

Readings Overview

Table of contents

Defining Computational Social Science	1
Prediction and Explanation	2
Simulations and Agent-based Models (ABMs)	2
Ethics and Best Practices	2
Text as Data	3
Experiments and Causal Inference	3
Network Analysis	3
Crowds and Communities	4
Wrapping Up	4

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.

- Lazer et al. (2009): Authored by several prominent names in the social sciences, this paper presents a vision for computational social science as a field of study and some of the challenges facing the field. [\[link\]](#)
- “Introduction” (2018): Introduction from probably the best textbook on computational social science out there. [\[link\]](#)
- King, Pan, and Roberts (2013): Example of an excellent computational social science paper that uses data to explain a social process. [\[link\]](#)

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?

- Wallach (2018): Perspective on the differences between machine learning and computational social science and why it matters. [\[link\]](#)
- “Observing Behavior” (2018): Discusses characteristics of big data and three “research strategies” for working with it: observations, forecasting, and quasi-experiments. [\[link\]](#)
- Hofman et al. (2021): Perspective on the differences between explanation and prediction and possible ways to integrate the two approaches in computational social science. [\[link\]](#)

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.

- Conte and Paolucci (2014): Proposes an interdisciplinary approach that combines ABM and CSS for advancing the computational study of social phenomena. [\[link\]](#)
- Smirnov, Oprea, and Strohmaier (2023): A paper which integrates computational analysis with an agent-based model to demonstrate the potential impacts of its findings. [\[link\]](#)

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” (2018): Practical examples and advice for ethical approaches to computational social science research. [\[link\]](#)
- Charlotte Jee (2019): A short article about how easy it is to deanonymize data. [\[link\]](#)
- Zook et al. (2017): Suggestions to practitioners for how to approach and think about ethical issues when working with big data. [\[link\]](#)
- Lazer et al. (2014): A classic example of “big data hubris”, where one might be tempted to ignore foundational issues just because they have access to a lot of data. [\[link\]](#)

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.

- DiMaggio (2015): Discusses the differences in how social scientists and computer scientists approach textual data and offers suggestions that may help bridge the gap. [\[link\]](#)
- Jensen et al. (2012): Applies textual analysis to the *Congressional Record* and compares it to Google Books to determine the relationship of polarization with “elite discourse”. [\[link\]](#)

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” (2018): How can we answer cause-and-effect questions using computational social science? Salganik discusses ways that digital media can facilitate experiments and how causal experiment design can be used on existing data. [\[link\]](#)
- Grimmer (2015): Grimmer argues that approaches from both computer science and the social sciences are needed to use big data toward solving large problems. In particular, this paper emphasizes that description, while underappreciated, is still an important part of the scientific process. [\[link\]](#)
- Chandrasekharan et al. (2017): Applies matching to a dataset of Reddit activity to evaluate the effectiveness of Reddit’s quarantine policy. [\[link\]](#)

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.

- Dodds, Muhamad, and Watts (2003): Uses email to replicate the “small-world” experiment of Milgram (1967). [\[link\]](#)
- Barberá et al. (2015): Uses Twitter data to study the influence of “peripheral” participants on the spread of social movement, finding that despite being less active, such users can be just as important as “core” users. [\[link\]](#)

- Bail et al. (2018): Do people tend to become less polarized when exposed to opposing views? This study paid active Twitter users to follow a bot which reposted opposing viewpoints. It finds that exposure to opposing views does not reduce polarization, and in fact can increase it. [\[link\]](#)

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” (2018): Categorizes three types of collaborative processes: human computation, open call, and distributed data collection. [\[link\]](#)
- Shaw and Hill (2014): A study on participation inequalities in online peer-production communities applying a political theory from 1911. [\[link\]](#)
- Muchnik, Aral, and Taylor (2013): A sort of social process audit on a news aggregation website, finding an asymmetric effect where users self-correct negative scores, but do not do the same on positive scores. [\[link\]](#)

Wrapping Up

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

- van Atteveldt and Peng (2018): An overview of the ways that computational techniques are changing communication research, with an emphasis on many of the themes and challenges discussed throughout this course. [\[link\]](#)
- Olteanu et al. (2019): An overview of many of the biases that can be introduced by social data. [\[link\]](#)

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. “Exposure to Opposing Views on Social Media Can Increase Political Polarization.” *Proceedings of the National Academy of Sciences* 115 (37): 9216–21. <https://doi.org/10.1073/pnas.1804840115>.

Barberá, Pablo, Ning Wang, Richard Bonneau, John T. Jost, Jonathan Nagler, Joshua Tucker, and Sandra González-Bailón. 2015. “The Critical Periphery in the Growth of Social Protests.” *PLOS ONE* 10 (11): e0143611. <https://doi.org/10.1371/journal.pone.0143611>.

