SOC576
COMPUTATIONAL SOCIOLOGY
Syllabus

Thomas Davidson

Spring 2021

COURSE DESCRIPTION

This course is designed to introduce students to computational methods and their applications to sociological research. We will discuss the computational toolkit from the bottom up, beginning with the fundamentals of programming and data analysis and management. Once these fundamentals are in place, we will turn to different methods for collecting data: application programming interfaces, web-scraping, and online experiments. The remainder of the course will focus on computational methods for data analysis. First, we will cover various methods for the quantitative analysis of text data including word embeddings, topic modeling, and supervised learning. Second, we will discuss supervised machine learning in more depth, assessing the relationship between prediction and explanation in social science, discussing bias and other limitations of these methods, as well as the opportunities these techniques present to work with images and multimodal data. In the last two weeks of the course we will discuss recent advances in social network analysis and agent-based modeling techniques. Throughout the course students will gain hands-on experience with these different techniques, as well as an understanding of how these techniques are being used in cutting-edge sociological research. Overall this course will provide students with a strong conceptual foundation in computational sociology and the ability to apply various techniques for data collection and analysis in their own research. All assignments will be conducted using the R programming language.

PREREQUISITES AND PREPARATION

This course is designed for students without any experience using computational methods or advanced statistics. Nonetheless, the course will proceed more efficiently if students without any such experience are willing to undertake some independent learning prior to the beginning of the course. In particular, I recommend students familiarize themselves with Github, the R programming language, the RStudio computing environment, and RMarkdown documents, as we will be using these tools throughout the course. We will review these topics over the first few weeks of class, but the more familiar students are with these tools, the more time we can spend focusing on their sociological applications.

Chris Bail (a sociologist at Duke University) recently published a series of video tutorials (https://sicss. io/boot_camp) and exercises designed to introduce these tools. I highly recommend that students watch these videos and review the relevant materials in *R for Data Science* (details on the book below) prior to the beginning of the semester.

ASSESSMENT

There will be two types of assessment used in the course. There will be several homework assignments throughout the semester that will be designed to enable students to become familiar with the various methodological techniques covered in the course. A tentative schedule of these assignments can be found in the course outline below. Students will also write an empirical paper over the course of the semester. The paper will involve the collection of original data and preliminary analyses using one or more of the approaches covered in the course. I intend for the paper to be an opportunity for students to develop the basis for a qualifying exam, master's thesis, or dissertation chapter. Students are expected to make progress on the paper over the course of the semester. At various points in the semester I will provide feedback on the paper proposal, the methodology for data collection, and the data analysis (see course outline for a timeline).

READINGS

There are weekly reading assignments for this course. These readings include methodological texts, reviews of relevant methodological and theoretical considerations, and examples of how sociologists and other social scientists apply computational approaches in their research. Given the complexity and unfamiliarity of some of the approaches we will cover in the course, I have included a diverse set of readings for each topic. Some students may find the technical readings more useful whereas others may benefit from the more contextualized applications.

Require texts and useful readings

* indicates a required text. All required texts and useful reading are available for free online on the listed websites.

- *Matthew Salganik. 2017. *Bit by Bit* (https://www.bitbybitbook.com/en/1st-ed/preface/)
- *Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. (*R4DS*). O'Reilly Media, Inc. (https://r4ds.had.co.nz/)
- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Springer Texts in Statistics. (Entire book can be downloaded as a PDF via Rutgers Library)
- Manning, Christopher D., Hinrich Schütze, and Prabhakar Raghavan. 2008. *Introduction to Information Retrieval*. Cambridge University Press. (https://nlp.stanford.edu/IR-book/information-retrieval-book.html)
- Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton University Press. (https://socviz.co/)

COURSE POLICIES

The Rutgers Sociology Department strives to create an environment that supports and affirms diversity in all manifestations, including race, ethnicity, gender, sexual orientation, religion, age, social class, disability status, region/country of origin, and political orientation. This class will be a space for tolerance, respect, and mutual dialogue. Students must abide by the Code of Student Conduct at all times, including during lectures and in participation online.

All students must abide by the university's Academic Integrity Policy. Violations of academic integrity will result in disciplinary action.

