

MS&E334: The Structure of Social Data**Fall 2018**

Units: 3

Grading: Letter or Credit/No Credit

Lectures: Gates B12

Lecture hours: Tuesday/Thursday: 10:30am-11:50am

Instructor:

Johan Ugander

Assistant Professor

Management Science & Engineering

Email: jugander @ stanford.edu

Office: Huang 357

Office hours: By appointment

Short description:

Social networks have a rich history of study across many disciplines. Recent opportunities afforded by large-scale online instrumentation and experimentation have begun to provide a rich view of their structure and role in diverse social and economic domains. This course provides a survey of recent research in the study of social networks and large-scale social and behavioral data. Topics will include network models based on random graphs and their properties; centrality and ranking on graphs; ranking from comparisons; social influence and homophily; experimentation and causal inference on networks; heavy-tailed statistical distributions.

Evaluation:

- 2 problem sets (20% each)
- Reaction paper/referee report (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: Graph models

Week 3: Graph centrality and ranking

Week 4: Ranking from comparisons and choice modelling

Week 5: Permutation data, temporal network data

Week 6: Homophily, social influence, contagion

Week 7: Causal inference on networks

Week 8: Friendship paradox, Small-world phenomena

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 social graphs and 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. Submissions should be either PDFs or Jupyter Notebooks. Let me know if there's another format you'd strongly prefer and I may allow it.

Problem Set 1 will be handed out on September 27 and due on October 11.

Problem Set 2 will be handed out on October 16 and due on October 30.

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 pages in length. You should view this assignment as a variation on a "referee report" when peer reviewing (even though the majority of papers are published). Part of the goal is to propose additional directions to take published work. You can work in groups of up to 3 people on this assignment. The reaction paper is intended to help you formulate a topic for the course project, and your group for the reaction paper can (but doesn't have to) become your group for the project (which also can be done in groups of up to 3).

I recommend that you read at least two closely related papers, at least one of which is not linked from the course home page (e.g. something in the bibliography of the first paper), and pick one to center your report on. What is the main technical content of the paper? Why is it interesting in relation to the corresponding section of the course? What are the weaknesses of the paper, 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 November 6th.

After the reaction paper has been submitted, each group will schedule a 30-minute meeting with me (through Canvas) **between November 6 and November 12** to brainstorm about 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 brief project proposal will be due one week after the scheduled meeting. This is meant to be a brief description of what you're intending for the project – 1-2 pages in length, with a 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. Permission to do “one or the other but not both” is likely to be granted, but please ask me. 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 11th, 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 structural measures for diverse ranking/predicting problems on graphs.
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>