discussion led by students
   Some key techniques in readings will be covered in pre-recorded lectures
Discussion Led by Students

Five key topics:

1. **Social Influence**: emotion contagion, weak and strong ties, social comparison
2. **Language and Attitude Change**: argumentation, deception, persuasion
3. **Fake News and Misinformation**: rumors, deepfake, prebunking
4. **Prosocial Behavior**: politeness, positive reframing, social support
5. **Prejudice and Stigma**: microaggression, bias, stigma and social movement
Grading

Project (60%)
Presentation or Scribe (10%)
Reading Responses (15%)
Quizzes (15%)
Class Participation (5%)
Project

Group Project
  1~3 people per group
  Please discuss your project idea with instructor/TA early in the course

Literature Review (10%)
Experiment Protocol (15%)
Final Paper + Poster Presentation (30%)
Presentation

2 students to work together and deliver a lecture on a given topic

Work together to well cover the material
Make easy-to-understand slides
20-minute presentation, followed by 25-minute discussion
Be prepared for Q&As
Please send your draft slides to the instructor/TA 2 days before the class
Presentation: **Dos and Don’ts**

**Dos**
- Coherent
- Interactive and engaging

**Don’ts**
- Simply summarizing the content
- No question/interest from the audience
Scribe (equivalent to Presentation)

2 students scribe a discussion session or a video lecture

A blog post that summarizes the topic, the reading, and the discussion content (e.g., a 10-min read for getting to know ways to combat misinformation)

Will be released on the course website

Due one week after each discussion session
Reading Responses

Reply to Ed discussion posts about the reading assigned for a particular class

No need to do reading response if you are the discussion leader or scribe

You’re also welcome to post any other thoughts about the readings

- critique certain features of the papers
- identify potentially important issues not covered in the papers
- suggest new research questions stimulated by the papers
- think about new ways to improve the work
Quiz

48 hours to finish the quiz once it is released

Duration is 1 hour

Only 1 attempt is allowed

No collaboration will be permitted
Logistics and Other Information

Course Contacts:

**Webpage:** materials and announcements

**Ed discussion:** discussion forum

**Other personal issues:** email cs224c-staff@lists.stanford.edu

Computing Resources:

Experiments can take up to hours, even with efficient computation
Academic Integrity

This class abides by the Honor Code

We take academic integrity **seriously**

You are encouraged to discuss readings/project with your classmates; however, what you hand in should be your own work

  - Okay to use open-source software, however, do acknowledge
  - Copying/reusing code is not allowed; strict action will be taken if similarities are found
  - Copying content from other published work (without citing it) or ChatGPT is also not allowed, and is considered plagiarism
Late Policy

Reading responses are due at 5pm PT on the day before the class.

Presentations need to be sent to the instructor or TA by 5pm PT 2 days before the relevant class meeting.

4 late days to use in total.
Course Materials

Readings are available on the course website.
Readings are subject to change, so always double check

No official text books

Lecture slides will be made available on the course website
Expectation and Prerequisites

Prerequisites

- **CS 106B is strictly required**; Programming background
- Basics in machine learning and data science
- Passion for topics on **Social + NLP 😊**

Expectation

- High-quality course project
- Read research papers from different research fields and venues (e.g., ACL, EMNLP, NAACL, CSCW, SIGCHI, Science, etc.)
Books to Check Out (Optional)

An Introduction to Statistical Learning by James, Witten, Hastie, and Tibshirani
Bit by Bit: Social Research in the Digital Age by Sagalnik
Networks, Crowds, and Markets by Easley and Kleinberg
Six Degrees by Duncan Watts
On Individuality and Social Forms by Georg Simmel
Writing for Social Scientists by Howard Becker
Natural Language Processing with Python by Steven Bird, Ewan Klein, and Edward Lope
Introduction to Computational Social Science
Computational Social Science

“A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors”
The Cross-Disciplinary Flavor of CSS

Cross-disciplinary research and application field with theoretical and methodological aspects in computational and social sciences.

