Project Proposal
Causal Inference II Spring 2022

February 18, 2022

Team

Reetahan Mukhopadhyay, rm3873

Working Title

Empirical Application of Estimating Peer Influence on Social Networks Using Embeddings

Topic

Applied Project | Contagion, Interference, and Social Networks

Problem Statement

Input: Datasets that represent some form of a social network of individuals, which includes data on, broadly, a type of decision they have made.
Output: An estimate of the direct influence that peers had on individuals in those networks.

This project aims to identify the direct effect that the neighbors of an individual has on the outcome of that individual in real-world social networks. Data will need to be acquired to represent some observed social networks and some sort of observed "decisions" that individuals in the network have made, and based on this, we can generate a causal estimand of the average outcome for an individual had the "treatment", or the decisions of their neighbors, been set to a particular outcome. This project will develop code that can execute the necessary procedures to perform upon the data to acquire this estimand, and report the results such as the estimand, confidence intervals for this estimand and other possible related data, on the acquired datasets.

This requires doing a backdoor adjustment to account for the effect that homophily, the latent traits the people who are connected in a social setting may tend to share, may have on the decisions they make. This can be accomplished by introducing embeddings of the individuals in the social network, into the causal model.

Preliminary Work

In [1], the authors attempt to estimate the causal effect that an individual neighbor's have upon the decisions of an individual in a social network. The key issue they attempt to overcome is the effect that the unobserved confounder that homophily has on the individual's decisions. Broadly, they develop a non-parametric approach to perform the necessary backdoor adjustment, that does not rely on specific information about the form of the graph underlying the network, and of course, the treatment assignment and outcome processes. This involves learning an embedding λ of the nodes in our network into $\mathbb{R}^k$ that they use to block the problematic backdoor path. The paper accomplishes several theoretical goals in the approach to this problem. First, it formalizes a causal estimand ψ_n that, averaged for every individual i in a population of size n, looks at the expected outcome Y_i for an individual for a given network G_n, a given treatment (decision of peers) $T = t^*$, and given unobserved attributes of an individual that may impact an individual's decision C_i. It proves the validity of this estimand by showing that ψ_n converges to ψ in probability under certain conditions. It also proves the causal identification of this effect provided a black box embedding $\lambda_i : i \in G_n \to \mathbb{R}^k$. Finally, it provides a possible method for actually providing this estimation.

However, this paper is preliminary work that does not provide experimental results (and it does note this). They also note one missing piece in the theory is that they do not demonstrate asymptotic normality for the estimator. In addition to the experimentation, I would need to complete the theory by demonstrating this, which would then allow for doing inference and constructing confidence intervals for the estimator, using normal approximations. Additionally, I would consider refining or altering the process they developed in section 5 for getting this estimator, possibly around specifying the embedding methodology. Essentially, this project is essentially meant to build upon and complete the work that [1] sets out to accomplish.

Additionally, I have read additional papers to provide additional background and perspectives on the matter that are listed in the references section such as [4], [5] and [2], including some of the key references that [1] builds upon or contrasts their approach to, such as [3] and [6].

Personal Background

The broad area of studying effects in social networks, communities, influence and contagion is an interest of mine. I took the Intro to Networks and Crowds course with Professor Augustin Chaintreau last semester, did very well and enjoyed the course. I started work on a thesis project with the professor this semester, and we are currently looking at potential bias and unfairness in multi-hop referral systems. Additionally, I took Unsupervised Learning with Professor Nakul Verma last semester in which there was heavy discussion as well as problem sets regarding different types of embeddings that I think may be useful. Overall, I believe integrating the power of causality tools into this field can be instrumental for better understanding some of the network effects that we have observed.

References

[1] Veitch Cristali. Using Embeddings to Estimate Peer Influence on Social Networks. In *WHY-21 Workshop at NeurIPS 2021*, 2021.

[2] Ruocheng Guo. *Learning Causality with Networked Observational Data*. PhD thesis, Arizona State University, 2021.

[3] Shalizi McFowland III. Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations. *Journal of the American Statistical Association*, June 2021.

[4] Peixoto. Disentangling Homophily, Community Structure, and Triadic Closure in Networks. *Physical Review X*, January 2022.

[5] Sweet Spencer, Junker. Faster MCMC for Gaussian Latent Position Network Models. November 2021.

[6] Blei Veitch, Wang. Using Embeddings to Correct for Unobserved Confounding in Networks. December 2019.