In accordance with University policy, if you have a documented disability and require accommodations to obtain equal access in this course, please contact me during the first week of classes. Students with disabilities must be registered with the Office of Student Disability Services and must provide verification of their eligibility for such accommodations.

I will also be making additional accommodations due to the COVID-19 pandemic. If you or your family are affected in any way that impedes your ability to participate in this course, please contact me as soon as you can so that we can make necessary arrangements.

COURSE OUTLINE

This outline is tentative and subject to change.

Week 1

Introduction to Computational Sociology

Readings

- *R4DS*: Preface, C2-6
- *Bit by Bit*, C1
- Lazer, David, et al. 2009. "Life in the Network: The Coming Age of Computational Social Science." Science 323 (5915): 721–23. https://doi.org/10.1126/science.1167742.
- Golder, Scott A., and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research." Annual Review of Sociology 40 (1): 129–52. https://doi.org/10.1146/annurev-soc-071913-043145.
- Bail, Christopher A. 2014. "The Cultural Environment: Measuring Culture with Big Data." Theory and Society 43 (3–4): 465–82. https://doi.org/10.1007/s11186-014-9216-5.
- Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. 2020. "Computational Social Science and Sociology." Annual Review of Sociology 46 (1): annurev-soc-121919-054621. https://doi.org/10.1146/annurev-soc-121919-054621.

Week 2

Data Structures

Readings

- *R4DS*: C7-11
- Freese, Jeremy. 2007. "Replication Standards for Quantitative Social Science: Why Not Sociology?" Sociological Methods & Research 36 (2): 153–72. https://doi.org/10.1177/0049124107306659.
- Freese, Jeremy, and David Peterson. 2017. "Replication in Social Science," Annual Review of Sociology.
- Liu, David, and Matthew Salganik. 2020. "Successes and Struggles with Computational Reproducibility: Lessons from the Fragile Families Challenge," Socius.

Week 3

Programming Fundamentals

Readings

- *R4DS*: C14-17, 21-24

Assignment 1: The computational toolkit.

Week 4

Data Collection I: APIs

Readings

Note: I have included readings that use a range of different APIs including Reddit (Hamilton et al.), Spotify (Askin and Mauskampf), Facebook (Davidson and Berezin), Google Trends (Davidson and Berezin; Bail, Brown, and Wimmer), Twitter (Mitts), and YouTube (Munger and Phillips). My intent is for you to see the breadth of research making use of APIs rather than to go in-depth into any one study. Make sure to read Dean Freelon's article on the "post-API age."

- *Bit by Bit*, C2
- Hamilton, William, Justine Zhang, Cristian Danescu-Niculescu-Mizil, Dan Jurafsky, and Jure Leskovec. 2017. "Loyalty in Online Communities." ICWSM, 10.
- Askin, Noah, and Michael Mauskapf. 2017. "What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music." American Sociological Review 82 (5): 910–44. https://doi.org/10.1177/0003122417728662.
- Davidson, Thomas, and Mabel Berezin. 2018. "Britain First and the UK Independence Party: Social Media and Movement-Party Dynamics." Mobilization: An International Quarterly 23 (4): 485–510. https://doi.org/10.17813/1086-671X-23-4-485.
- Bail, Christopher, Taylor Brown, and Andreas Wimmer. 2019. "Prestige, Proximity, and Prejudice: How Google Search Terms Diffuse across the World." American Journal of Sociology 124 (5): 1496–1548. https://doi.org/10.1086/702007.
- Mitts, Tamar. 2019. "From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West." American Political Science Review 113 (1): 173–94. https://doi.org/10.1017/S0003055418000618.
- Munger, Kevin, and Joseph Phillips. 2020. "Right-Wing YouTube: A Supply and Demand Perspective." The International Journal of Press/Politics, 34.
- Freelon, Deen. 2018. "Computational Research in the Post-API Age." Political Communication 35 (4): 665–68. https://doi.org/10.1080/10584609.2018.1477506.

