Prof. Brennan Klein
b.klein@northeastern.edu
Network Science Institute
Northeastern University

CNET 5052
Advanced Tools for Complex Network Analysis

Spring 2026; 4 credits
Tues: 1:30 – 4:50pm
Jan. 7 – Apr. 24, 2026
177 Huntington #226

Summary

This course extends the foundations of CNET 5051 into a set of advanced, research-facing tools for complex network analysis. Topics emphasize modern workflows for network inference and modeling (e.g., link prediction, sparsification, Bayesian/EM-style reasoning, stochastic block models and model fitting), computational methods for structure (e.g., distances, spectral tools, motifs, signed networks), and dynamics and simulation (e.g., reconstruction, games on networks, agent-based models). A parallel goal throughout the semester is to develop good research habits: reproducible code, clear documentation, defensible evaluation, and careful interpretation. Course materials (notebooks, readings, assignments, and code templates) will be distributed through a dedicated GitHub repository. Students conclude the semester with a final project in the form of a short research-style paper that presents a network-science question, method, or empirical study with clear results and limitations.

Course Learning Outcomes

1. Build reproducible network-analysis workflows in Python, including clear project structure, documentation, and version-controlled code suitable for research collaboration.
2. Implement and evaluate methods for network structure and inference, including graph distances, link prediction, and sparsification/sampling, with appropriate baselines and metrics.
3. Formulate and fit probabilistic and generative network models (e.g. stochastic block models), and interpret results with attention to uncertainty, model assumptions, and diagnostics.
4. Apply computational tools for network structure beyond standard metrics, including spectral methods, motifs, and signed-network analysis.
5. Design and analyze network dynamics and simulation studies (e.g., reconstruction problems, games on networks, agent-based models).
6. Produce a research-style final project that combines data, methods, results, and interpretation into a reproducible repository and a well-structured paper with proper citation practices.

Coursework, Class Structure, Grading

This is a once-weekly, hands-on, code-forward course focused on developing comfort, fluency, and independence with computational workflows in network science. Each class meeting blends conceptual discussions, notebook-driven demonstrations, short implementation exercises, and guided time for students to deepen their computational practice.

Grading will be based on the following:

- **Participation & Engagement (10%)** – Attendance, class participation, contribution to discussions.
- **Assignments (45%)** – A mix of programming exercises, write-ups, and theoretical exercises.
- **Final Project (45%)** – Proposal (5%), mid-semester update presentation (5%), final paper + reproducible repository (25%), final presentation (10%).

Final Project Details

The final project is a research-style project designed to mirror how network science work is actually done: you will pose a question (or evaluate a method), assemble or generate data, implement an analysis pipeline, report results, and communicate limitations. Projects may be methodological (e.g., comparing techniques, extending existing tools, theoretical work, etc.) or applied (e.g., a focused empirical study of an online, biological, spatial, or infrastructure network). The emphasis is on clarity, defensible evaluation, and reproducibility.

Project milestones

- **Tue, Jan 27 (in class): Proposal + short presentation.** Submit a brief (up to 1 page) proposal and give a short (no more than 5 min) in-class overview of your plan.
- **Tue, Feb 17 (in class): Mid-semester update presentations.** Present progress, obstacles, and preliminary results to receive feedback (5 min).
- **Tue, Apr 21 (in class): Final project presentations.** Present your completed work and receive peer/instructor feedback (12 min, +3 min Q&A).

Final submission package. The final project submission must include:

- **A reproducible GitHub repository** containing:
  - A clear `README` describing the project, how to reproduce results, and how data are obtained.
  - A reproducible environment specification (e.g., `requirements.txt` or `environment.yml`).
  - Code and/or notebooks that run end-to-end (data $\rightarrow$ results $\rightarrow$ figures/tables).
  - Proper attribution for any external code, data, or tools used.
- **A research paper** (PDF), typically 8-12 pages, written for a scientific audience.
- **A final presentation** (in class) that communicates motivation, methods, key results, limitations.

