Kellogg School of Management + McCormick School of Engineering

Emerging Topics in Computational Social Science

Winter 2025

Monday time TBD

*Draft syllabus – subject to change

Professor Dashun Wang
Dashun.wang@kellogg.northwestern.edu

Course Overview

Social scientists increasingly have access to datasets of unparalleled scope and complexity. At the same time, there have been remarkable development in fields like network science, data science, and machine learning over the past decade, which offer us a wide range of tools to help us make sense of this data with growing accuracy and robustness. Together, the new data and computational methods offer researchers opportunities to explore and understand human behavior at an unprecedented level of scale and detail, fueling the emergence of an exciting, interdisciplinary field called computational social science.

This course surveys the emerging frontiers in the field, open to students from both computational and social science backgrounds. For those new to the social sciences, this course is an opportunity to see where your computer science and statistical skills can go, with innovative applications to problems of massive societal interest. For those new to computational methods, this is a chance to develop the tools necessary to make new and exciting contributions, tools that will shape the originality and power of your work for years to come.

Requirements

- You must read articles marked with (*) prior to each class;
- You must complete all assigned homework;
- You must actively participate in class discussion, including presenting papers to the class;
- You must complete a replication research project;

Student Evaluation

- Homework: 30%
- Class attendance & participation: 20%
- Class Project (mid + final): 10%+40%

Administration

- The readings, syllabus, homework assignments, and class slides are available through Canvas.
- There are no "official" office hours; please feel free to make appointments with the course instructors.

Course Participation Expectations

- Students are expected to come to class having read the required reading thoroughly and prepared to discuss the papers' theoretical contributions, research designs, data, and findings.
- To facilitate vibrant and engaged discussion in which all students participate, **the instructors reserve the right to cold call students in class. This serves as your warning!** If this will be an issue or you need disability accommodations, please make an appointment with Professors Wang or Furnas to discuss your needs and circumstances.
- Students will present a paper from the course syllabus to the class. Grades for this will be reflected in the overall participation grade. Students will be asked their assignment of

preference. In week three, students will be told the week they will be presenting. We will do our best to accommodate students' preferences.

- *Presentation Expectations:*
  - Choose one of the papers from the week you are assigned, and create slides, practice your presentation, and present this paper to the class. You must choose from the selection papers the instructors provide following week 2 (those for which students have written referee reports as part of HW II). You may not present and review the same paper.
  - Prepare to present this paper as if it were your own research project, and be prepared to field detailed questions about the study, as you would if you were presenting your own work at a seminar or job talk.
  - **These presentations should be roughly 30 minutes, including time for questions.**
  - It is highly recommended that students read the supplementary materials closely for the papers they are assigned to present.

Schedule Overview

Class 1 – Jan 6 - Course Overview & Introduction (Dashun) - HW: Ref report I
Class 2 – Jan 13 - Teams (Dashun) – HW: Ref report II
Class 3 – Jan 27 - Networks 1 (Dashun) – HW: visualization
Class 4 – Feb 3 - Networks 2 (Dashun)
Class 5 – Feb 10 - Mid-term presentation (Idea pitch); Failure
Class 6 – Feb 17 - Text as Data (Zander)
Class 7 – Feb 24 - Social media (student presentations + TBD)
Class 8 – Mar 3 - AI
Class 9 – Mar 10 - Final Presentations

Due Dates

Class 5	Idea Pitch (<1 page + 5 minutes presentation)
Class 9	Project Presentation (15-20 minutes)
Week 12	Project Writeup (8-10 pages)

Assignments

1. Peer review Part 1

Select and read two recent computational social science papers, together with their peer review reports, published in the following journals: *Nature*, *Nature Human Behaviour, Nature Communications, Nature Machine Intelligence, Nature Computational Science*.

Note: Most but not all recent research in these journals includes peer review information. Make sure you select papers with referee reports.

Submit the author, title, journal, and DOI of the papers you read.

2. Peer Review Part 2

Choose a paper from the selected course readings from weeks 6, 7, or 8. The papers chosen for peer review will also be those that are later presented by other students – you may not review and present the same paper.

