Final Project Proposal

Calvin Walker

Problem: Classic methods for causal inference in economic and other social science research often rely on the Stable Unit Treatment Value Assumption (SUTVA), which requires that the response of an individual only depends on the treatment to which they were assigned, i.e. if a unit's friend is in the treatment group, their treatment only effects them. However, there are numerous situations of interest to researchers where this may not be a plausible assumption, and being able to separate the effects of the treatment assigment and peer influence have important implications.

Graphs are a common, and intuitive way to express social interactions between individuals—offering an attractive model for learning, and properly acounting for peer influence in experimental settings. There is a growing body of literature ... *dicuss current approaches here*
research questions

Approach: My approach to the problem will be to model peer influence using a probabilistic graphical model such as a markov random field, where edges and potentials represent social connection and influence between individuals in the graph. I will investigate ways to learn the potentials in this graph, and how to use this learned graph to estimate the effect of peer influence on the outcome of interest. Formulating this explicitly will most likely be the greatest technical challenge, but it seems achievable using the course material.

Evaluating Success: I will first evaluate the performance of my approach by testing it on a semi-synthetic dataset, which is common in the related literature. In practice, this will mean taking an existing dataset of a social network/graph, and simulating treatment effects and peer influence using some data generating process, so see if my approach can recover the true effects of treatment and peer influence. Next, I hope to test my approach empirically on experimental data, but I am still searching for a suitable dataset. The only software I expect to use is Python for data manipulation and learning.

Timeline: Given that there is about a month remaining before the final project presentations, I hope to spend the next week continuing my literature review, choosing both datasets, and finalizing the general structure of my approach to the problem. Then, the next couple weeks will be spent explicitly defining my solution, and running the necessary experiments on the data to test its efficacy. Finally, the last week will be spent fixing errors, and summarizing my findings.