| **CS 134: Network Science** | Syllabus |
| --- | --- |
| Prof. Michael Mitzenmacher and Prof. Yaron Singer | Fall 2018 |

# 1 Overview

This course focuses on analyzing, predicting, and engineering processes in networks. Throughout the course we will use fundamental concepts from computer science, probability, applied mathematics, microeconomics, and mathematical sociology so that students are equipped to analyze processes on large network data sets effectively.

Topics discussed include graph models, the small world phenomena, richer-get-richer effects, power-laws, sampling biases in data, contagion and influence, ranking in the web, and community detection. Concepts and tools covered include basic graph theory, random graphs, distributed algorithms, statistical machine learning, optimization, approximation algorithms, spectral analysis, and clustering.

# 2 Basic Information

**Contact Information:**


- **Professor Michael Mitzenmacher**
  Email: michaelm@eecs.harvard.edu
- **Professor Yaron Singer**
  Email: yaron@seas.harvard.edu


**Course homepage:** Most information about the course can be found on the course Canvas site.

# 3 Prerequisites

To enjoy and succeed in this course, you will need to be comfortable with some basic math and programming. We will assume familiarity with probability theory, calculus, linear algebra, and the basics of theoretical computer science (e.g. complexity theory, asymptotic runtime analysis), as well as basic proficiency in Python.

Courses that provide good preparation for calculus and linear algebra would be Math 21a or 23a, discrete math at the level of CS 20, and basic probability at the level of Stat 110. Algorithms at the level of CS 124 would be helpful but not necessary. For programming, you should have basic programming knowledge at the level of CS 50 and know how to write multi-file programs in Python.

We have prepared a diagnostic quiz that is available on the course Canvas page. **Submission is mandatory.** Any submission (even an empty one) will receive full credit. We encourage you to

complete all the questions on the quiz. This will allow you to (a) make an honest self-assessment that will help you determine whether you have the sufficient background to take the class and (b) refresh basic concepts that will be used in class. Solutions will be published after the submission deadline. You should, of course, feel free to discuss the quizzes with the TF's.

## 4 Learning Objectives

**The language and basic tools of networks** On successfully completing the course, students should have a solid grasp of the vocabulary, central concepts, and some basic facts relating to networks. The study of networks draws on a variety of disciplines: mathematics, statistics, computer science, economics, and sociology. Thus, students should be fluent in combining and reasoning about terms and ideas coming from each of these.

**Puzzles and big ideas.** The study of networks is full of great ideas that can be summarized in a short story form. We aim to give students the power to tell these short stories well to their colleagues and friends when they're relevant, and to apply them in research and business.

**Measuring and modeling networks in the wild.** The highest learning outcome will involve students combining multiple concepts they learn to address practical challenges that come up in measuring and modeling networks.

**Beyond networks.** Although the entire course is focused on networks, the morals and tools generalize to data science at large. After completing the course, students should feel comfortable with modeling real world phenomena, analyzing large data sets, designing and applying algorithms in probabilistic models, and predicting outcomes of complex processes.

## 5 Logistics

**Lecture notes.** Lecture notes will be released after class and will summarize definitions and main ideas. In general, we strongly encourage you to attend lecture, as the lecture notes are not necessarily self-contained and do not aim to substitute for class attendance.

**Recordings.** Lecture and section recordings will be available to Extension school students on the Extension Canvas page.

**Problem sets.** There will be weekly problem sets. Problem sets will typically be released on Wednesday evenings and will be due *the following Wednesday* at 11:59:00 AM **sharp**. Solutions to problem sets should be submitted via Canvas. Solutions to problem sets will be posted shortly after the deadline. Submission deadlines are **strict**, and submissions submited after 11:59:00 AM will receive no credit. Except for unusual circumstances we will not accept late submissions. To

prove that we're not all evil however, we will drop the lowest score of your problem sets (thus you can drop one problem sets and still earn a perfect score on the problem sets).

