

STAT-GB.4310: Statistics for Social Data

Description

Statistical methods for describing and utilizing nontraditional data modalities arising from social processes. Half of the course will be devoted to network data (communications, recommendations, and transactions) and the other half will be devoted to textual data (words, phrases, and documents).

Objectives

The objectives of the course are as follow:

- to learn about approaches for describing and utilizing network and text data;
- to understand these approaches from within an inferential statistical framework;
- to get hands-on experience in applying these methods, through the homework assignments and the class project.

Prerequisites

The course will assume some prior knowledge of linear algebra, probability, and statistical inference. You should be familiar with eigenvectors, likelihoods, and confidence intervals. Homework assignments will require a small amount of programming in Python and R; you can take the course if you know one of these programming languages and you are willing to learn the other.

Course staff

Instructor: Patrick Perry
E-mail: pperry@stern.nyu.edu
Office: KMC 8-63
Office Hours: TBD

Meeting time and place

Tuesdays, 2:00–5:00, in the Stat/OM conference room (KMC 8-170).

Course webpage

<http://ptrckpry.com/course/ssd>

Homework assignments

There will be regular homework assignments. Each assignment will involve applied data analysis and theoretical problem solving.

Readings

TBD. Some mixture of textbook chapters and research articles.

Tentative schedule

Date	Topic
2/02	Case study: <i>The Federalist</i>
2/09	Phrase and sentence segmentation
2/16	Markov models for text
2/23	Matrix decompositions
3/01	Topic models
3/08	Vector space text representations
3/15	<i>Spring Break (no class)</i>
3/22	Sentiment analysis
3/29	Exponential random graph models
4/05	Latent space network models
4/12	Point process models for interactions
4/19	Community detection
4/26	Network sampling
5/03	Network experiments