Syllabus

Syllabus: Computational Methods

for Social and Political Science,

Bocconi, Spring 2025

Recommended Reading:

1. Christian, B., & Griffiths, T. (2016). *Algorithms to live by: The computer science of human decisions*. Macmillan.
2. O'Neil, C. (2016). Weapons of Math Destruction. Penguin.
3. Hovy, D. (2022). *Lecture Notes* (http://dirkhovy.com/portfolio/papers/download/nlpss.pdf). Bocconi University.

Recommended data:

- Inside AirBnB data set (https://insideairbnb.com/)

Week 1: Introduction to Computational Social Science

Topics: Introductions and course overview. What is Computational Social Science? Good computing practices. Processing, cleaning, and visualizing data with Pandas.

- Slides (https://drive.google.com/file/d/1JBGGuWt-xBMXrZ6OpC_ICTsVW3ih_xnQ/view?usp=drive_link)
- Background notebook (https://colab.research.google.com/drive/1KfI2Ly4HDKXPYZaVN6lOXmpb6lav32I7?usp=sharing)
- Practice set (https://colab.research.google.com/drive/1DzDrkZ6vpKys9WSPn19tYTQAcgu9tamy?usp=sharing)
- Practice Set SOLUTIONS (https://colab.research.google.com/drive/1ilc2K-mx67MRVwG1GNReOEpULcSN1VM2?usp=sharing)

Background reading:

1. Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Jebara, T. (2009). Life in the network: the coming age of computational social science

(https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2745217/). *Science*.
2. Edelman, A., Wolff, T., Montagne, D. & Bail, C. A. (2020). Computational Social Science and Sociology (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8612450/). *Annual Review of Sociology.*
3. Lazer, D. & and Radford, J. (2017). Data ex Machina: Introduction to Big Data (https://www.annualreviews.org/doi/abs/10.1146/annurev-soc-060116-053457). *Annual Review of Sociology*.
4. Anderson, C. (2008). The End of Theory. The Data Deluge Makes the Scientific Method Obsolete. (https://www.wired.com/2008/06/pb-theory/) WIRED.
5. Salganik, M. (2017). Bit by Bit: Social Research in the Digital Age. Chapters 1 (https://www.bitbybitbook.com/en/1st-ed/introduction/), 2 (https://www.bitbybitbook.com/en/1st-ed/observing-behavior/).
6. Watts, D. (2007). A twenty-first century science (https://www.nature.com/articles/445489a). *Nature*.
7. Bail, C. A. (2024). Can Generative AI improve social science? (https://www.pnas.org/doi/10.1073/pnas.2314021121) *Proceedings of the National Academy of Sciences*, *121*(21), e2314021121.

Week 2: Data Collection and Processing

Topics: Collecting and analyzing social media data. The ethics of collecting online data. First steps in machine learning with sklearn. Web scraping and working with APIs.

- Slides (https://drive.google.com/file/d/1WE4sySUCia18H0uLSYLzbVPTJF-nEF63/view?usp=drive_link)
- Background notebook web scraping (https://colab.research.google.com/drive/1CfMmLMowHVod6ulKMt7HI50rspl3PVQD?usp=sharing)
- Background notebook (https://colab.research.google.com/drive/1y2R6ZHcEcyr7OSsaB3zO9JWetAdCtEJl?usp=sharing) ML
- Practice Set (https://colab.research.google.com/drive/1IDmp9_IAURt-LwMND0vHFOuS2NeNw-kK?usp=sharing) ML
- Practice Set ML SOLUTIONS (https://colab.research.google.com/drive/1mbDZt1vT3LrKQcSU5BzyCOrdS2qylC5-?usp=sharing)

Background reading:

1. Bail, C. A. (2014). The cultural environment: Measuring culture with big data (https://link.springer.com/article/10.1007/s11186-014-9216-5). *Theory and Society*
2. Freelon, D. (2018). Computational research in the post-API age (https://www.tandfonline.com/doi/full/10.1080/10584609.2018.1477506). *Political Communication*
3. Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems (https://www.science.org/doi/10.1126/science.aal3856). *Science*.
4. Shmueli, G. (2010). To Explain or to Predict? (https://www.stat.berkeley.edu/~aldous/157/Papers/shmueli.pdf) Statistical Science, Vol. 25, No. 3, 289–310.
5. Salganik, M. (2017). Bit by Bit: Social Research in the Digital Age. Chapter 6. (https://www.bitbybitbook.com/en/1st-ed/ethics/)
6. Schneider, D., & Harknett, K. (2019). What's to like? Facebook as a tool for survey data collection (https://journals.sagepub.com/doi/full/10.1177/0049124119882477). *Sociological Methods &*

Research

Week 3: Text as Data

Topics: Word and document embeddings, text classification, Topic Models.

