G8325 Fall 2015 Topics in Advanced Statistics

Analyzing and Modeling Relational Data

Department of Statistics, Columbia University

Course Information

- Thursdays 2:30-4:30 PM, Room 1025 SSW
- Instructor: Tian Zheng (Office hours: Mondays 1:30-2:30 PM, or by appointments; Room 1007, SSW). Email: tian.zheng@columbia.edu
- Course website: http://courseworks.columbia.edu

Description

Relational data to refer to numerical data that describe the relations among a collection of entities. The relations can be dichotomous, discrete-valued or continuous-valued. The most common type of relational data is a network, described by a list of egos and an adjacency matrix among the egos. In this course we will explore current statistical models and methods for relational data, especially on network analysis. we will cover major models for relational data/networks, collection and estimation/inference of network data, visualization, major applications such as in social network analysis and biology. Analyzing and modeling real complex relational data sets will be part of this course.

Textbook

No textbook required. We will use lecture notes and relevant journal articles, which will be posted on courseworks.


Assignments and grading

Each registered student is required to give an in-class presentation and submit a final report based on a literature review + analysis/modeling project that includes a review of a current topic in the literature and application of related methods to a real data set. The presentations will be scheduled in the 2nd half of the semester. Grading is based on this project and in-class participation. Students can choose any statistical computational package. Lecture examples will be given using R.