Evaluation criteria. Projects will be assessed based on the clarity and specificity of the research question and the motivation for the design choices that follow from it. Work should demonstrate methodological correctness, including appropriate use of course tools and accurate implementation. Projects should also include a defensible evaluation strategy—with sensible baselines, well-chosen metrics, and validation or robustness checks that support the claims being made. Strong projects interpret results carefully, making clear what the findings do and do not imply, and explicitly discussing limitations. Reproducibility is essential: repositories should be well organized and documented, with enough information for another reader to rerun the analysis and recover the main results. Finally, projects will be evaluated on communication quality, including the structure and readability of the paper, the clarity of figures and tables, and the effectiveness of the final presentation.

Course Materials

There is no single textbook that covers the scope of this course. Instead, students will work with a combination of open-source texts, research articles, and software tools. All required readings, notebooks, assignments, and code templates will be available through the course GitHub repository.

Resources

- *Python & Data Science:*
  - VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, Inc. `https://jakevdp.github.io/PythonDataScienceHandbook/`

- Severance, C. (2016). Python for Everybody: Exploring Data using Python 3. Charles Severance. `https://do1.dr-chuck.com/pythonlearn/EN_us/pythonlearn.pdf`
- Downey, A. (2012). Think Python: How to Think Like a Computer Scientist. `https://www.greenteapress.com/thinkpython/thinkpython.pdf`

- *Network Science & Complex Systems:*
  - Barabási, A.L. & Pósfai, M. (2016). *Network Science*. Cambridge University Press. `https://networksciencebook.com/`
  - Thurner, S., Hanel, R., & Klimek, P. (2018). Introduction to the Theory of Complex Systems. Oxford University Press. `https://academic.oup.com/book/25504`
  - Klein, B., Smith, A., Chinazzi, M., Zhang, Q., et al. (2025) Network Science Data & Models Python Textbook — `https://network-science-data-and-models.github.io/phys7332_fa25/README.html`

Software and Data

- Python (e.g. `numpy`, `pandas`, `matplotlib`, `networkx`, `statsmodels`, `scikit-learn`, among others) and Jupyter notebooks, distributed through the course GitHub repository.

Instructors

Brennan Klein is core faculty at the Network Science Institute and Assistant Teaching Professor in the Department of Physics. He is the program director of the MS in Complex Network Analysis at Northeastern University. Prof. Klein is also the director of the Complexity & Society Lab, which is focused on two broad research areas: 1) Information, emergence, and inference in complex systems: developing tools and theory for characterizing dynamics, structure, and scale in networks, and 2) Public health and public safety: drawing on complex systems science to document—and fight against—emergent or systemic disparities in society, especially as they relate to public health and public safety. As of 2025, he is also the director of NetSI Sport, an interdisciplinary research group focusing on complex systems-inspired approaches to sports analytics. In 2023, Prof. Klein was awarded the René Thom Young Researcher Award, given to a researcher to recognize substantial early career contributions and leadership in research in Complex Systems-related fields. Prof. Klein is the Data for Justice Fellow at the Institute on Policing, Incarceration & Public Safety at Harvard University's Hutchins Center for African & African American Research. He received a PhD in Network Science in 2020 from Northeastern University and earned his BA in Cognitive Science & Psychology from Swarthmore College in 2014. Website: `brennanklein.com`.

Milo Trujillo is a Postdoctoral Research Fellow and Associate Director of the Communication Media and Marginalization Lab at the Network Science Institute. His primary interest is in how the structure of online platforms, including both their technical design and social policies, influences online group behavior. These topics include content moderation and deplatforming, the emergence of alt-tech, decentralized social platforms, and the governance of open source software. Dr. Trujillo received a PhD in Complex Systems and Data Science in 2024 from the University of Vermont, and received M.S. and B.S. degrees in computer science and a B.S. in Science and Technology Studies from Rensselaer Polytechnic Institute in 2020 and 2018. Website: backdrifting.net.