- Slides (https://drive.google.com/open?id=1wcg00tgZMGMlihFq_rkG8vT37oIuuy0o)
- Background notebook (https://colab.research.google.com/drive/1dTpRfmAuEQ1vzw3EwFEKZrQoKQlytuNO?usp=sharing)
- Practice notebook (https://colab.research.google.com/drive/11a5LXwxArtKEoCvkIPxSIklJyJED3IC7?usp=sharing)
- Practice notebook SOLUTIONS (https://colab.research.google.com/drive/1CT6xLYWuNyzFes5hiuGeY3YK_y3TJvXv?usp=sharing)

Background reading:

1. Lecture Notes (http://www.dirkhovy.com/portfolio/papers/download/nlpss.pdf), chapter I.2, chapter I.3, chapter II.4
2. Jurafsky and Martin, Chapter 6 (https://web.stanford.edu/~jurafsky/slp3/6.pdf)
3. Blei et al. (2003) "Latent Dirichlet Allocation" (https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf)
4. Bianchi et al. (ACL 2021) "Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence" (https://aclanthology.org/2021.acl-short.96.pdf)
5. Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). (https://www.pnas.org/content/pnas/115/16/E3635.full.pdf)Word embeddings quantify 100 years of gender and ethnic stereotypes. (https://www.pnas.org/content/pnas/115/16/E3635.full.pdf) *Proceedings of the National Academy of Sciences*, *115*(16), E3635-E3644.
6. Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). (https://journals.sagepub.com/doi/pdf/10.1177/0003122419877135)The geometry of culture: Analyzing the meanings of class through word embeddings. (https://journals.sagepub.com/doi/pdf/10.1177/0003122419877135) *American Sociological Review*, *84*(5), 905-949.
7. Bhatia, S. (2017). (https://www.sas.upenn.edu/~bhatiasu/Bhatia%202017%20PsychReview.pdf)Associative judgment and vector space semantics. (https://www.sas.upenn.edu/~bhatiasu/Bhatia%202017%20PsychReview.pdf) *Psychological Review, 124*(1), 1–20.

Week 4: Language Models

Topics: The Transformer architecture, language models. Use in surveys and online experiments, also via agent-based models.

- Slides (https://drive.google.com/file/d/1wX5vkHQDIUHV5j-2cAeHV8grzgJkUEYa/view?usp=drive_link)
- Background notebook (https://colab.research.google.com/drive/1n7_UCldzNvAJml76noBV105GFNiS3GKi?usp=sharing)

- Practice notebook (https://colab.research.google.com/drive/1HfyA-XgvNkLWt0gGiaxvdSqvaSPr5n1v?usp=sharing)
- Practice notebook SOLUTIONS (https://colab.research.google.com/drive/1VKY1Iadfi40ZQxaCjPnk5y6R6xLvQkvj?usp=sharing)

Background reading:

1. Lecture Notes (http://www.dirkhovy.com/portfolio/papers/download/nlpss.pdf), chapters III.14.3, III.14.4 and III.14.5
2. Jurafsky and Martin, Chapter 10 (https://web.stanford.edu/~jurafsky/slp3/10.pdf), Chapter 11 (https://web.stanford.edu/~jurafsky/slp3/11.pdf)
3. Vaswani et al (NeurIPS 2017): "Attention is all you need" (https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)
4. Bainbridge, W. S. (2007). The scientific research potential of virtual worlds (https://www.science.org/doi/10.1126/science.1146930). *Science*.
5. Salganik, M. J., Dodds, P. S. & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. (https://www.princeton.edu/~mjs3/salganik_dodds_watts06_full.pdf) *Science*.
6. Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions (https://www.nber.org/papers/w23180). *The quarterly journal of economics*.
7. Manning, B. S., Zhu, K., & Horton, J. J. (2024). Automated social science: Language models as scientist and subjects (https://www.nber.org/system/files/working_papers/w32381/w32381.pdf). National Bureau of Economic Research.
8. Aher, G. V., Arriaga, R. I., & Kalai, A. T. (2023, July). Using large language models to simulate multiple humans and replicate human subject studies. (https://proceedings.mlr.press/v202/aher23a/aher23a.pdf) In: International Conference on Machine Learning (pp. 337-371). PMLR.
9. Horton, J. J. (2023). *Large language models as simulated economic agents: What can we learn from homo silicus?* (https://www.nber.org/papers/w31122) (No. w31122). National Bureau of Economic Research.
10. Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. (https://www.cambridge.org/core/journals/political-analysis/article/out-of-one-many-using-language-models-to-simulate-human-samples/035D7C8A55B237942FB6DBAD7CAA4E49) *Political Analysis*, *31*(3), 337-351.
11. Kozlowski, A. C., & Evans, J. A. Simulating Subjects: The Promise and Peril of AI Stand-ins for Social Agents and Interactions. (https://osf.io/vp3j2/download/?format=pdf)
12. Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2024). Can large language models transform computational social science? (https://aclanthology.org/2024.cl-1.8) Computational Linguistics, 50(1), 237-291.
13. Park, J. S., Popowski, L., Cai, C., Morris, M. R., Liang, P., & Bernstein, M. S. (2022, October). Social simulacra: Creating populated prototypes for social computing systems. (https://dl.acm.org/doi/10.1145/3526113.3545616) In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (pp. 1-18).
14. Salganik, M. (2017). Bit by Bit: Social Research in the Digital Age. Chapter 4. (https://www.bitbybitbook.com/en/1st-ed/running-experiments/)

Week 5: Networks

Topics: Networks, graphs, and the small-world theory: modeling the spread of ideas, diseases, and the web. Measuring polarization and echo chambers. Social contagion processes.

