SOSC 4300 / SOSC 5500: Computational Social Science

Fall 2020

Lecture Time: Tuesday 9:00 - 11:50AM
Tutorial Time: Friday 6:00PM - 6:50PM

Tentative, this version prepared on August 30, 2020

	Instructor	Teaching Assistant
	ZHANG, Han	**Peng, Wenwei**
Office	Room 2379, Academic Building, Lift 15	Room 3001, Academic Building, Lift 4
Email	zhangh@ust.hk	wshaaa@connect.ust.hk
Office Hour	Tuesday 2:00PM to 3:00PM	TBD
Zoom for office hour		

Prerequisites

- Students are expected to be familiar with the materials covered in basic statistics (e.g., SOSC 2400 for UG students and SOSC 5090 for PG students). Students with statistics knowledge but do not meet prerequisite can **seek instructor's approval for enrollment**.
- Students should also have basic literacy in at least one statistical programming language. We will use R in tutorials. You can also use other programming languages such as Python, Matlab, Julia, etc., as long as you can finish course assignment and projects with the codes.

Goals

Upon finishing the course, students should be able to:

1. Describe the opportunities and challenges of social research in the age of big data
2. Evaluate research on social phenomena from different fields, including social sciences and computer science/data science.
3. Practice the essential techniques to analyze social big data
4. Propose research questions that are suited to be examined by computational methods with big data
5. (For PG students): write a research article that utilizes the techniques and methods of computational social sciences to address social science problems.

Grading

Your score will be accessed based on the following five components (no mid-term and final exams):

	UG students	PG students	Due
Attendence and participation in class activities	10%	10%	
Coding exercises and short written assignments	20% (independent)	20% (independent)	Two weeks
Literature review			
Report	15% (3-4 people)	20% (3-4 people)	TBD
Presentation	5% (3-4 people)	10% (1-2 people)	TBD
Final Paper/Project			
Presentation	10% (3-4 people))	10% ((3-4 people))	Dec 1
Write-up	40% (3-4 people)	40% (1-2 people)	Dec 15

Attendance and participation in class activities

- Based on class attendance and involvement in lecture and tutorial (e.g., ask questions during lectures and come to office hours).

Coding exercises and short written assignments

- These exercises test your knowledge of analyzing data using statistical software. Each exercise is due in **two weeks** after the release of assignment.

Literature review

Select a substantive research topic and summarize how the literature has used computational methods and/or big data are being used to study this particular research area.

- Some examples of research areas:
  - Sociological topics: internal migration, international migration, social inequality, race and ethnicity relations, happiness,
  - Political science topics: government performance, government policy (and its effectiveness), election, social movements
  - Economics: measuring economic growth with big data
  - History: anything related to historical phenomena
  - Psychology: measuring personality with big data
  - Communication and information science: fake news
- It is better for you to select a research areas that are similar to your final research paper. Students can discuss with instructors and TA for possible topics or feasibility.
- UG students should form groups of 3 or 4 participants. PG students can work individually or work with another PG student.
- Your performances will be accessed in two ways
  - **Report**: each literature review report should contain at least **10 pages, 12 points, double space**.
  - **Presentation of literature review** (**20 minutes**): each student/group needs to present their literature reviews in class. Use the follow items for templates:
    - What is the research area you have chosen, and why it's important
    - How people studied it traditionally (e.g., what data they use, what methods they use)
    - What are the advantages of using computational social science?
    - Some good examples
    - What are shortcomings of using computational social science methods to study this phenomena?

Final paper/project

Each student/group needs to choose a research topic and write a research paper (a conventional assignment) **using computational social science methods or digital data**. This research article should follow the format of a standard research article, with the following components: introduction; review of past studies; research methods and data; results; conclusions. UG students can form groups up to 4 people. PG students can work individually or work with another PG student.

- Presentation (**20 minutes**):
- Final paper/project: **20 pages, 12 points, double space, including Tables, Figures and References**.

Alternatively, you can choose a project by building a demo/website/app that are useful for communicating some fancy social science data to laymen. Image your audience as people with no academic background, and you need to persuade them your stories with visualization and explanations. Some ideas of cool demo/projects can be found here: `https://projects.fivethirtyeight.com/`. If you choose this option, please discuss your ideas with the instructor at early stage. Your performance will be evaluated based on the following criteria:

- Presentation (**15 minutes**). Show case your project in front of the class. Focusing on whether your topic is interesting enough to attract your classmates' interest.
- Final project: submit the project as a public accessible website/app that allow others (especially outside of the class) to access your output.

Grading policies

- **Late delivery** of due items will be marked down 75% if received within 1 day of the due date, and 50% if received within 3 days of the due date; you will receive zero credit if the due item is not delivered within 3 days of the due date. Contact the instructor if there are rare unforeseen circumstances.
- If you want to dispute a grade, please submit your argument in writing along with your assignment. We will evaluate the merit of your argument as well as perform a full reassessment of your entire assignment. This means that your grade may end up lower than it was originally.
- Final papers are checked by anti-plagiarism software. Students should take steps to avoid plagiarism and copying. For confirmed cases of plagiarism, severe sanctions including but not limited to a failure grade may be imposed.

Course Outline (Tentative)

The course materials will be drawn from lecture slides and assigned readings. Readings are available at Canvas. You are **required to read the readings before the start of each class (except the first class)**. Optional readings are for students who are interested to read more on the topic.