You will write a referee report for the paper you choose.
Reminder: A good structure for a referee report is as follows:
1. A 1-2 paragraph summary of the paper (question, data/methods, findings).

2. Comments and suggestions. This is the heart of the report and should begin with your major concerns / suggestions first and work down to more minor / smaller concerns / suggestions. It is helpful for the authors and the editor to do this as a numbered list, which allows them to respond to and highlight the points you make in a clear fashion.

Referee reports in computational social sciences are typically 2-3 pages in length.

3. Network Visualization

A series of simple lab sessions for network visualization exercises.

4. Final Project Assignment: Independent Research Project

Overview

The final project is an independent research project, aimed at contributing to your dissertation. You will work either solo or in teams of two, pairing with a peer from a different school or program to encourage interdisciplinary perspectives and approaches. The goal is to produce original research that demonstrates your combined expertise, promotes cross-disciplinary collaboration, and has the potential to be expanded into a future paper.

Objectives

1. **Develop a Research Question:** Clearly define a unique and compelling research question that reflects your combined interests and expertise.
2. **Literature Review:** Summarize and synthesize relevant existing literature to provide context and justification for your research question.
3. **Original Research:** Conduct original theoretical or empirical work that incorporates and benefits from your diverse academic backgrounds.

Outputs

1. **Research Write-Up (8-10 pages):** The write-up should include:

- Introduction and rationale for the research question.
- Detailed literature review.
- Methodology and data sources, highlighting the interdisciplinary approach.
- Analysis and findings.
- Conclusion and potential areas for future research.

2. **In-Class Presentation (15 minutes):** Toward the end of the quarter, each team will present their research findings to the class. The presentation should be concise, engaging, and reflect your collaborative work, summarizing your research question, methodology, key findings, and the significance of your work.

Data Sets

Teams are strongly encouraged to utilize one of the following datasets for their project:

- **SciSciNet:** A comprehensive dataset focused on scientific and scholarly research.
- **Overton:** A rich database of policy documents, references, and analyses.
- **BHHT (Brief History of Human Time):** cross-verified database of 2.2 million notable individuals using several Wikipedia editions and Wikidata.
  - Including related projects such as the Pantheon.

These datasets are chosen for their breadth and depth, offering a solid foundation for your analysis. If your team prefers a different dataset, you must jointly seek and obtain the instructor's approval before proceeding.

Course Details Subject to Change

Please note that the specifics of this course syllabus are subject to change in the case of unforeseen circumstances. Instructors will notify students of any changes as soon as possible. Students will be responsible for abiding by the changes.

Academic integrity

Students in this course are required to comply with the policies found in the booklet, "Academic Integrity at Northwestern University: A Basic Guide." All papers submitted for credit in this course must be submitted electronically unless otherwise instructed by the professor. Your written work may be tested for plagiarized content. For details regarding academic integrity at Northwestern or to download the guide, visit: https://www.northwestern.edu/provost/policies-procedures/academic-integrity/index.html

Any form of cheating, including improper use of content generated by artificial intelligence, constitutes a violation of Northwestern's academic integrity policy.

Accessibility

Northwestern University is committed to providing the most accessible learning environment as possible for students with disabilities. Should you anticipate or experience disability-related barriers in the academic setting, please contact AccessibleNU to move forward with the university's established accommodation process (e: accessiblenu@northwestern.edu; p: 847-467-5530). If you already have established accommodations with AccessibleNU, please let me know as soon as possible, preferably within the first two weeks of the term, so we can work together to implement your disability accommodations. Disability information, including academic accommodations, is confidential under the Family Educational Rights and Privacy Act.