- Slides (https://drive.google.com/file/d/1I84WWcRJC4W2iOZxHc8ecXG9aMfEsp-e/view?usp=drive_link)
- Background notebook (https://colab.research.google.com/drive/1ftArxVdVSdWUhvh7GG0Lld95J5Q8Uu0C?usp=sharing)
- Practice notebook (https://colab.research.google.com/drive/1drC5sfyBejkg1Nyc3-_sndn7DKIUvSkM?usp=sharing)
- Practice notebook SOLUTIONS (https://colab.research.google.com/drive/1nU3RfxUBbXW3Br-w92hG-r3eTIL35yv5?usp=sharing)

Background reading:

1. S. Milgram (1967) "The small world problem (http://snap.stanford.edu/class/cs224w-readings/milgram67smallworld.pdf)," Psychology Today.
2. Watts, D. (2016). "How small is the world, really? (https://medium.com/@duncanjwatts/how-small-is-the-world-really-736fa21808ba)" Blog post.
3. Bail, C. A., Guay, B., Maloney, E., Combs, A., Hillygus, D. S., Merhout, F., ... & Volfovsky, A. (2020). Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017 (https://www.pnas.org/doi/10.1073/pnas.1906420116). *Proceedings of the national academy of sciences*.
4. Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? (https://journals.sagepub.com/doi/abs/10.1177/0956797615594620) *Psychological science*.
5. Sekara, V., Stopczynski, A., & Lehmann, S. (2016). Fundamental structures of dynamic social networks (https://www.pnas.org/doi/full/10.1073/pnas.1602803113). *Proceedings of the national academy of sciences.*
6. Johnson, N. F., Velásquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., ... & Lupu, Y. (2020). The online competition between pro-and anti-vaccination views (https://www.nature.com/articles/s41586-020-2281-1). *Nature*.
7. Johnson, N. F., Leahy, R., Restrepo, N. J., Velasquez, N., Zheng, M., Manrique, P., ... & Wuchty, S. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology (https://www.nature.com/articles/s41586-019-1494-7). *Nature*.
8. Ananthasubramaniam, A., Jurgens, D., & Romero, D. M. (2024). Networks and identity drive the spatial diffusion of linguistic innovation in urban and rural areas. (https://www.nature.com/articles/s44260-024-00009-9) *npj Complexity*, *1*(1), 14.

Week 6: Maps

Topics: Working with geolocated data and maps. Python libraries and data formats.

- Slides (https://drive.google.com/file/d/13hkCCapLFrAUlSkK1pZf4nRLSCbj4_Pd/view?usp=sharing)
- Background notebook (https://colab.research.google.com/drive/1a5hyVNMEOA6fHdTDGEdFTcykxajo9OJT?usp=sharing)

- Practice notebook (https://colab.research.google.com/drive/1uPUxmlf8lSCU5Gi9Y6GM587K0Rkpsl98?usp=sharing)
- Practice notebook SOLUTIONS (https://colab.research.google.com/drive/1lLKKxPKV8AK06dL8Ch5HB5i0ed7ilUt5?usp=sharing)

Background reading:

1. Schneider, T. (2015). "Analyzing 1.1 Billion NYC Taxi and Uber Trips, with a Vengeance (http://toddwschneider.com/posts/analyzing-1-1-billion-nyc-taxi-and-uber-trips-with-a-vengeance)." toddwschneider.com. Nov 17th. Retrieved September 29, 2018.
2. Purschke, C., & Hovy, D. (2019). (http://www.dirkhovy.com/portfolio/papers/download/loerres_draft.pdf)Lörres, Möppes, and the Swiss. (Re)Discovering regional patterns in anonymous social media data. (http://www.dirkhovy.com/portfolio/papers/download/loerres_draft.pdf) *Journal of Linguistic Geography, 7*(2), 113-134. doi:10.1017/jlg.2019.10
3. Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. (https://www.science.org/doi/pdf/10.1126/science.aaf7894) *Science*, *353*(6301), 790-794.
4. Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., & Fei-Fei, L. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States (https://www.pnas.org/doi/pdf/10.1073/pnas.1700035114). *Proceedings of the National Academy of Sciences*, *114*(50), 13108-13113.
5. Khachiyan, A., Thomas, A., Zhou, H., Hanson, G., Cloninger, A., Rosing, T., & Khandelwal, A. K. (2022). Using neural networks to predict microspatial economic growth.