Topics in Computational Social Science: Option Reading List

Note: starred (*) papers are for presentation to the group by the assigned student for that week. These are the core readings for the week.

Week 1: Introduction to Computational Social Science - Discussion

- * Kashyap, Ridhi, R. Gordon Rinderknecht, Aliakbar Akbaritabar, Diego Alburez-Gutierrez, Sofia Gil-Clavel, André Grow, Jisu Kim et al. "Digital and computational demography." In *Research Handbook on Digital Sociology*, pp. 48-86. Edward Elgar Publishing, 2023.
- * Lazer, David, Eszter Hargittai, Deen Freelon, Sandra Gonzalez-Bailon, Kevin Munger, Katherine Ognyanova, and Jason Radford. 2021. 'Meaningful Measures of Human Society in the Twenty-First Century'. *Nature* 595(7866):189–96. doi: [10.1038/s41586-021-03660-7](#).
- * Edelmann, Achim, Tom Wolff, Danielle Montagne, and Christopher A. Bail. "Computational social science and sociology." *Annual review of sociology* 46, no. 1 (2020): 61-81.
- Lazer, David M. J., Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts, Alondra Nelson, Matthew J. Salganik, Markus Strohmaier, Alessandro Vespignani, and Claudia Wagner. 2020. 'Computational Social Science: Obstacles and Opportunities'. *Science* 369(6507):1060–62. doi: [10.1126/science.aaz8170](#).
- Salganik, Matthew J. 2019. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Chapter 1.

Week 2: Social data in the digital age - opportunities and challenges - Lab
Note: No paper presentations this week.

- Ruggles, Steven. 2014. 'Big Microdata for Population Research'. *Demography* 51(1):287–97. doi: [10.1007/s13524-013-0240-2](#).
- * Kashyap, Ridhi. 2021. 'Has Demography Witnessed a Data Revolution? Promises and Pitfalls of a Changing Data Ecosystem'. *Population Studies* 75(sup1):47–75. doi: [10.1080/00324728.2021.1969031](#).

- * Salganik, Matthew J. 2019. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Chapter 2.
- 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. doi: 10.1126/science.1248506.
- * Ruggles, Steven, Catherine A. Fitch, and Evan Roberts. 2018. 'Historical Census Record Linkage'. *Annual Review of Sociology* 44(1):19–37. doi: 10.1146/annurev-soc-073117-041447.
- Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt. 2022. 'Social Capital I: Measurement and Associations with Economic Mobility'. *Nature* 608(7921):108–21. doi: [10.1038/s41586-022-04996-4](10.1038/s41586-022-04996-4).
- * Bruch, Elizabeth E., and M. E. J. Newman. 2018. 'Aspirational Pursuit of Mates in Online Dating Markets'. *Science Advances* 4(8):eaap9815. doi: 10.1126/sciadv.aap9815.
- Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. 'Fake News on Twitter during the 2016 U.S. Presidential Election'. *Science* 363(6425):374–78. doi: 10.1126/science.aau2706.

Week 2: Collecting Facebook data using an Application Program Interface (API) [LAB 1]

- * Fatehkia, Masoomali, Ridhi Kashyap, and Ingmar Weber. "Using Facebook ad data to track the global digital gender gap." *World Development* 107 (2018): 189-209.
- Zagheni, Emilio, Ingmar Weber, and Krishna Gummadi. 2017. 'Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants'. *Population and Development Review* 43(4):721–34.
- Araujo, Matheus, Yelena Mejova, Ingmar Weber, and Fabricio Benevenuto. "Using Facebook ads audiences for global lifestyle disease surveillance: Promises and limitations." In *Proceedings of the 2017 ACM on Web science conference*, pp. 253-257. 2017.
- Collect original data on the count of Facebook users broken down by characteristics using the Facebook marketing API