Office Hours

TBD

Accessibility and Accommodations

Northeastern is committed to providing equal educational opportunities for all students. Students who require accommodations for a documented disability should contact the Disability Resource Center as early as possible to ensure that appropriate arrangements can be made. Once you have documentation, please share your accommodation letter with me so we can discuss how best to support your learning.

Health and Wellness

Your well-being is essential to your success in this course. If you are experiencing stress, anxiety, illness, or other challenges, I encourage you to reach out for support. Northeastern's University Health and Counseling Services and WeCare office provide confidential resources for medical and mental health needs. Please also feel free to communicate with me if circumstances are affecting your coursework; I will work with you to connect you to the right resources.

Academic Integrity

All students are expected to uphold Northeastern University's Academic Integrity Policy, which prohibits cheating, plagiarism, fabrication, unauthorized collaboration, and other forms of academic dishonesty. You are responsible for ensuring that your work reflects your own effort and analysis, even when you consult outside resources such as peers, published materials, or AI tools. Proper citation is required whenever you use code, data, text, or ideas that are not your own. Questions about what counts as appropriate collaboration or citation should be raised with me directly. Suspected violations will be referred to the Office of Student Conduct and Conflict Resolution. More information can be found here: `https://osccr.sites.northeastern.edu/academic-integrity-policy/`.

All student records and coursework in this class are handled in compliance with the Family Educational Rights and Privacy Act. Please use your Northeastern email account for all course communications.

Policy on Artificial Intelligence and Large Language Models

This course recognizes the potential of artificial intelligence (AI) tools—such as ChatGPT, Copilot, Claude, and other text or code generators—to support learning, creativity, and efficiency. You are encouraged to use AI when it adds value to your learning process, provided that its use is transparent, relevant, and critically evaluated. AI can help brainstorm ideas, debug code, generate visualizations, or give writing feedback, but it is not a substitute for your own analysis or reasoning.

Guidelines for Use

- AI use will vary depending on the assignment. Labels will be provided to indicate whether AI use is prohibited, permitted, encouraged, or required, depending on the learning objectives.
- For assignments where AI use is allowed: cite the tool, include information about the prompt or queries you used, and briefly explain how it contributed to your work. This is not meant to police your prompts, but rather to crowdsource and share effective strategies for navigating the tool.
- You remain responsible for the accuracy, originality, and integrity of all submitted work. AI tools are known to make errors, invent references, or introduce bias. Verification is your responsibility.

Learning Orientation

Think of AI as a *ladder, not a crutch*. Its purpose is to extend your abilities, not to replace the productive struggle of problem-solving. Over-reliance on AI will limit your growth, while thoughtful use can accelerate your improvement on a range of quantitative and qualitative skills. Throughout the semester, we will highlight best practices for integrating AI into analysis, coding, and communication in ways that strengthen—not weaken—your understanding.

Schedule

Schedule and topics may be adjusted with reasonable notice. Friday entries are events (not class meetings).

Week 1: Introduction, Growth, Distances

Tuesday, January 13, 2026 – Class 1: Introduction, Network Growth, Graph Distances *Both*

- Course overview; computational expectations; what "advanced tools" means in practice.
- Network growth models (with an emphasis on implementable generative processes).
- Graph distances at scale: shortest paths, efficiency, diameter, and practical approximations.
- Final project examples + structured brainstorming.

Friday, January 16, 2026 – **Assignment 1 announced**

Week 2: Link Prediction and Sparsification

Tuesday, January 20, 2026 – Class 2: Link Prediction & Sparsification *Klein*

- Link prediction as inference: scores, features, and evaluation (with attention to leakage).
- Similarity-based predictors and baselines; where they work and where they fail.
- Sparsification/sampling for scale: what structure is preserved, what is distorted, and why.
- Connections to homophily and robustness (as framing for the homework).

Week 3: Bayesian Thinking for Network Inference

Tuesday, January 27, 2026 – Class 3: Bayesian Methods & Expectation Maximization *Both*

- **Project idea due (in class):** short write-up (up to 1 page) + brief presentation.
- Bayes' rule, likelihood, priors, and posteriors (as tools, not ideology).
- A compact view of latent-variable models and EM as an inference pattern.
- How probabilistic framing changes link prediction and uncertainty reporting.