Week 5

Data Collection II: Webscraping

Readings

- 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 (02): 326–43. https://doi.org/10.1017/S0003055413000014.
- Fiesler, Casey, Nate Beard, and Brian C Keegan. 2020. "No Robots, Spiders, or Scrapers: Legal and Ethical Regulation of Data Collection Methods in Social Media Terms of Service." In Proceedings of the Fourteenth International AAAI Conference on Web and Social Media, 187–96.
- *Bit by Bit*, C6

Assignment 2: Collecting and storing data

Week 6

Data Collection III: Online experiments and surveys

Readings

- *Bit by Bit*, C3-4
- Salganik, Matthew, Peter Sheridan Dodds, and Duncan Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." Science Reports 311: 854–56.
- Munger, Kevin. 2016. "Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment." Political Behavior, November. https://doi.org/10.1007/s11109-016-9373-5.
- Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." International Journal of Forecasting 31 (3): 980–91. https://doi.org/10.101

6/j.ijforecast.2014.06.001.

Paper proposal due

Week 7

Natural Language Processing I: Fundamentals

Readings

- *Introduction to Information Retrieval*, pp. 117-126.
- Grimmer, Justin, and Brandon Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." Political Analysis 21 (3): 267–97. https://doi.org/10.1093/pan/mps028.
- Evans, James, and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." Annual Review of Sociology 42 (1): 21–50. https://doi.org/10.1146/annurev-soc-081715-074206.
- Danescu-Niculescu-Mizil, Cristian, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. "No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities." In Proceedings of the 22nd International Conference on World Wide Web, 307–318. http://dl.acm.org/citation.cfm?id=2488416.
- Nelson, Laura 2017. "Computational Grounded Theory: A Methodological Framework." Sociological Methods & Research, November. https://doi.org/10.1177/0049124117729703.

Week 8

Natural Language Processing II: Word Embeddings

Readings

- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." In Advances in Neural Information Processing Systems, 3111–3119. http://papers.nips.cc/paper/5021-distributed-representations.
- Hamilton, William, Jure Leskovec, and Dan Jurafsky. 2016. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change." In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1489–1501.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." In Proceedings of NAACL-HLT 2019, 4171–86. ACL.
- Kozlowski, Austin, Matt Taddy, and James Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings." American Sociological Review, September, 000312241987713. https://doi.org/10.1177/0003122419877135.

Week 9

Natural Language Processing III: Topic Models

Readings

- Blei, David 2012. "Probabilistic Topic Models." Communications of the ACM 55 (4): 77. https://doi.org/10.1145/2133806.2133826.
- Mohr, John, and Petko Bogdanov. 2013. "Introduction—Topic Models: What They Are and Why They Matter." Poetics 41 (6): 545–69. https://doi.org/10.1016/j.poetic.2013.10.001.
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." Poetics 41 (6): 570–606. https://doi.org/10.1016/j.poetic.2013.08.004.

- Roberts, Margaret, Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses: Structural Topic Models for Survey Responses." American Journal of Political Science 58 (4): 1064–82. https://doi.org/10.1111/ajps.12103.
- Karell, Daniel, and Michael Freedman. 2019. "Rhetorics of Radicalism." American Sociological Review 84 (4): 726–53. https://doi.org/10.1177/0003122419859519.

Assignment 3: Natural language processing

Week 10

Machine Learning I: Prediction and Explanation

Readings

- Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." Annual Review of Sociology 45: 27–45.
- Hofman, Jake, Amit Sharma, and Duncan Watts. 2017. "Prediction and Explanation in Social Systems." Science 355 (6324): 486–488. https://doi.org/10.1126/science.aal3856.
- Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." Journal of Economic Perspectives 31 (2): 87–106. https://doi.org/10.1257/jep.31.2.87.
- Davidson, Thomas. 2019. "Black-Box Models and Sociological Explanations: Predicting High School Grade Point Average Using Neural Networks." Socius: Sociological Research for a Dynamic World 5 (January): 237802311881770. https://doi.org/10.1177/2378023118817702.
- Hanna, Alex. 2013. "Computer-Aided Content Analysis of Digitally Enabled Movements." Mobilization: An International Quarterly 18 (4): 367–388.