Week 3: Machine learning/prediction - discussion

- * Salganik, Matthew J., Ian Lundberg, …, and Sara McLanahan. 2020. 'Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration'. Proceedings of the National Academy of Sciences 117(15):8398–8403. doi: 10.1073/pnas.1915006117.
- Lundberg, Ian, Jennie E. Brand, and Nanum Jeon. 2022. 'Researcher Reasoning Meets Computational Capacity: Machine Learning for Social Science'. *Social Science Research* 108:102807. doi: 10.1016/j.ssresearch.2022.102807.
- Lundberg, Ian, Rachel Brown-Weinstock, Susan Clampet-Lundquist, Sarah Pachman, Timothy J. Nelson, Vicki Yang, Kathryn Edin, and Matthew J. Salganik. 2024. 'The Origins of Unpredictability in Life Outcome Prediction Tasks'. *Proceedings of the National Academy of Sciences* 121(24):e2322973121. doi: 10.1073/pnas.2322973121.
- * Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. 'Predicting Poverty and Wealth from Mobile Phone Metadata'. *Science* 350(6264):1073–76. doi: 10.1126/science.aac4420.
- Chi, Guanghua, Han Fang, Sourav Chatterjee, and Joshua E. Blumenstock. 2022. 'Microestimates of Wealth for All Low- and Middle-Income Countries'. *Proceedings of the National Academy of Sciences* 119(3):e2113658119. doi: 10.1073/pnas.2113658119.
- Molina, Mario, and Filiz Garip. 2019. 'Machine Learning for Sociology'. *Annual Review of Sociology* 45(1):27–45. doi: 10.1146/annurev-soc-073117-041106.
- * Hofman, Jake M., Duncan J. Watts, Susan Athey, Filiz Garip, Thomas L. Griffiths, Jon Kleinberg, Helen Margetts, Sendhil Mullainathan, Matthew J. Salganik, Simine Vazire, Alessandro Vespignani, and Tal Yarkoni. 2021. 'Integrating Explanation and Prediction in Computational Social Science'. *Nature* 595(7866):181–88. doi: [10.1038/s41586-021s-03659-0](10.1038/s41586-021s-03659-0).

Integrating explanation and prediction in computational social science | Nature

Week 4: Machine learning [LAB 2]

Apply machine learning algorithms and Facebook user counts to predict overall levels of internet adoption and digital divisions

- Breen, Casey F., Masoomali Fatehkia, Jiani Yan, Xinyi Zhao, Douglas R. Leasure, Ingmar Weber, and Ridhi Kashyap. 2024. 'Mapping Subnational Gender Gaps in Internet and Mobile Adoption Using Social Media Data'.

Week 5: Non-probability sampling - discussion

- * Stedman, Richard C., Nancy A. Connelly, Thomas A. Heberlein, Daniel J. Decker, and Shorna B. Allred. 2019. 'The End of the (Research) World As We Know It? Understanding and Coping With Declining Response Rates to Mail Surveys'. *Society & Natural Resources* 32(10):1139–54. doi: [10.1080/08941920.2019.1587127](#).
- * Lehdonvirta, Vili, Atte Oksanen, Pekka Räsänen, and Grant Blank. 2021. 'Social Media, Web, and Panel Surveys: Using Non-Probability Samples in Social and Policy Research'. *Policy & Internet* 13(1):134–55. doi: [10.1002/poi3.238](#).
- Elliott, Michael R., and Richard Valliant. 2017. 'Inference for Nonprobability Samples'. *Statistical Science* 32(2):249–64.
- * Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. 'Forecasting Elections with Non-Representative Polls'. *International Journal of Forecasting* 31(3):980–91. doi: [10.1016/j.ijforecast.2014.06.001](#).
- Dutwin, David, and Trent D. Buskirk. 2017. 'Apples to Oranges or Gala versus Golden Delicious?: Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples'. *Public Opinion Quarterly* 81(S1):213–39. doi: [10.1093/poq/nfw061](#).
- Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. 'Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls'. Political Analysis 12(4):375–85. doi: 10.1093/pan/mph024.
- Breen, Casey F., Ayesha S. Mahmud, and Dennis M. Feehan. 2022. 'Novel Estimates Reveal Subnational Heterogeneities in Disease-Relevant Contact Patterns in the United States'. *PLOS Computational Biology* 18(12):e1010742. doi: [10.1371/journal.pcbi.1010742](#).

