

Syllabus

Introduction to Computational Social Science: Theory and Practice

Course description

This course is an introduction to the theory and practice of computational social science (CSS), an interdisciplinary field at the intersection of computer science, statistics, and the social sciences. CSS researchers apply computational methods to study social phenomena. The course will cover a range of topics—including text analysis, simulations, and network analysis—with a continuous focus on the epistemological approach of the methods and the processes behind the data.

Audience

This course is catered to advanced undergraduates in the social sciences who have some familiarity with computer programming. The exercises and activities may require students to use programming languages unfamiliar to them, but the focus of the course remains conceptual rather than on the technical details of the methods.

Format

The course is designed to be taught in a flipped classroom format, with students reading the assigned readings before class and then discussing the readings or working on exercises in class. The syllabus is designed to be taught over the course of a quarter (ten weeks).

Learning Objectives

By the end of the course, students should be able to:

- Define computational social science as a field and explain how it differs from other fields
- Explain the strengths and weaknesses of different computational social science methods

- Understand how to apply computational social science methods to answer research questions
- Evaluate computational social science research

Activities

Week	Topic	Description
1	Introduction	Finding CSS Research. Based on the descriptions we have read, find a peer-reviewed paper you think fits the description of computational social science and discuss why.
2	Prediction and Explanation	Same dataset, two stories. In this lab, we will use the same data both to try to explain and predict and discuss the limitations and advantages of each approach.
3	Simulations and Agent-based Models (ABMs)	Agent-based models in NetLogo. We will first step through the basics of creating agent-based models. Then, you will be tasked with exploring the examples on NetLogo Web for one which demonstrates a social process. Discuss why modeling may or may not be a good approach to this domain.
4	Ethics and Best Practices	Deanonimization activity. Find your “ individual risk ” from Rocher et. al (2019).
5	Text as Data	NLP project. Replication of the Mosteller and Wallace (1963) identification of Federalist Papers authors using latent text data.
6	Experiments and Causal Inference	Design a Virtual Experiment. Students design a hypothetical experiment to answer a social science question, addressing issues like sample selection, control groups, and causal inference. They then present and critique each other’s designs.
7	Network Analysis	Lab. Introduction to working with network data with classic data sets such as Zachary’s karate club and the florentine families dataset (Padget & Ansell, 1993).
8	Crowds and Communities	Community Analysis Project. Select an online community and discuss how one might use computational tools to analyze its dynamics, norms, and behavior. Then, we will work together using data derived from Pushshift on Reddit.
9	Wrapping Up	Student Presentations
10		Student Presentations

Readings

All selected readings are open access.

Defining Computational Social Science

What exactly is “computational social science”?

We define computational social science as a field of study which uses computational methods to study social phenomena. We also discuss the history of the field and some of the challenges it faces including access to data and institutional structures.

- David Lazer et al., “Computational Social Science,” *Science* 323, no. 5915 (February 2009): 721–23, <https://doi.org/10.1126/science.1167742>.
- “Introduction,” in *Bit by Bit: Social Research in the Digital Age* (Princeton: Princeton University Press, 2018), 1–12.
- Gary King, Jennifer Pan, and Margaret E. Roberts, “How Censorship in China Allows Government Criticism but Silences Collective Expression,” *American Political Science Review* 107, no. 2 (May 2013), <https://doi.org/10.1017/S0003055413000014>.

Prediction and Explanation

On computational social science’s epistemological perspectives.

A continuous tension we see is how different stakeholders view the value of data. Some may be interested in models which give us insight into future events, while other may be interested in models which help us understand the underlying mechanisms of a process. In what ways do prediction and explanation differ? In what cases might we wish to use each approach?

- Hanna Wallach, “Computational Social Science $\neq$ Computer Science + Social Data,” *Communications of the ACM* 61, no. 3 (February 2018): 42–44, <https://doi.org/10.1145/3132698>.
- “Observing Behavior,” in *Bit by Bit: Social Research in the Digital Age* (Princeton: Princeton University Press, 2018), 13–83.
- Jake M. Hofman et al., “Integrating Explanation and Prediction in Computational Social Science,” *Nature* 595, no. 7866 (July 2021): 181–88, <https://doi.org/10.1038/s41586-021-03659-0>.

Simulations and Agent-based Models (ABMs)

How can we use computer simulations to study social phenomena from the “bottom up”?

We discuss the role of simulations and when they may be useful in the development and explanation of theories or in forecasting.

- Rosaria Conte and Mario Paolucci, “On Agent-Based Modeling and Computational Social Science,” *Frontiers in Psychology* 5 (2014), <https://doi.org/10.3389/fpsyg.2014.00668>.
- Ivan Smirnov, Camelia Oprea, and Markus Strohmaier, “Toxic Comments Are Associated with Reduced Activity of Volunteer Editors on Wikipedia,” *PNAS Nexus* 2, no. 12 (December 2023): pgad385, <https://doi.org/10.1093/pnasnexus/pgad385>.

