

MS&E334: Topics in Social Data**Fall 2017**

Units: 3

Grading: Letter or Credit/No Credit

Lectures: 380-381U (Building 380, Room 381U, 1st floor)

Lecture hours: Tuesday/Thursday: 3:00pm-4:20pm

Instructor:

Johan Ugander

Assistant Professor

Management Science & Engineering

Email: jugander @ stanford.edu

Office: Huang 357

Office hours: By appointment

Short description:

This course provides a in-depth survey of methods research for the analysis of large-scale social and behavioral data. There will be a particular focus on recent developments in discrete choice theory and preference learning. Connections will be made to graph-theoretic investigations common in the study of social networks. Topics will include random utility models, item-response theory, ranking and learning to rank, centrality and ranking on graphs, and random graphs. The course is intended for Ph.D. students, but masters students with an interested in research topics are welcome. Recommended: 221, 226, CS161, or equivalents.

Evaluation:

- 2 problem sets (20% each)
- reaction paper (20%)
- project proposal + project (40%)

Prerequisites:

This course is intended for doctoral students and an adequate mathematical and programming background is mandatory. The course will assume a solid background in probability, statistics, algorithms, graphs, and linear algebra, as well as basic programming experience (the ability to manipulate network datasets).

Basic overview:

Week 1: Introduction and Overview

Week 2: Random graph models

Week 3: Graph centrality and ranking

Week 4: Choice models and ranking from comparisons

Week 5: Models of ranking data

Week 6: Social processes, influence maximization

Week 7: Causal inference on networks

Week 8: Friendship paradox, small-worlds

Week 9: Dissecting complex papers

Week 10: Project presentations

Some topics may receive a longer or shorter treatments depending on audience interest at the onset of the course. A detailed list of references will be posted on the course homepage as the course progresses.

Problem Sets:

The first part of the course will feature two problem sets, where the goal is to practice skills that will be employed during the project component. The goal is to familiarize you with (a) computational aspects of working with real data and (b) with analyzing mathematical tools involved in modeling such data.

You may discuss the problem sets with other students in the class, but since the goal is to practice skills, the actual writing up of the solutions must be done separately. In particular, this means that your solution should not word-for-word resemble another student's. In short, students are expected to abide by Stanford's honor code.

**Problem Set 1 will be handed out on October 3 and due on October 12.
Problem Set 2 will be handed out on October 12 and due on October 24.**

Reaction Paper and Project Proposal:

The course includes a lot of material from the past 5-10 years; this means that much of it exists mainly in the form of papers on the web, and the existing literature raises a lot of interesting issues that have yet to be explored.

As a way to get everyone thinking about the research issues underlying the course, and to prepare for the project, there will be a short reaction paper of roughly 5-7 pages in length. You can work in groups of up to 3 people on the paper. The reaction paper is intended to help you formulate a topic for the course project, and your group for the reaction paper can become your group for the project (which also can be done in groups of up to 3).

You should read at least two closely related papers, at least one of which is not linked from the course home page. What is main technical content of the papers? Why is it interesting in relation to the corresponding section of the course? What are the weaknesses of the papers, and how could they be improved? What are some promising further research questions in the direction of the papers, and how could they be pursued? The last two questions should receive the most attention in your paper.

The reaction paper will be due on October 31st.

After the reaction papers have been submitted, each group will schedule a 30-minute meeting with me **between October 31 and November 7** to discuss what they've read and their project ideas. If some group members are more excited about certain questions or ideas than others, or find other students to partner with, it is possible to form new groups (including individual groups) for the project.

A project proposal will be due one week after your scheduled meeting. This is meant to be a brief description of what you're intending for the project -- about 2 pages in length, with a discussion of relevant background work and tentative plans for how you'll proceed. If your project is based on your reaction paper, then you don't need to repeat things you've said in the reaction paper -- it's enough to describe how you're planning to turn the ideas from the reaction paper into a larger project.

Project:

The basic genres of project are the following:

- An empirical evaluation of an algorithm, model, or measure on an interesting dataset. The papers linked from the course home page suggest some possible domains; you can also assemble your own data.

- A theoretical project that considers an algorithm, model, or measure in the area of some course topic, and derives rigorous results about it.
- An extended, critical survey of one the course topics, going into significant depth and offering a novel perspective on the area.

As with the reaction paper, the project should contain at least some amount of mathematical analysis, and some experimentation on real or synthetic data. The result of the project will typically be a 10-15 page paper, describing the approach, the results, and the related work. The references on the course home page will serve as examples of what such papers tend to look like; of course, the overall form of the paper will depend on the nature of the project.

The project paper will be due December 12th, but before that there will be presentations of the projects in class during the last week (**Dec 5 and 7**). The exact schedule for the project presentations will be worked out later in the semester.

Learning outcomes:

1. Students should develop a familiarity with relevant structural properties of empirical social networks, and how different graph models capture or don't capture these properties.
2. Students should be able to weigh advantages and disadvantages of different observational and experimental study designs that examine/test mechanisms in social systems.
3. Students should be able to employ choice models and ranking models under different assumptions about human behavior.
4. Students should be able to critically read research papers in the field to identify strengths and potential weaknesses, and to be able to design tests of potential weaknesses.

Students with Documented Disabilities:

Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty dated in the current quarter in which the request is made. Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations. For more information: <http://studentaffairs.stanford.edu/oae>