Chandrasekharan, Eshwar, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. “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 (CSCW): 1–22. <https://doi.org/10.1145/3134666>.

- Charlotte Jee. 2019. “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/>.
- Conte, Rosaria, and Mario Paolucci. 2014. “On Agent-Based Modeling and Computational Social Science.” *Frontiers in Psychology* 5. <https://doi.org/10.3389/fpsyg.2014.00668>.
- “Creating Mass Collaboration.” 2018. In *Bit by Bit: Social Research in the Digital Age*, 231–80. Princeton: Princeton University Press.
- DiMaggio, Paul. 2015. “Adapting Computational Text Analysis to Social Science (and Vice Versa).” *Big Data & Society* 2 (2): 2053951715602908. <https://doi.org/10.1177/2053951715602908>.
- Dodds, Peter Sheridan, Roby Muhamad, and Duncan J. Watts. 2003. “An Experimental Study of Search in Global Social Networks.” *Science* 301 (5634): 827–29. <https://doi.org/10.1126/science.1081058>.
- “Ethics.” 2018. In *Bit by Bit: Social Research in the Digital Age*, 281–354. Princeton: Princeton University Press.
- Grimmer, Justin. 2015. “We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together.” *PS: Political Science & Politics* 48 (1): 80–83. <https://doi.org/10.1017/S1049096514001784>.
- Hofman, Jake M., Duncan J. Watts, Susan Athey, Filiz Garip, Thomas L. Griffiths, Jon Kleinberg, Helen Margetts, et al. 2021. “Integrating Explanation and Prediction in Computational Social Science.” *Nature* 595 (7866): 181–88. <https://doi.org/10.1038/s41586-021-03659-0>.
- “Introduction.” 2018. In *Bit by Bit: Social Research in the Digital Age*, 1–12. Princeton: Princeton University Press.
- Jensen, Jacob, Suresh Naidu, Ethan Kaplan, Laurence Wilse-Samson, David Gergen, Michael Zuckerman, and Arthur Spirling. 2012. “Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech [with Comments and Discussion].” *Brookings Papers on Economic Activity*, 1–81. <https://www.jstor.org/stable/41825364>.
- King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. “How Censorship in China Allows Government Criticism but Silences Collective Expression.” *American Political Science Review* 107 (2). <https://doi.org/10.1017/S0003055413000014>.
- Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. “The Parable of Google Flu: Traps in Big Data Analysis.” *Science* 343 (6176): 1203–5. <https://doi.org/10.1126/science.1248506>.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, et al. 2009. “Computational Social Science.” *Science* 323 (5915): 721–23. <https://doi.org/10.1126/science.1167742>.
- Muchnik, Lev, Sinan Aral, and Sean J. Taylor. 2013. “Social Influence Bias: A Randomized Experiment.” *Science* 341 (6146): 647–51. <https://doi.org/10.1126/science.1240466>.
- “Observing Behavior.” 2018. In *Bit by Bit: Social Research in the Digital Age*, 13–83. Princeton: Princeton University Press.
- Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. “Social Data:

- Biases, Methodological Pitfalls, and Ethical Boundaries.” *Frontiers in Big Data* 2. <https://doi.org/10.3389/fdata.2019.00013>.
- “Running Experiments.” 2018. In *Bit by Bit: Social Research in the Digital Age*, 147–229. Princeton: Princeton University Press.
- Shaw, Aaron, and Benjamin Mako Hill. 2014. “Laboratories of Oligarchy? How the Iron Law Extends to Peer Production.” *Journal of Communication* 64 (2): 215–38. <https://doi.org/10.1111/jcom.12082>.
- Smirnov, Ivan, Camelia Oprea, and Markus Strohmaier. 2023. “Toxic Comments Are Associated with Reduced Activity of Volunteer Editors on Wikipedia.” *PNAS Nexus* 2 (12): pgad385. <https://doi.org/10.1093/pnasnexus/pgad385>.
- van Atteveldt, Wouter, and Tai-Quan Peng. 2018. “When Communication Meets Computation: Opportunities, Challenges, and Pitfalls in Computational Communication Science.” *Communication Methods and Measures* 12 (2-3): 81–92. <https://doi.org/10.1080/19312458.2018.1458084>.
- Wallach, Hanna. 2018. “Computational Social Science $\neq$ Computer Science + Social Data.” *Communications of the ACM* 61 (3): 42–44. <https://doi.org/10.1145/3132698>.
- Zook, Matthew, Solon Barocas, danah boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, et al. 2017. “Ten Simple Rules for Responsible Big Data Research.” *PLOS Computational Biology* 13 (3): e1005399. <https://doi.org/10.1371/journal.pcbi.1005399>.