Syllabus for CSCI 5352
Network Analysis and Modeling
Fall 2020

Lectures: Monday, Wednesday, Friday, from 1:50pm – 2:40pm

Zoom: Unpublished. Please email.

Lecturer: Dan Larremore

Office:	Online via Zoom — Link TBD
Email:	daniel.larremore@colorado.edu
Web Page:	LarremoreLab.github.io/5352
Office Hours:	TBD or by appointment.

Description: This graduate-level course will examine modern techniques for analyzing and modeling the structure and dynamics of complex networks. The focus will be on statistical algorithms and methods, and both lectures and assignments will emphasize model interpretability and understanding the processes that generate real data. Applications will be drawn from computational biology and computational social science. No biological or social science training is required. (Note: this is not a scientific computing course, but there will be plenty of computing for science.)

Prerequisites (recommended): CSCI 3104 (undergraduate algorithms) and either CSCI 3022 (data science) or APPM 3570 (applied probability), or equivalent preparation.

An adequate mathematical and programming background is mandatory. The concepts and techniques covered in this course depend heavily on basic statistics (distributions, Monte Carlo techniques), scientific programming, and calculus (integration and differentiation). Students without sufficient preparation will struggle to keep up with the lectures and assignments. Students without proper preparation may audit the course. Those who audit the course are considered full participants but their homeworks may not always be graded.

Required Texts:
(1) *Networks: An Introduction* 2nd Edition by M.E.J. Newman
(2) *Pattern Recognition and Machine Learning* by C.M. Bishop

Overview:

- Mostly lecture-style class, with some guest lectures and some class discussions.
- Problem sets (12 total, worth 60% of grade) due every week throughout the semester.
- Class project (worth 40% of grade) due at end of semester.
- No exams.
- Networks are cool.

Tentative schedule:

Week 1	Introduction and overview
Week 2	Measures of structural importance
Week 3	Random graphs I: homogeneous degrees
Week 4	Random graphs II: heterogeneous degrees
Week 5	Large-scale structure I: modularity, assortativity, homophily
Week 6	Large-scale structure II: stochastic block models
Week 7	Spreading processes on networks
Week 8	Ranking in directed networks and pairwise comparisons
Week 9	Wrangling network data I: sampling
Week 10	Wrangling network data II: auxiliary information
Week 11	Spatial networks
Week 12	Growing networks
—	Fall break
Week 13	Dynamic networks
Weeks 14–15	Project presentations

Deadlines:

1. CP, topic proposal — Oct. 9 (Friday)
2. CP, short presentations — Nov 30, Dec. 2, 4, 7, 9, 11
3. CP, final write up — Dec. 14 (Monday, 11:59am)

Lecture recordings:

- Live lectures via Zoom: Unpublished. Please email.
- Archive of Lecture Recordings: see Canvas.

If you need help with getting Zoom up and running, please visit
http://www.colorado.edu/oit/services/conferencing-services/web-conferencing-zoom

Coursework and grading:

Most of the class will be standard graduate-style lectures by me. These will be supplemented by guest lectures on special or advanced topics, and class discussions of selected papers drawn from the network science literature. Many lectures will have associated lecture notes, which I will post on the class website after class.

Grades will be assigned based on (i) performance on the problem sets, and (ii) performance on the class project. Problem sets will develop and extend selected class topics and will introduce additional topics not covered in class. There are no written examinations in the course. Students are expected to spend serious quality time on the problem sets and project.

Problem sets (PS): The 12 problem sets, due every week throughout the first 12 weeks of the semester, will include a mixture of mathematical, programming, and data analysis problems.

Programming and data analysis problems may be completed in a programming language of your choice. I recommend using something like Python or Matlab, which have good support for data analysis and visualization. Familiarity with `Mathematica` may be useful for some of the mathematical problems, and you are free to use Mathematica in any way to complete calculations.

