40989 – Computational Social Science – a.y. 2020-21

Mission & Content summary

A growing proportion of people's daily lives is spent online. These activities leave behind granular, time-stamped digital footprints of human behaviour and interactions. Improvements in data storage and the development of computational techniques allow to leverage this unprecedented amount of data to address long-standing social scientific questions. At the same time, as more and more of the human experience is shaped by information system, platforms and algorithms, novel opportunities for understanding arise, as well as relevant social and ethical challenges. The goal of 40989 – CSS is to a) equip students with concepts and techniques that are needed to access and analyse new data; b) formulate frontier research questions and answer them with techniques learned in the course; c) develop a critical understanding of the social and ethical dilemmas that characterise digital societies and research.

Instructors

The instructor for 40989 – CSS are Nicolò Cavalli (nicolo.cavalli@unibocconi.it) and Dirk Hovy (dirk.hovy@unibocconi.it).

Teaching methods

The course will follow a 'learning-by-doing' approach that employs a 'flipped classroom' model to place emphasis on gaining hands-on experience in data collection and analysis. Students are expected to do the required readings and coding exercises for each week. The lectures will build upon the content of the readings with a series of data challenges that will introduce new statistical and programming concepts, which will then be applied to the analysis of data from published research papers or common tasks in computational social science. Loosely speaking, the structure of each session is composed of three parts: a class discussion of reading materials and assignments; an introduction, led by the instructor, on how these methods are embedded in the broader computational social science literature; a practical component that will demonstrate how to use specific tools. The first block will be based on a set of readings from textbooks and introductory manuals. All students in the class are expected to at least have skimmed the papers and be ready to ask questions or engage in the discussion. This part will assume that all students have done the readings and thus we will move quickly and focus on the most complex aspects, how these methods are embedded int he broader computational social science literature, and strengths and weaknesses that perhaps were not obvious from the readings (second part). The third part of each class will be applied. The instructor will provide code and data that illustrate the application of a specific computational social science method, followed by a "coding challenge" where students work together to apply what they have learned to a new dataset. The code and output (log file) of their coding challenge needs to be submitted before the beginning of the following class.

Computing environment

40989 - CSS promotes a flexible, 'problem-based' approached to the choice of the computing environment. Some activities will require coding in either R or Python. Students are welcome to use other languages that are commonly used by computational social scientists that may allow them to efficiently solve problems, but we cannot guarantee that we can support them. Before the start of the first class, students are required to install a modern, stable-release version of R and RStudio, as well as Anaconda with Python 3.8 and make sure it is running.

Suggested background knowledge

There is no compulsory pre-requisite for this course. Knowledge of programming in R and Python is welcome. To refresh R, we recommend the free RStudio Primers, which can be supplemented by the open access book R for Data Science by Garrett Grolemund and Hadley Wickham. RStudio Primers cover 6 topics: The Basics, Working with Data, Visualize Data, Tidy Your Data, Iterate, and Write Functions. The course will also assume some basic familiarity with Python. If you are looking to learn Python, or to refresh the basics, the official Python website lists a number of excellent online tutorials and textbooks. Students who already feel comfortable with these topics do not need to complete these primers.

Assessment

Class participation (10%)
Class participation is not compulsory. Students are however expected to attend every session and do the assigned readings before each session. You should come to class with questions and ready to engage in a discussion about that week's topic.

Coding challenges (30%)
At the end of each class, the instructor will share a short problem set based on the content of that class. There will be some time during class to start working on this exercise, and to address any question, but most of it should be completed after class. Each of these coding challenges will be due before the beginning of the following class, when the solutions (or a sketch of the solution) will be posted. A complete submission includes both the R or Python script with the code written by the student (and comments with interpretation of the results, answering the questions) as well as a log/html file showing the output of compiling the code. Submission will not be graded but submitting at least three of them is compulsory.

Research project (60%)
Students are required to submit an original research paper (10 pages, excluding references) that employs any of the methods introduced in the course. The goal of this exercise is to demonstrate that you have the ability to conduct research in computational social science. This research paper can be an individual or group project (up to 3 people).