Friday, January 30, 2026 – **Assignment 1 due**

Week 4: Communities Revisited and the SBM as a Generative Object

Tuesday, February 3, 2026 – Class 4: Communities revisited; SBM (forward process) *Both*

- Community structure: "algorithmic" vs "model-based" perspectives.
- Stochastic Block Models as a data-generating story (and what that implies).
- What SBMs can/can't represent; why degree correction matters (conceptually).

Friday, February 6, 2026 – **Assignment 2 announced**

Week 5: Fitting SBMs in Practice with `graph-tool`

Tuesday, February 10, 2026 – Class 5: `graph-tool` + SBM inference *Klein*

- Practical SBM fitting workflows in `graph-tool`.
- Model selection / complexity control (intuition + what the software is optimizing).
- Interpreting partitions responsibly: uncertainty, stability, and diagnostics.

Week 6: Spatial Networks + Mid-Semester Project Updates

Tuesday, February 17, 2026 – Class 6: Spatial networks (and intermediate project presentations) *Klein*

- **Intermediate project update presentations (in class).**
- Embedding networks into space: distance effects, spatial statistics, and null models.
- Properties of spatial networks and what changes when geometry matters.

Friday, February 20, 2026 – **Assignment 2 due**

Week 7: Machine Learning Workflows for Network Data

Tuesday, February 24, 2026 – Class 7: Machine Learning from a Data Science Lens *Trujillo*

- End-to-end ML pipelines for network problems: features, splits, baselines, metrics.
- When "standard" ML assumptions break on network data (dependence, sampling, leakage).

Friday, February 27, 2026 – **Assignment 3 announced**

Spring Break

Tuesday, March 3, 2026 – No class – *Spring Break*

Week 8: Topics in Big Data for Network-Scale Questions

Tuesday, March 10, 2026 – Class 8: Big Data; HyperLogLog *Trujillo*

- Streaming constraints and approximate computation as a design choice.
- HyperLogLog for approximate distinct counting: intuition and implementation.
- Where sketches plug into network analysis workflows (and where they don't).

Friday, March 13, 2026 – **Assignment 3 due**

Week 9: Network Dynamics and Reconstruction

Tuesday, March 17, 2026 – Class 9: Network dynamics and network reconstruction *Klein*

- Dynamics on networks as computational objects (simulation and inference).
- Reconstruction problems: partial observation, missing edges, and temporal evidence.
- Connecting mechanistic models to data and to evaluation.

Friday, March 20, 2026 – **Assignment 4 announced**

Week 10: Games on Networks and Agent-Based Models

Tuesday, March 24, 2026 – Class 10: Games on networks and ABMs *Klein*

- Games on networks: strategic interaction with topology as structure.
- Agent-based models on networks: design patterns, debugging, and interpretation.
- What "mechanism" buys you (and what it doesn't) in network settings.

Week 11: Spectral Methods

Tuesday, March 31, 2026 – Class 11: Spectral methods *Klein*

- Laplacians, eigenvectors, and what spectra say about structure.
- Spectral clustering (conceptual and computational view).
- Spectral ideas as "tools you can reuse" across network tasks.

Friday, April 3, 2026 – **Assignment 4 due**

Week 12: Motifs and Signed Networks

Tuesday, April 7, 2026 – Class 12: Motifs & Signed Networks *Klein*

- Motifs: counting, null models, and what "significance" really means.
- Signed networks: balance, structure, and analysis tools for positive/negative ties.

Week 13: Flexible Topics / Tooling Comparisons

Tuesday, April 14, 2026 – Class 13: Flexible Topics *Both*

- Student-driven topics based on project needs and open questions from the semester.
- Tooling comparisons and practical workflow choices (when/why to use what).

Week 14: Final Project Presentations

Tuesday, April 21, 2026 – Class 14: Final project presentations *Both*

- **Final project paper + repository due**
- Project presentations + feedback.
- Synthesis and wrap-up.