- Problem sets will be due roughly one week after they are assigned.
- **Solutions must** be in PDF format (e.g., typeset using LaTeX), should include all necessary details for me to follow the logic. *Non-PDF files will receive an automatic 0 score.*
- **Solutions must** be submitted to me on the day they are due.
- **Solutions must** have your first and last name and problem set number *inside* your file (imagine printing the document). *Unidentifiable submissions will receive no credit.*
- **Solutions must** include your source code for your algorithms (do not include library files), appended **to the end** of your submission. Do not submit them as a separate file.
- **Late assignments will not be accepted**. *Late submissions will receive an automatic 0.*
- **Collaboration is encouraged** on the problem sets. However, you may not copy (in any way) from your collaborators and you must respect University academic policies at all times. To be clear: you may discuss the problems verbally, but you must write up your solutions separately. If you do discuss the problems with someone (and you are encouraged to!), you must then list and describe the extent of your collaboration in your solutions (a footnote is fine). *Copying from any source, in any way, including the Web but especially from another student (past or present), is strictly forbidden.*
- If you are unsure about whether something is permitted under these rules, ask me well before the deadline.
- I do not expect every algorithm to be coded from scratch, but I do expect you to do a substantial amount of coding yourself. It is okay to use publicly available libraries that do standard numerical network calculations. If you use such libraries, *you must state in your solution to the corresponding problem* which libraries you used and what you used them for. You are still required to submit the rest of your code; do not submit code for the libraries.

Class project (CP): The purpose of the class project is to explore a research question of your own devising related to network analysis and modeling. Class projects may be done in teams of 1–3. If you choose to work with others, you are responsible for finding those people and ensuring an even division of work. The deliverables for the class project are (i) a short (10 minute-ish) in-class presentation of your project results and (ii) a 10-page writeup of your project results (via email).

To facilitate this component of the class, each team must submit (via email) a short project proposal. This will provide each team with about 8 weeks to complete the project. The proposal must include (i) the names of the individuals in your team, (ii) two paragraphs describing any background material, including necessary references to the scientific literature, and what you are specifically going to do, and (iii) a brief description of any data you plan to use.

The best project topics are those on which you can make good progress in 6 weeks. I am happy to provide feedback on your project ideas, and if your proposal is inappropriately scoped, I may ask you to revise it.

Grading: Grades will be assigned based on a 60% problem sets, and 40% project division.

Letter grades will not be assigned until after all work for the semester has been submitted and graded. In the meantime, only numerical grades will be tracked.

Advice for writing up your solutions:
Your solutions for the problem sets should have the following properties. I will be looking for these when I grade them:

1. **Clarity**: Your solutions should be both clear and concise. The longer it takes me to figure out what you're trying to say, the less likely you are to receive full credit. The more clear you make your thought process, the more likely you are to get full credit.
2. **Completeness**: Full credit is based on (i) sufficient intermediate work and (ii) the final answer. For many problems, there are multiple paths to the correct solution, and I need to understand exactly how you arrived at the solution. A heuristic for deciding how much detail is sufficient: if you were to present your solution to the class and everyone understood the steps and could repeat them themselves, then you can assume it is sufficient.
3. **Succinctness**: Solutions should be long enough to convince me that your answer is correct, but no longer. More than half a page of dense algebra, more than a few figures or more than a page or two per problem is probably not succinct. Clearly indicate your final answer (circle, box, underline, whatever). Rewriting your solutions, with an eye toward succinctness, before submitting will help. Strive for maximum understanding in minimum space.
4. **Numerical experiments**: Some programming problems will require you to conduct numerical experiments using random number generators. One run is not a result. Your goal is to

produce beautifully smooth central tendencies and you should average your measured quantity over as many independent trials as is necessary to get something smooth. Further, your results should span several orders of magnitude. I recommend a dozen or so measurement values across the x-axis, distributed logarithmically, e.g., $n = \{2^4, 2^5, 2^6, \ldots\}$.

Solutions that use a numerical experiment but fail to adequately explain the experimental design will receive an automatic 0.

5. **Source code**: Your source code for all programming problems must be included at the *end* of your solutions. Code must include copious comments explaining the sub-algorithms and must be run-able; that is, if I try to compile and run it, it should work as advertised.

6. **Data analysis**: In presenting results from analyzing real data, you should always briefly describe the data to the reader. Explain what the network is (what is a vertex and when are two vertices connected) and what any network meta-data (vertex attributes, edge weights, etc.) means. Try also to explain what questions you are investigating, and how your results address those questions.