Week 6: Social networks - discussion

- Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. 'Network Analysis in the Social Sciences'. *Science* 323(5916):892–95. doi: [10.1126/science.1165821](#).
- * Watts, Duncan J. 2004. 'The "New" Science of Networks'. *Annual Review of Sociology* 30(1):243–70. doi: [10.1146/annurev.soc.30.020404.104342](#).
- Eagle, Nathan, Alex (Sandy) Pentland, and Dav[The "New" Science of Networks | Annual Reviews](#)id Lazer. 2009. 'Inferring Friendship Network Structure by Using Mobile Phone

Data'. *Proceedings of the National Academy of Sciences* 106(36):15274–78. doi: 10.1073/pnas.0900282106.

- Kossinets, Gueorgi, and Duncan J. Watts. 2006. 'Empirical Analysis of an Evolving Social Network'. *Science* 311(5757):88–90. doi: 10.1126/science.1116869.
- Feehan, Dennis M., and Curtiss Cobb. 2019. 'Using an Online Sample to Estimate the Size of an Offline Population'. *Demography* 56(6):2377–92. doi: 10.1007/s13524-019-00840-z.
- * Granovetter, Mark S. 1973. 'The Strength of Weak Ties'. *American Journal of Sociology* 78(6):1360–80.
- * Salganik, Matthew J., and Douglas D. Heckathorn. 2004. 'Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling. *Sociological Methodology* 34(1):193–240.
- Goel, Sharad, and Matthew J. Salganik. 2010. 'Assessing Respondent-Driven Sampling'. *Proceedings of the National Academy of Sciences* 107(15):6743–47. doi: 10.1073/pnas.1000261107.
- * Centola, Damon, and Michael Macy. 2007. 'Complex Contagions and the Weakness of Long Ties'. *American Journal of Sociology* 113(3):702–34. doi: 10.1086/521848.

Week 7: Non-probability sampling [LAB]

Apply poststratification and inverse-probability weighting to a non-probability sample to estimate mortality rates using network survival method

- Baker, Reg, J. Michael Brick, Nancy A. Bates, Mike Battaglia, Mick P. Couper, Jill A. Dever, Krista J. Gile, and Roger Tourangeau. 2013. 'Summary Report of the AAPOR Task Force on Non-Probability Sampling'. Journal of Survey Statistics and Methodology 1(2):90–143. doi: 10.1093/jssam/smt008.

Week 8: Agent-based modelling and microsimulation - discussion
- * Bruch, Elizabeth, and Jon Atwell. "Agent-based models in empirical social research." *Sociological methods & research* 44, no. 2 (2015): 186-221.
- Spielauer, Martin. "What is social science microsimulation?" *Social Science Computer Review* 29, no. 1 (2011): 9-20.
- Zagheni, Emilio. "The impact of the HIV/AIDS epidemic on kinship resources for orphans in Zimbabwe." *Population and Development Review* 37, no. 4 (2011): 761-783.

- * Kashyap, Ridhi, and Francisco Villavicencio. "The dynamics of son preference, technology diffusion, and fertility decline underlying distorted sex ratios at birth: A simulation approach." *Demography* 53, no. 5 (2016): 1261-1281.
- Grow, André, and Jan Van Bavel. "Assortative mating and the reversal of gender inequality in education in Europe: An agent-based model." *PloS one* 10, no. 6 (2015): e0127806.
- *Alburez‑Gutierrez, Diego, Carl Mason, and Emilio Zagheni. "The "sandwich generation" revisited: Global demographic drivers of care time demands." *Population and Development Review* 47, no. 4 (2021): 997-1023.