Prohibition of Recording of Class Sessions by Students

Unauthorized student recording of classroom or other academic activities (including advising sessions or office hours) is prohibited. Unauthorized recording is unethical and may also be a violation of University policy and state law. Students requesting the use of assistive technology as an accommodation should contact AccessibleNU. Unauthorized use of classroom recordings – including distributing or posting them – is also prohibited. Under the University's Copyright Policy, faculty own the copyright to instructional materials – including those resources created specifically for the purposes of instruction, such as syllabi, lectures and lecture notes, and presentations. Students cannot copy, reproduce, display, or distribute these materials. Students who engage in unauthorized recording, unauthorized use of a recording, or unauthorized distribution of instructional materials will be referred to the appropriate University office for follow-up.

Statement of Inclusivity

This course strives to be an inclusive learning community, respecting those of differing backgrounds and beliefs. As a community, we aim to be respectful to all students in this class, regardless of race, ethnicity, socio-economic status, religion, gender identity or sexual orientation.

Support for Wellness and Mental Health

Northwestern University is committed to supporting the wellness of our students. Student Affairs has multiple resources to support student wellness and mental health. If you are feeling distressed or overwhelmed, please reach out for help. Students can access confidential resources through the Counseling and Psychological Services (CAPS), Religious and Spiritual Life (RSL) and the Center for Awareness, Response and Education (CARE). All Northwestern students are also eligible to access support at no cost though TimelyCare, a virtual mental health platform that provides counseling, health coaching and 24/7 on-demand services.

Additional information on the resources mentioned above can be found here:

https://www.northwestern.edu/counseling/

https://www.northwestern.edu/religious-life/

https://www.northwestern.edu/care/

https://www.northwestern.edu/studentaffairs/timelycare.html

Weekly Reading

Class 1 Course Overview

(*) Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis et al. "Computational social science." *Science* 323, no. 5915 (2009): 721-723.
(*) Lazer, David MJ, Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon et al. "Computational social science: Obstacles and opportunities." *Science* 369, no. 6507 (2020): 1060-1062.
(*) Liu, Lu, Benjamin F. Jones, Brian Uzzi, and Dashun Wang. "Data, measurement and empirical methods in the science of science." *Nature Human Behaviour* (2023): 1-13.
Hofman, Jake M., Duncan J. Watts, Susan Athey, Filiz Garip, Thomas L. Griffiths, Jon Kleinberg, Helen Margetts et al. "Integrating explanation and prediction in computational social science." *Nature* 595, no. 7866 (2021): 181-188.

Class 2 Teams

(*) Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. "The increasing dominance of teams in production of knowledge." *Science* 316, no. 5827 (2007): 1036-1039.
(*) Wu, Lingfei, Dashun Wang, and James A. Evans. "Large teams develop and small teams disrupt science and technology." *Nature* 566, no. 7744 (2019): 378-382.
(*) Woolley, Anita Williams, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. "Evidence for a collective intelligence factor in the performance of human groups." *Science* 330, no. 6004 (2010): 686-688.
Chowdhary, Sandeep, Luca Gallo, Federico Musciotto, and Federico Battiston. "Team careers in science: formation, composition and success of persistent collaborations." *arXiv preprint arXiv:2407.09326* (2024).
Risha, Zak, Yiling Lin, Erin Leahey, and Lingfei Wu. "Replacing the Renaissance Man: Are Teams More than the Sum of Their Parts?." *arXiv preprint arXiv:2304.14518* (2023).

Class 3+4 Networks

(*) Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world'networks." *Nature* 393, no. 6684 (1998): 440-442.

(*) Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *Science* 286, no. 5439 (1999): 509-512.
Albert, Réka, Hawoong Jeong, and Albert-László Barabási. "Error and attack tolerance of complex networks." *Nature* 406, no. 6794 (2000): 378-382.
Kleinberg, Jon M. "Navigation in a small world." *Nature* 406.6798 (2000): 845-845.
Girvan, Michelle, and Mark EJ Newman. "Community structure in social and biological networks." *Proceedings of the national academy of sciences* 99, no. 12 (2002): 7821-7826