7. **Figures**: Always label your axes and always label your data series. Avoid having a lot of wasted whitespace in your figures (choose appropriate x- and y-ranges). Know what message you want the reader to take away from your figure, and be sure your figure accomplishes it clearly.

   Figures that lack axis or data series labels will receive an automatic 0.

Suggestions: Suggestions for improvement are welcome at any time. Any concern about the course should be brought first to my attention. Further recourse is available through the office of the Department Chair or the Graduate Program Advisor, both accessible on the 7th floor of the Engineering Center Office Tower.

COVID-19: Please see the final page of the syllabus for this important Addendum.

Honor Code: As members of the CU academic community, we are all bound by the CU Honor Code. I take the Honor Code very seriously, and I expect that you will, too. Any significant violation will result in a failing grade for the course and will be reported. As a small bonus for having read the syllabus carefully, I will award some extra credit points if you send me an email containing a photo of a panda (any kind is fine). Here is the University's statement about the matter:

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code (honor@colorado.edu); 303-492-5550). Students who are found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found at the Honor Code Office website.
https://www.colorado.edu/sccr/honor-code.

Accommodation for Disabilities: If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website. Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition or injury, see Temporary Medical Conditions under the Students tab on the Disability Services website
https://www.colorado.edu/disabilityservices/students/temporary-medical-conditions.

This course requires the use of the Zoom conferencing tool which is currently not accessible to users using assistive technology. If you use assistive technology to access the course material, please contact your faculty member immediately to discuss.

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, I will make reasonable efforts to accommodate such needs if you notify me of their specific nature by the end of the 3rd week of class. See full details at
http://www.colorado.edu/policies/observance-religious-holidays-and-absences-classes-andor-exams

Classroom Behavior: Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. For more information, see the policies on classroom behavior and the Student Code of Conduct.
https://www.colorado.edu/sccr/
https://www.colorado.edu/policies/student-classroom-and-course-related-behavior

Sexual Misconduct, Discrimination, Harassment and/or related retaliation: The University of Colorado Boulder (CU Boulder) is committed to fostering a positive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct intimate partner abuse (including dating or domestic violence), stalking, protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or cureport@colorado.edu. Information about the OIEC, university policies, anonymous reporting, and the campus resources can be found on the OIEC website.
https://cuboulder.qualtrics.com/jfe/form/SV_0PnqVK4kkIJIZnf
https://www.colorado.edu/oiec/

Please know that faculty and instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about options for reporting and support resources.

Syllabus Statements

Requirements for COVID-19

As a matter of public health and safety due to the pandemic, all members of the CU Boulder community and all visitors to campus must follow university, department and building requirements, and public health orders in place to reduce the risk of spreading infectious disease. Required safety measures at CU Boulder relevant to the classroom setting include:

- maintain 6-foot distancing when possible,
- wear a face covering in public indoor spaces and outdoors while on campus consistent with state and county health orders,
- clean local work area,
- practice hand hygiene,
- follow public health orders, and
- if sick and you live off campus, do not come onto campus (unless instructed by a CU Healthcare professional), or if you live on-campus, please alert CU Boulder Medical Services.

Students who fail to adhere to these requirements will be asked to leave class, and students who do not leave class when asked or who refuse to comply with these requirements will be referred to Student Conduct and Conflict Resolution. For more information, see the policies on COVID-19 Health and Safety and classroom behavior and the Student Code of Conduct. If you require accommodation because a disability prevents you from fulfilling these safety measures, please see the "Accommodation for Disabilities" statement on this syllabus.

Before returning to campus, all students must complete the COVID-19 Student Health and Expectations Course. Before coming on to campus each day, all students are required to complete a Daily Health Form.

Students who have tested positive for COVID-19, have symptoms of COVID-19, or have had close contact with someone who has tested positive for or had symptoms of COVID-19 must stay home and complete the Health Questionnaire and Illness Reporting Form remotely. In this class, if you are sick or quarantined**, please stay off campus, and discuss with me an potential impacts on coursework as early as possible**.