Ethics and Best Practices

What are the pitfalls and potential ethical issues in computational social science research?

We discuss such challenges for computational social science in practice as reidentification, potential effects on privacy, and how more data alone does not solve study design problems.

- “Ethics,” in *Bit by Bit: Social Research in the Digital Age* (Princeton: Princeton University Press, 2018), 281–354.
- Charlotte Jee, “You’re Very Easy to Track down, Even When Your Data Has Been Anonymized,” *MIT Technology Review* (<https://www.technologyreview.com/2019/07/23/134090/youre-very-easy-to-track-down-even-when-your-data-has-been-anonymized/>, July 2019).
- Matthew Zook et al., “Ten Simple Rules for Responsible Big Data Research,” *PLOS Computational Biology* 13, no. 3 (March 2017): e1005399, <https://doi.org/10.1371/journal.pcbi.1005399>.
- David Lazer et al., “The Parable of Google Flu: Traps in Big Data Analysis,” *Science* 343, no. 6176 (March 2014): 1203–5, <https://doi.org/10.1126/science.1248506>.

Text as Data

Methods for working with text data.

A lot of social data is encoded within unstructured text. This module is more practical than theoretical and focuses on strategies to extract data from text using natural language processing and modern, vector-based approaches.

- Paul DiMaggio, “Adapting Computational Text Analysis to Social Science (and Vice Versa),” *Big Data & Society* 2, no. 2 (December 2015): 2053951715602908, <https://doi.org/10.1177/2053951715602908>.

- Jacob Jensen et al., “Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech [with Comments and Discussion],” *Brookings Papers on Economic Activity*, 2012, 1–81, <https://www.jstor.org/stable/41825364>.

Experiments and Causal Inference

How can we answer cause-and-effect questions using computational social science?

Experiments allow the researcher to manipulate independent variables and observe the effect on dependent variables. However, experiments are not always possible. Causal inference provides a framework to answer causal questions even when experiments are not possible.

- “Running Experiments,” in *Bit by Bit: Social Research in the Digital Age* (Princeton: Princeton University Press, 2018), 147–229.
- Justin Grimmer, “We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together,” *PS: Political Science & Politics* 48, no. 1 (January 2015): 80–83, <https://doi.org/10.1017/S1049096514001784>.
- Eshwar Chandrasekharan et al., “You Can’t Stay Here: The Efficacy of Reddit’s 2015 Ban Examined Through Hate Speech,” *Proceedings of the ACM on Human-Computer Interaction* 1, no. CSCW (December 2017): 1–22, <https://doi.org/10.1145/3134666>.

Network Analysis

Much social data is produced in the context of networks of relationships. This section introduces the basic concepts of network analysis, and provides a few examples of how it is used in the social sciences.

- Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts, “An Experimental Study of Search in Global Social Networks,” *Science* 301, no. 5634 (August 2003): 827–29, <https://doi.org/10.1126/science.1081058>.
- Pablo Barberá et al., “The Critical Periphery in the Growth of Social Protests,” *PLOS ONE* 10, no. 11 (November 2015): e0143611, <https://doi.org/10.1371/journal.pone.0143611>.
- Christopher A. Bail et al., “Exposure to Opposing Views on Social Media Can Increase Political Polarization,” *Proceedings of the National Academy of Sciences* 115, no. 37 (September 2018): 9216–21, <https://doi.org/10.1073/pnas.1804840115>.

Crowds and Communities

A lot of social data is not produced in isolation, but rather in the context of communities with their own norms and practices. We discuss how to think about communities and crowds, and how to study them.

- “Creating Mass Collaboration,” in *Bit by Bit: Social Research in the Digital Age* (Princeton: Princeton University Press, 2018), 231–80.
- Aaron Shaw and Benjamin Mako Hill, “Laboratories of Oligarchy? How the Iron Law Extends to Peer Production,” *Journal of Communication* 64, no. 2 (2014): 215–38, <https://doi.org/10.1111/jcom.12082>.
- Lev Muchnik, Sinan Aral, and Sean J. Taylor, “Social Influence Bias: A Randomized Experiment,” *Science* 341, no. 6146 (August 2013): 647–51, <https://doi.org/10.1126/science.1240466>.

Wrapping Up

We synthesize the main themes of the course and discuss the future of computational communication research.

- Wouter van Attevelde and Tai-Quan Peng, “When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science,” *Communication Methods and Measures* 12, no. 2-3 (April 2018): 81–92, <https://doi.org/10.1080/19312458.2018.1458084>.
- Alexandra Olteanu et al., “Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries,” *Frontiers in Big Data* 2 (2019), <https://doi.org/10.3389/fdata.2019.00013>.