Related fields:
- NLP/ML/CV, Data science
- Communication
- Human computer interaction
- Sociological, psychological, economics
- Political science, social science
Online Interaction Generates Big Unstructured Text Data

**Volume**
- 2 Wikipedia revisions per second

**Velocity**

**Variety**
- Tweets, articles, slangs, news, etc
Online Interaction in Text Format Grows Exponentially

between human and human
between human and machines
Opportunities: Data & Social Phenomena

Data

- Speech data is expensive; social media data is a good proxy
- Personal conversations
- Socially grounded data
- Evolution of new words and slang

Social phenomena

- Hate speech, fake news, misinformation, online counseling..
Opportunities and Benefits

1. Provide information about social relationships (e.g., emails)
2. Analyze how group interactions predict individual behaviors

Coordination of the user (as speaker) and, respectively, towards the user (as target) in the months before and after status change occurs.

Opportunities and Benefits

1. Provide information about social relationships (e.g., emails)
2. Analyze how group interactions predict individual behaviors
3. **Understand how the structures of society change evolve over time**
Opinion Polarization In The Digital Era

Democrats and Republicans More Ideologically Divided than in the Past

Distribution of Democrats and Republicans on a 10-item scale of political values

1994

MEDIAN Democrat
Consistently liberal

MEDIAN Republican
Consistently conservative

2004

MEDIAN Democrat
Consistently liberal

MEDIAN Republican
Consistently conservative

2014

MEDIAN Democrat
Consistently liberal

MEDIAN Republican
Consistently conservative

Source: 2014 Political Polarization in the American Public
Notes: Ideological consistency based on a scale of 10 political values questions (see Appendix A). The blue area in this chart represents the ideological distribution of Democrats; the red area of Republicans. The overlap of these two distributions is shaded purple. Republicans include Republican-leanin independents; Democrats include Democratic-leanin independents (see Appendix B).

PEW RESEARCH CENTER

https://www.pewresearch.org/politics/2017/10/05/1-partisan-divides-over-political-values-widen/

Opportunities and Benefits

1. Provide information about social relationships (e.g., emails)
2. Analyze how group interactions predict individual behaviors
3. Understand how the structures, network of society change evolve over time
4. Large-scale tracing of people’s movements and physical proximities

Opportunities and Benefits

1. Provide information about social relationships (e.g., emails)
2. Analyze how group interactions predict individual behaviors
3. Understand how the structures, network of society change evolve over time
4. Large-scale tracing of people’s movements and physical proximities
5. Offer channels for understanding what people say and how they connect
6. Understand the impact of users’ digital activities on everything from their moods, political ideology, to their health
Community structure of political blogs

"Being A Voter”
Vs
“Voting”

(Bryan Et Al., 2011)

Image: https://commons.wikimedia.org/wiki/File:Police_body_cam.png
Opportunities and Benefits

1. Provide information about social relationships (e.g., emails)
2. Analyze how group interactions predict our power and performance
3. Understand how the structures, network of society change evolve over time
4. Large-scale tracing of people’s movements and physical proximities
5. Offer channels for understanding what people say and how they connect
6. Understand the impact of users’ digital activities on everything from their moods, political ideology, to their health

7. **Analyze how technology affects the society as a whole**
Source: https://www.wsj.com/articles/deepfake-videos-are-ruining-lives-is-democracy-next-1539595787
Computational Social Science in a nutshell

Data → Algorithm → Problem → Knowledge → Impact
Risks

1. The potential risk to individuals and corporations in the sharing of personal data by private companies
2. Robust models of collaboration and data sharing between industry and academia
3. Potential risks of de-anonymization
4. Ethical concerns & Institutional Review Boards

Ethics in CSS

IRB is a floor not a ceiling
Put yourself in everyone else’s shoes
Think of research ethics as continuous not discrete
Always
  Design ethically thoughtful research
  Explain your decisions to others

Credit to Matthew Salganik @ Princeton
Challenges

1. The complexity of the theoretical issues confronting social science
2. The difficulty in obtaining the relevant observational data
3. The difficulty of manipulating large scale social organizationals experimentally

4. The complexity and difficulty in computationally, scientifically and rigorously modeling such problems and data
Methods Covered in CS224C

1. Working with social data
2. Inferring sentiment and affect
3. Topic modeling for the social sciences
4. Deep learning for computational social science
5. Data annotation
6. Statistical hypothesis testing
7. Casual inference
8. Word embedding meets social applications
What’s Next?

Sign up for presentation/scribe!
Sign up for Ed discussion!