Paper initial data collection due

Week 11

Machine learning II: Challenges

Readings

- Salganik, Matthew, Ian Lundberg, Alexander Kindel, et al. 2020. "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration." Proceedings of the National Academy of Sciences.
- Buolamwini, Joy, and Timnit Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In Proceedings of Machine Learning Research, 81:1–15.
- Davidson, Thomas, Debasmita Bhattacharya, and Ingmar Weber. 2019. "Racial Bias in Hate Speech and Abusive Language Detection Datasets." In Proceedings of the Third Workshop on Abusive Language Online, 25–35. Florence, Italy: Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-3504.
- De-Arteaga, Maria, Alexey Romanov, Hanna Wallach, Jennifer Chayes, Christian Borgs, Alexandra Chouldechova, Sahin Geyik, Krishnaram Kenthapadi, and Adam Tauman Kalai. 2019. "Bias in Bios: A Case Study of Semantic Representation Bias in a High-Stakes Setting." In Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19, 120–28. Atlanta, GA, USA: ACM Press. https://doi.org/10.1145/3287560.3287572.
- Gonen, Hila, and Yoav Goldberg. 2019. "Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But Do Not Remove Them." In Proceedings of NAACL_HLT, 609–14. Minneapolis, Minnesota: Association for Computational Linguistics.

Week 12

Machine learning III: Image classification

Readings

- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." In Advances in Neural Information Processing Systems, 1097–1105. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.
- Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." Science 353 (6301): 790–94. https://doi.org/10.1126/science.aaf7894.
- Gebru, Timnit, Jonathan Krause, Yilun Wang, Duyun Chen, Jia Deng, Erez Lieberman Aiden, and Li Fei-Fei. 2017. "Using Deep Learning and Google Street View to Estimate the Demographic Makeup of Neighborhoods across the United States." Proceedings of the National Academy of Sciences 114 (50): 13108–13. https://doi.org/10.1073/pnas.1700035114.
- Zhang, Han, and Jennifer Pan. 2019. "CASM: A Deep-Learning Approach for Identifying Collective Action Events with Text and Image Data from Social Media." Sociological Methodology 49 (1): 1–57. https://doi.org/10.1177/0081175019860244.
- Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, et al. 2017. "Mastering the Game of Go without Human Knowledge." Nature 550 (7676): 354–59. https://doi.org/10.1038/nature24270.

Assignment 4: Machine learning

Week 13

Social Network Analysis

Readings

- Watts, Duncan. 1999. "Networks, Dynamics, and the Small-World Phenomenon." American Journal of Sociology 105 (2): 493–527. https://doi.org/10.1086/210318.
- Aral, Sinan, Lev Muchnik, and Arun Sundararajan. 2009. "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks." Proceedings of the National Academy of Sciences 106 (51): 21544–21549.
- González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. 2011. "The Dynamics of Protest Recruitment through an Online Network." Scientific Reports 1 (1): 197. https://doi.org/10.1038/srep00197.
- Goel, Sharad, Ashton Anderson, Jake Hofman, and Duncan J. Watts. 2015. "The Structural Virality of Online Diffusion." Management Science, July, 150722112809007. https://doi.org/10.1287/mnsc.2015.2158.
- Block, Per, Christoph Stadtfeld, and Tom Snijders. 2019. "Forms of Dependence: Comparing SAOMs and ERGMs From Basic Principles." Sociological Methods & Research 48 (1): 202–39. https://doi.org/10.1177/0049124116672680.

Paper preliminary analyses due

Week 14

Agent-Based Models

Readings

- Macy, Michael, and Robert Willer. 2002. "From Factors to Factors: Computational Sociology and Agent-Based Modeling." Annual Review of Sociology 28 (1): 143–66. https://doi.org/10.1146/annurev.soc.28.110601.141117.
- Bruch, Elizabeth, and Jon Atwell. 2015. "Agent-Based Models in Empirical Social Research." Sociological Methods & Research 44 (2): 186–221. https://doi.org/10.1177/0049124113506405.
- Centola, Damon, and Michael Macy. 2007. "Complex Contagions and the Weakness of Long Ties." American Journal of Sociology 113 (3): 702–34. https://doi.org/10.1086/521848.
- Centola, Damon. 2015. "The Social Origins of Networks and Diffusion." American Journal of Sociology 120 (5): 1295–1338. https://doi.org/10.1086/681275.
- DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why Do Liberals Drink Lattes?" American Journal of Sociology 120 (5): 1473–1511. https://doi.org/10.1086/681254.

Assignment 5: Social networks and agent-based models

Work on final projects

Final paper due (date to be confirmed)