Syllabus

Week 1 – Introduction to Computational Social Science (N. Cavalli)

Topics: Introductions and course overview. What is Computational Social Science? Good computing practices. Introduction to version control and GitHub.

Discussion readings:

- Edelman, A., Wolff, T., Montagne, D. & Bail, C. A. (2020). Computational Social Science and Sociology, *Annual Review of Sociology.*
- Lazer, D. & and Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology*.
- 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. *Science*.
- Salganik, M. (2017). Bit by Bit: Social Research in the Digital Age. Chapters 1, 2.

- Watts, D. (2007). A twenty-first century science. *Nature*.

Week 2 – Automated Collection of Web and Social Data (N. Cavalli)

Topics: Working with APIs. Collecting and analysing social media data. The ethics of collecting online data. Platforms and challenges of a post-API world. Scraping social media data in table format and unstructured format.

Discussion readings:

- Bail, C. A. (2014). The cultural environment: Measuring culture with big data. *Theory and Society*
- Freelon, D. (2018). Computational research in the post-API age. *Political Communication*
- Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*.
- Salganik, M. (2017). Bit by Bit: Social Research in the Digital Age. Chapter 6.
- Schneider, D., & Harknett, K. (2019). What's to like? Facebook as a tool for survey data collection. *Sociological Methods & Research*

Week 3 – Data visualization, working with images and maps (D. Hovy)

Topics: Dimensionality reduction. Images as matrices. Visualizing data: geomapped data.

Background readings:

- Hovy, D. (2020). Text Analysis in Python for Social Scientists – Discovery and Exploration. Cambridge University Press. *Relevant chapters provided on BlackBoard.*
- Hovy, D., & Purschke, C. (2018). Capturing regional variation with distributed place representations and geographic retrofitting. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4383-4394).

Week 4 – Advanced Text Analysis (D. Hovy)

Topics: Statistical principles in machine learning. Algorithmic bias and challenges.
Ethics of working with large data. Representing and analysing text corpora with embeddings. Sequential models. Natural language processing with neural language models.

Background readings:

- Hovy, D. (2020). Text Analysis in Python for Social Scientists – Discovery and Exploration. Cambridge University Press. *Relevant chapters provided on BlackBoard.*
- Grimmer, J. (2015). We are all social scientists now: How big data, machine learning, and causal inference work together. *PS, Political Science & Politics*, *48*(1), 80.
- Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*(6213), 1063-1064.
- Shmueli, G. (2010). To explain or to predict? *Statistical science*, *25*(3), 289-310.
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100-1122.

- Rheault, L., & Cochrane, C. (2019). Word embeddings for the analysis of ideological placement in parliamentary corpora. *Political Analysis*, *27*.
- Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635-E3644.

Week 5 – Social networks, echo chambers and polarization (D. Hovy & N. Cavalli)

Topics: Network analysis. Measuring polarization and echo chambers. Social contagion processes.

Background readings:

- 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. *Proceedings of the national academy of sciences*.
- 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?. *Psychological science*.
- Sekara, V., Stopczynski, A., & Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proceedings of the national academy of sciences.*
- 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. *Nature*.
- Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, *39*(3), 629-649.
- Ugander, J., Backstrom, L., Marlow, C., & Kleinberg, J. (2012). Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, *109*(16), 5962-5966.
- 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. *Nature*.

Week 6 – Online experiments and mass collaboration (N. Cavalli)

Topics: Online experiments. Mass collaboration. Statistical issues in computational social sciences. Algorithmic bias and challenges.

Background readings:

- Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science*.
- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The quarterly journal of economics*.
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*.
- Salganik, M. J., Dodds, P. S. & Watts, D. J. (2006). "Experimental study of inequality and unpredictability in an artificial cultural market." *Science*.
- Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., ... & McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences.*

Honor code

Bocconi University conceives of education as an ongoing process that stretches across a person's entire professional life. The University hopes that the entire Bocconi community will respect the values of fairness and correctness associated with it, values which inspire and guide the conduct of all community members as they pursue common objectives and a shared mission. The Bocconi University Honor Code is published at http://www.unibocconi.eu/honorcode. We encourage all students to read it.