**Programming.** Most problem sets will include programming assignments. The programming will be relatively light, where the idea is to use simple scripts to analyze real-world data sets and apply some of the algorithms you learn in class. You're expected to have taken CS 50 or have similar background and experience. You are expected to know how to code in Python. We will not teach Python or programming related material, but we will cover programming in some sections to help you with the coding exercises on the problem sets.

**Homework Canvas submission.** Solutions to all non-coding problems should be submitted in a PDF file uploaded to the corresponding assignment on Canvas. For the coding problem, unless specified otherwise, you should include a short 1-2 paragraph write-up describing your results in your PDF file, and also upload a Python file code submission to Canvas. **No iPython notebooks will be accepted.** While you are encouraged to use Latex to write your solutions, the staff will also accept PDF files containing scanned images of your write-ups. Note that any submissions that are deemed illegible or unacceptably messy by the course staff will receive a zero.

**Sections.** We will have sections on a weekly basis taught by the teaching fellows. Sections will include exercises relevant to the problem sets, and also introduce coding concepts that will play a role in the programming assignment for that week. Sections are not mandatory, but we strongly encourage you to participate. Section times and locations will be determined in the first week of the coruse. **We will aim to schedule sections to take place on Wednesday and Thursdays**.

**Exams.** There will be two exams (midterms) taken in class. There will not be programming questions *per se* on these exams, but there will be related questions (writing pseudocode, using insights from programming exercises completed in the course, etc.).

**Grading.** The grading for the course consists of differently-weighted pieces:

- 5%: class participation;
- 45%: weekly problem sets (lowest score pset dropped; no excused lateness)
- 20%: first exam
- 30%: second exam

**Piazza.** We will be using Piazza as a tool for students to ask questions about the material covered in class and about assignments. Of course, students also have the opportunity, and are encouraged, to go to office hours to ask questions.

## 6 Resources

There will be comprehensive lecture notes that will be released after class. In addition, students may find it useful to use relevant textbooks.

**Textbooks.** Most of the material of the course can be found in:

- *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* by David Easley and Jon Kleinberg, Cambridge University Press, 2010. A version of the book is available for free online:

  `http://www.cs.cornell.edu/home/kleinber/networks-book/`

Another recommended, textbook:

- *Social and Economic Networks* by Matthew Jackson, Princeton University Press, 2008.

**Additional resources:** We will also post additional material on the course Canvas page as the course progresses.

## 7 Course outline

**Preliminaries**

- Graph definitions and basic properties
- The Karate Club network
- Randomized algorithms
- Random graphs

**Small world networks**

- Small-world phenomena
- The structure of small world networks
- Navigation in a small world

**Structure of networks**

- Power law and heavy tailed degree distributions
- Preferential attachment and the rich-get-richer model

**Statistical biases in networks analysis**

- The friendship paradox, or why your friends have more friends than you
- The friendship paradox in power-law networks

**Strategic interactions**

- Game theory definitions and basics
- Solving games: elimination of dominated strategies and Nash equilibrium
- Network games

**Influence in networks**

- Models of social influence in networks

**Diffusion of information**

- Cascading behavior in networks
- Random walks in graphs
- Influence maximization

**Link analysis and web Search**

- Spectral Analysis of Networks
- Pagerank, HITS algorithms

**Clustering and communities in networks**

- Unsupervised machine learning
- Clustering in networks
- Triangles in graphs

## 8 Action items

- **Diagnostic quiz:** Submission of the diagnostic quiz on Canvas is mandatory, though all submissions will receive full credit. If you have any doubt as to whether your background in a particular area (math, programming, economics) suffices, you are encouraged to take the diagnostic quiz to see whether this course is a good fit for you. The diagnostic quiz is due **Wednesday, September 12th, 2018 AD at 11:59:00 AM**. The diagnostic quiz will be posted on Canvas by Wednesday September 5th.
- **Problem set due in one week:** The first problem set is due **Wednesday, September 12th, 2018 AD at 11:59:00 AM**. The problem set will be posted on Canvas by Wednesday September 5th.