Syllabus for CSCI 5352
Network Analysis and Modeling
Fall 2016

Lectures: Monday and Wednesday from 9:30am – 10:45am in ECCS 1B12

Lecturer: Aaron Clauset

Office:	ECOT 743
Email:	aaron.clauset@colorado.edu
Web Page:	`http://tuvalu.santafe.edu/~aaronc/courses/5352/`
Office Hours:	Monday 1:15–2:45pm 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 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.

Required Texts:
(1) *Networks: An Introduction* 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 (6 total, worth 50% of grade) due every 2 weeks throughout the semester.
- Class project (worth 30% of grade) due at end of semester.
- No exams.
- Networks are cool.

Piazza Class Discussion: We will use Piazza for class discussion and Q&A. The system is designed to help you get help from classmates and myself. Rather than emailing questions to the teaching staff, please post your questions on our Piazza forum.

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	Large-scale structure III: more block models
Week 8	Wrangling network data I: sampling
Week 9	Wrangling network data II: auxiliary information
Week 10	Spatial networks
Week 11	Growing networks
Week 12	Dynamic networks
Week 13	Advanced topics
Week 14	Fall break
Weeks 15–16	Project presentations

Deadlines:

	assigned	due
PS 1	Aug. 22 (Monday)	Sept. 6 (Tuesday)
PS 2	Sept. 7 (Wednesday)	Sept. 20 (Tuesday)
PS 3	Sept. 21 (Wednesday)	Oct. 4 (Tuesday)
CP, topic proposal		Oct. 5 (Wednesday)
PS 4	Oct. 5 (Wednesday)	Oct. 18 (Tuesday)
PS 5	Oct. 19 (Wednesday)	Nov. 1 (Tuesday)
PS 6	Nov. 2 (Wednesday)	Nov. 15 (Tuesday)
CP, short presentation		Dec. 5, 6
CP, final write up		Dec. 12 (Monday, 11:00am)

Lecture live steam, via Zoom: Meeting ID 952307376

- Join via web browser: `https://cuboulder.zoom.us/j/952307376`
- Join via Zoom app (using meeting ID)
- Join via phone: 1-646-558-8656 or 1-408-638-0968

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: Attendance to the lectures is required.

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, (ii) performance on the class project, and (iii) class attendance and participation. 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 6 problem sets, due every 2 weeks throughout 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 `Matlab` or Python, 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 two weeks after they are assigned (schedule given above).
- **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 via email by 11:59pm the day they are due.
- **Solutions must** have your first and last name and problem set number in 2 places: (i) in the filename of your submission, e.g., `ClausetAaron-PS1.pdf`, and (ii) 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 time 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 (on-campus) a 20% attendance, 45% problem sets, and 35% project division, or (off-campus) 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.

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. Here is the University's statement about the matter:

All students of the University of Colorado at Boulder are responsible for knowing and adhering to the academic integrity policy of this institution. Violations of this policy may include: cheating, plagiarism, aid of academic dishonesty, fabrication, lying, bribery, and threatening behavior. All incidents of academic misconduct shall be reported to the Honor Code Council (`honor@colorado.edu`; 303-735-2273). Students who are found to be in violation of the academic integrity policy will be subject to both academic sanctions from the faculty member and non-academic sanctions (including but not limited to university probation, suspension, or expulsion). Other information on the Honor Code can be found at `http://www.colorado.edu/policies/honor.html` and at `http://www.colorado.edu/academics/honorcode/`

Special Accommodations: If you qualify for accommodations because of a disability, please submit to your professor a letter from Disability Services in a timely manner (for exam accommodations provide your letter at least one week prior to the exam) so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities. Contact Disability Services at 303-492-8671 or by e-mail at `dsinfo@colorado.edu`.

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.

If you have a temporary medical condition or injury, see Temporary Medical Conditions: Injuries, Surgeries, and Illnesses guidelines under Quick Links at Disability Services website and discuss your needs with your professor.

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/fac_relig.html`.

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 differences of race, color, culture, religion, creed, politics, veterans status, sexual orientation, gender, gender identity and gender expression, age, disability, and nationalities. 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 at
`http://www.colorado.edu/policies/classbehavior.html`
`http://www.colorado.edu/studentaffairs/judicialaffairs/code.html#student_code`

Discrimination and Harassment: The University of Colorado Boulder (CU Boulder) is committed to maintaining a positive learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct, discrimination, harassment or related retaliation against or by any employee or student. CUs Sexual Misconduct Policy prohibits sexual assault, sexual exploitation, sexual harassment, intimate partner abuse (dating or domestic violence), stalking or related retaliation. CU Boulders Discrimination and Harassment Policy prohibits discrimination, harassment or related retaliation based on race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Individuals who believe they have been subject to misconduct under either policy should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127. Information about the OIEC, the above referenced policies, and the campus resources available to assist individuals regarding sexual misconduct, discrimination, harassment or related retaliation can be found at the `http://www.colorado.edu/odh`