Syllabus for CSCI 7000-003
Inference, Models and Simulation for Complex Systems
Fall 2010

Lectures:

Tuesdays and Thursdays from 11:00–12:15 in MUEN E114

Lecturer: Aaron Clauset

Office:	ECOT 743
Email:	aaron.clauset@colorado.edu
Web Page:	http://tuvalu.santafe.edu/~aaronc/courses/csci7000_2010.htm
Office Hours:	Tuesday 1:30–3:00 or by appointment

Description: This graduate-level topics course will cover a selection of recent developments in computational approaches to doing science with complex systems. It is not a scientific computing course. Topics will include statistical inference, the structure of complex networks, macro-phenomena in biological evolution and in wars and terrorism, simple mathematical models, and simulation techniques for more complicated models. The focus will be on using computational tools (algorithms) to do science (work with data; test hypotheses; build understanding; make predictions).

Prerequisites: No biological or social science background is necessary. A quantitative background is required, for instance, familiarity with basic statistics, calculus, and programming. If you don't have such a background, please see me.

Text: No textbook is required.
Networks: An Introduction by M.E.J. Newman, *All of Statistics* by L. Wasserman and any of the *Numerical Recipes* books are recommended

Coursework and grading: Attendance to the lectures is required.

The first half of the course will be lecture and problem-set driven.

The four problem sets will each have some mathematical and some programming problems. Problem sets are due roughly two weeks after they are assigned. Solutions should in PDF

format (e.g., typeset using LaTeX), should include all necessary details for me to follow the logic, and should be submitted via email by 11:59pm the day they are due. No late assignments will be accepted. Collaboration is allowed on the problem sets, but you're not allowed to simply copy solutions from another student. You must cite your collaborators in your solutions. Copying from any source, including the Web, is not allowed.

The second half of the class will revolve around crowd-sourced lectures on advanced topics and an independent project.

For the lectures, students will team up in small groups, choose a paper or topic related to inference, models or simulations for complex systems, and give a lecture describing the technique and/or results. We'll set the schedule for these early in the semester. Students should form their groups by the end of September and should meet with me soon afterward to choose a paper or topic.

Some lectures will have an accompanying reading assignment. I'll post the readings on the class website prior to the lecture.

The grade for the course will be 30% attendance, 30% problem sets, and 40% class project.

Potential crowdsource lecture topics: Students are free to choose any topic for their crowdsource lecture that is related to the general class topic. I'm happy to make specific suggestions of papers on any of the following topics.

- Statistical inference: the bootstrap, cross-validation, model averaging, regression models, non-parametric tests, statistical power, de-anonymizing SSNs and IP addresses.
- Structure of complex networks: Internet topology evolution, metabolic networks, genetic regulatory networks, road networks, graph alignment, signaling networks, network epidemiology, densification, social networks on the Web.
- Macro-phenomena: Google Insights, patterns in biological diversity, human population growth, Zipf's law, social behavior on the Web.
- Mathematical models: diffusion processes, branching processes, percolation, self-organized criticality, highly-optimized tolerance, random walks, power-law mechansims.
- Simulations: complex models, error bars and uncertainty, model comparison, sensitivity analysis, parameter spaces, parametric vs. non-parametric models, Occam's Razor.

Tentative schedule:

Week 1	Overview, probability distributions, likelihood functions, Poisson processes
Week 2	Power-law distributions
Week 3	Hypothesis tests, model plausibility, model comparison, the Bayesian fallacy
Week 4	Time series analysis, random walks
Week 5	Measures of network structure, algorithms
Week 6	Random graph models, small worlds, citation networks
Week 7	Modular and hierarchical structure, missing information, link prediction
Week 8	Macroevolution
Week 9	Terrorism and wars
Week 10	Crowdsource lectures
Week 11	Crowdsource lectures
Week 12	Crowdsource lectures
Week 13	Crowdsource lectures
Week 14	Fall break
Week 15	Project presentations
Week 16	Project presentations

Assignments:

	assigned	due
Problem set 1	August 26	September 15
Problem set 2	September 16	September 29
Problem set 3	September 30	October 13
Problem set 4	October 14	October 27
Final project write up		December 10

Class project: The assignment is to develop and complete a small research-style project based around the tools and concepts developed in the course. The project should focus on either inference, modeling or simulation. Students will work independently and will turn in a 10-page writeup of their project results at the end of the semester. Students should meet with me outside of class in the first few weeks of October (but no later than Oct. 21) to finalize the project topic.

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.