Class 5 Midterm presentation + Failure

(*) Wang, Yang, Benjamin F. Jones, and Dashun Wang. "Early-career setback and future career impact." *Nature Communications* 10, no. 1 (2019): 4331.
(*) Yin, Yian, Yang Wang, James A. Evans, and Dashun Wang. "Quantifying the dynamics of failure across science, startups and security." *Nature* 575, no. 7781 (2019): 190-194.
(*) Eskreis-Winkler, L., Woolley, K., Erensoy, E., & Kim, M. (2024). The exaggerated benefits of failure. *Journal of Experimental Psychology: General, 153*(7), 1920–1937. https://doi.org/10.1037/xge0001610

Mariani, Manuel S., Federico Battiston, Emőke-Ágnes Horvát, Giacomo Livan, Federico Musciotto, and Dashun Wang. "Collective dynamics behind success." *Nature Communications* 15, no. 1 (2024): 10701. (good review article)

Class 6 Text as Data (Zander)

(*) Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. "Text as data." *Journal of Economic Literature* 57, no. 3 (2019): 535-574.
(*) Rodriguez, Pedro L., and Arthur Spirling. "Word embeddings: What works, what doesn't, and how to tell the difference for applied research." *The Journal of Politics* 84, no. 1 (2022): 101-115.
Rodriguez, Pedro L., Arthur Spirling**,** and Brandon M. Stewart. "Embedding regression: Models for context-specific description and inference." *American Political Science Review* (2023): 1-20
Kozlowski, Austin C., Matt Taddy, and James A. Evans. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84, no. 5 (2019): 905-949.
Egami, Naoki, Musashi Hinck, Brandon Stewart, and Hanying Wei. "Using imperfect surrogates for downstream inference: Design-based supervised learning for social science applications of large language models." *Advances in Neural Information Processing Systems* 36 (2024).
Ash, Elliott, and Stephen Hansen. "Text algorithms in economics." *Annual Review of Economics* 15 (2023): 659-688.
Baden, Christian, Christian Pipal, Martijn Schoonvelde, and Mariken AC G. van der Velden. "Three gaps in computational text analysis methods for social sciences: A research agenda." *Communication Methods and Measures* 16, no. 1 (2022): 1-18.
Egami, Naoki, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart. "How to make causal inferences using texts." *Science Advances* 8, no. 42 (2022): eabg2652.

Kapoor, Sayash, Emily Cantrell, Kenny Peng, Thanh Hien Pham, Christopher A. Bail, Odd Erik Gundersen, Jake M. Hofman et al. "Reforms: Reporting standards for machine learning based science." *arXiv preprint arXiv:2308.07832* (2023).

Class 6 Social Media + Misinformation (Pennycook or Mohsen)

(*) Pennycook**,** G., Epstein, Z., Mosleh, M., Arechar, A.A., Eckles, D. and Rand, D.G., 2021. Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), pp.590-595.
(*) Lu, Y., & Shen, C., 2023. Unpacking Multimodal Fact-checking: Features and Engagement of Fact-checking Videos on Chinese TikTok (Douyin). *Social Media + Society* 9(1).
(*) González-Bailón, Sandra, David Lazer, Pablo Barberá, Meiqing Zhang, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio et al. "Asymmetric ideological segregation in exposure to political news on Facebook." *Science* 381, no. 6656 (2023): 392-398
(*) Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D., Marlow, C., Settle, J.E. and Fowler, J.H., 2012. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), pp.295-298.
Juul, J.L. and Ugander, J., 2021. Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences*, 118(46), p.e2100786118.
Stein, Jonas, Marc Keuschnigg, and Arnout van de Rijt. "Network segregation and the propagation of misinformation." *Scientific Reports* 13, no. 1 (2023): 917.
Mosleh, Mohsen, Qi Yang, Tauhid Zaman, Gordon Pennycook, and David G. Rand. "Differences in misinformation sharing can lead to politically asymmetric sanctions." *Nature* (2024): 1-8.
Budak, Ceren, Brendan Nyhan, David M. Rothschild, Emily Thorson, and Duncan J. Watts. "Misunderstanding the harms of online misinformation." *Nature* 630, no. 8015 (2024): 45-53.

Class 8 AI/LLM

Readings TBD

Class 9 Final presentation