Week 1 (Sep 8) : Digital Traces and Ethics

- Salganik, M. (2019). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Chapter 1 and 2. This book can be freely accessible at `https://www.bitbybitbook.com/en/1st-ed/preface/`
- Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science*, 369(6507), 1060–1062
- Salganik, M. (2019). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Chapter 6.
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), 8788–8790

Optional readings

- Lazer, D. & Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology*, 43(1), 19–39
- Golder, S. A. & Macy, M. W. (2014). Digital Footprints: Opportunities and Challenges for Online Social Research. *Annual Review of Sociology*, 40(1), 129–152

Week 2 (Sep 15) Prediction vs. Explanation

- Hofman, J. M., Sharma, A., & Watts, D. J. (2017). Prediction and explanation in social systems. *Science*, 355(6324), 486–488
- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, 133(1), 237–293

Week 3 (Sep 22) : Survey

- Salganik, M. (2019). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Chapter 3.
- Beauchamp, N. (2017). Predicting and Interpolating State-Level Polls Using Twitter Textual Data. *American Journal of Political Science*, 61(2), 490–503

Optional readings:

- Salganik, M. J. & Levy, K. E. C. (2015). Wiki Surveys: Open and Quantifiable Social Data Collection. *PLOS ONE*, 10(5), e0123483
- Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The Generalizability of Survey Experiments. *Journal of Experimental Political Science*, 2(02), 109–138

Week 4 (Sep 29) : Text (I)

- Grimmer, J. & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(03), 267–297
- Wilkerson, J. & Casas, A. (2017). Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges. *Annual Review of Political Science*, 20(1), 529–544
- Denny, M. J. & Spirling, A. (2018). Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*, 26(2), 168–189

Optional readings:

- Benoit, Kenneth (2020). Text as Data : An Overview. In L. Curini & Franzese, Robert (Eds.), *The SAGE Handbook of Research Methods in Political Science and International Relations*. SAGE Publications Ltd
- Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, 57(3), 535–574
- Goldberg, Y. (2015). A Primer on Neural Network Models for Natural Language Processing. *arXiv:1510.00726 [cs]*

Week 5 (Oct 6) : Text (II); Dictionary and Supervised

- Benoit, K., Conway, D., Lauderdale, B. E., Laver, M., & Mikhaylov, S. (2016). Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data. *American Political Science Review*, 110(2), 278–295
- Barberá, P., Boydstun, A. E., Linn, S., McMahon, R., & Nagler, J. (2020). Automated Text Classification of News Articles: A Practical Guide. *Political Analysis*, (pp. 1–24)

Week 6 (Oct 13) : Text (III); Unsupervised And Word Embeddings

- Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., & Tucker, J. A. (2019). Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data. *American Political Science Review*, 113(4), 883–901
- 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
- Klingenstein, S., Hitchcock, T., & DeDeo, S. (2014). The civilizing process in London's Old Bailey. *Proceedings of the National Academy of Sciences*, 111(26), 9419–9424

Optional readings

- Blei, D. M. (2012). Probabilistic Topic Models. *Commun. ACM*, 55(4), 77–84
- Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2014). *Structural Topic Models*. Technical report, Working Paper. Export BibTex Tagged XML
- Blaydes, L., Grimmer, J., & McQueen, A. (2018). Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds. *The Journal of Politics*, 80(4), 1150–1167

Week 7 (Oct 20) : Images and Spatial Data

- Zhang, H. & Pan, J. (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
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790–794
- 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. *Proceedings of the National Academy of Sciences*, 114(50), 13108–13113

Week 8 (Oct 27) : Network (I); small world

- Watts, D. J. (1999). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Alstyne, M. V. (2009). Computational Social Science. *Science*, 323(5915), 721–723

Week 9 (Nov 3) : Network (II); social influence vs. homophily

- Christakis, N. A. & Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine*, 357(4), 370–379
- Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12 (pp. 519–528). New York, NY, USA: ACM
- Lewis, K., Gonzalez, M., & Kaufman, J. (2012). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1), 68–72

Week 10 (Nov 10): Network (III): simulations and tipping points

- Schelling, T. C. (2006). *Micromotives and Macrobehavior*. W. W. Norton & Company

Week 11 (Nov 17): Causal inference: online field experiment

- Salganik, M. (2019). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press. Chapter 4.
- Chen, Y. & Konstan, J. (2015). Online field experiments: A selective survey of methods. *Journal of the Economic Science Association*, 1(1), 29–42

Week 12 (Nov 24): Warp ups; Presentations

Week 13 (Dec 1) : Presentation of final paper and projects

Tutorial Schedule (Tentative)

Week	Topic
1 (Sep 8)	Reproducible Research; Git and Project Management
2 (Sep 15)	Reproducible Research; literacy programming
3 (Sep 22)	Survey: offline and online
4 (Sep 29)	Text data: basic operations
5 (Oct 6)	Text data: dictionary counts
6 (Oct 13)	Text data: supervised learning
7 (Oct 20)	Text data: topic models
8 (Oct 27)	Network: basic operations
9 (Nov 3)	Networks: visualization
10 (Nov 10)	Networks: simulations
11 (Nov 17)	Networks: statistic
12 (Nov 24)	Online experiments
13 (Dec 1)	