Causal Effects Under Interference

Calvin Walker
University of Chicago
cswalker1@uchicago.edu

Abstract

In randomized experiments, it is often assumed that the response of a given unit only depends on its own characterisitcs and the treatment to which it is assigned. However, there are numerous situations of interest to researchers where this may not be a plausible assuption, particularly when there is social interaction between individuals in an experiment, and the ability to separately identify the effects of treatment assigment and peer influence have important implications. This paper presents a general framework for reliably identifying and estimating causal effects in experimental settings where there is potential interference between units. The framework consists of (i) learning the potential for interference between connected units, (ii) a probabilistic graphical model for estimating individuals' exposure to treated units, and (iii) methods that make use of estimated exposures to compute causal effects of interest. We then evaluate the performance of the proposed framework on by simulating synthetic experimental data on a variety of real world social networks.

1 Introduction

Reliable causal inference is central to a variety of diciplines in the natural and social sciences. While the Randomized Control Trial (RCT) has long been the gold standard in terms of causal inference, central to it is an assumption in the economic literature first coined by Rubin [1990] known as SUTVA, or the stable unit treatment value assumption, which states that there is no interference between units, and no hidden variations of treatment. Essentially, the outcome of each individual does not depend on the treatment assigment of others. When this assumption is violated, there are said to be spillover effects. And it often is. In an RCT on 61 million Facebook Users in the buildup to the 2020 United States Congressional elections, the treatment—sending a message that encouraged users to vote—was observed to "spillover" to the treated individual's closest Facebook friends Robert M. Bond [2012].

A number of recent works have sought to address estimation of both treatment and spillover effects in the presence of SUTVA violations, particularly in the context of an observed social network between individuals. Toulis and Kao [2013] seek to estimate spillover effects directly, developing both an experimental design using sequential randomization, and a Bayesian procedure that assumes a linear response. Other approaches consist of the researcher specifying an "exposure mapping" to categorize individuals, e.g. treated, neighbor of treated, etc. and then use inverse probability weighting based on the probability of assignment for each individual Aronow and Samii [2017]. Due to the inherent graphical nature of the problem, this paper seeks to examine how a probabilistic graphical model may be well suited to capture interference and spillover effects in a social network. We develop a flexible framework that defines a probabilistic graphical model over the social network, and uses this model to estimate unit level exposure to treated units, allowing us to compute causal effects of interest.

2 Setting and Assumptions

We consider the setting of a randomized experiment on a social network of N nodes. The network, $G = (V, E)$, is observed, and may or may not contain node-level covariates or edge strengths. Let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_N)$ be a treatment assigment vector over N units, where $Z_i \in \{0, 1\}$ specifies which of the possible treatments unit i receives. We assume that $P(Z_i = 1)$ is the same for all $i \in \{N\}$, so treatment assigment is unconfounded. We define $\mathbf{C} = (C_1, C_2, \ldots, C_N)$ to be a latent exposure vector, where $C_i \in \{0, 1\}$ specifies if unit i is exposed to the treatment via social interaction. Here, we assume that all treated units are exposed, such that $Z_i \times C_i = 1$, but that treatment units do not experience further spillover effects. If we could observe $\mathbf{C}$, then our work would be done, since we could partition the units into treatment, exposure, and control groups, and employ classic techniques to infer their differences in outcome.

Instead, we define $\pi_i = P(C_i = 1)$ to be the expsoure probability for unit i given some instantiation of the experiment $\mathbf{Z}, \mathbf{C}$. Since an individual's social exposure to the treatment depends on the random assigment in G, so does π_i. For simplicity, we assume that the effect of treatment assigment is homogeneous. Furthermore, we assume that social interaction, and thus social exposure to the treatment, only occurs through the edges E of G. However, we do not assume that spillover effects are homogeneous, and examine both this and the heterogeneous case in later sections.

3 Estimating Exposures and Causal Effects

We propose modeling the spread of influence in the social network as a Pairwise Markov Random Field, defined as $B = (P, H)$, where $H = (V, E)$ is an isomporphism of the social network G, given by the bijection $f : G(V) \mapsto H(V)$ where $f(v_i) = X_i \sim$ Bernoulli(p), i.e. H is an undirected graph with the skeleton of G where each node in $H(V)$ is a Bernoulli random variable. P, then, is defined as a Gibbs distribution that factorizes over H:

$$P(X_1, X_2, \ldots, X_n) = \frac{1}{Z} \prod_{i \in V} \phi_i(X_i) \prod_{(i,j) \in E} \phi_{i,j}(X_i, X_j)$$

By having each individual represented as a bernoulli random vairable in H, we can reformulate the problem of estimating the exposure probability for node i, $\pi_i(\mathbf{Z})$, into an inference problem, where we seek to infer marginal probability distributions for each node. Observe that for any treated unit, they are assumed to be already exposed, so $X_i = 1$ for all individuals i where $Z_i = 1$. This observation acts as our evidence in P when performing inference on the unknown marginal distributions of the other nodes in H. Formally, we infer the posterior distribution $P(X_i \mid e)$, where $e = \{X_i = 1 \mid Z_i = 1\}$. This reformulation allows us to use a host of existing inference algorithms to learn the posterior marginals. In practice, we use the Sum-Product Belief Propogation algorithm defined in Daphne Koller [2009]. In larger social networks where exact inference may be intractable, approximation alogrithms can be used, although we do not examine this case. Since nodes not influenced by the treatment group have the posterior marginal distribution: $P(X_i = 1) = P(X_i = 0) = 0.5$, we let $\pi_i = P(X_i = 1) - P(X_i = 0)$ to obtain easily interpretable exposure probabilities.

3.1 Unary and Edge Potentials

Having refomulated estimating $\pi_i(\mathbf{Z})$ as inference of marginals in a Pairwise MRF, properly defining the edge and unary potentials in H becomes an important specification. In this section, we dicuss several potential approaches. In the setting considered by this paper, treatment assigment is unconfounded, so the unary potentials $\phi_i(X_i)$, are uniform across the nodes such that $\phi_i(X_i = 1) = \phi_i(X_i = 0)\ \forall i$. However, in the case that treatment assigment is, for instance, confounded by covariates, it may make sense to use the unary potentials as a prior on similarity with the treatment group, since influence may flow to "similar" nodes more easily.

The edge potentials $\phi_i(X_i, X_j)$ have an even more intuitive interpretation, representing the potential for influence, or "strength" of a social connection between two nodes in H. Here, we propose two possible approaches for specifying the edge potentials. Both involve first specifying the probability of each edge in the social network. The first is that, given domain knowledge within the experiment, the researcher may have a prior on the distribution of edge probabilities, and can make use of this prior

belief. For instance, let $p_{i,j}$ be the probability of an edge between node i and node j. A reasonable prior could be $p_{i,j} \sim \text{Beta}(5, 5)$. This belief could be incorporated by initializing the edge potentials according to this distribution. The next is a latent variable model introduced by Handcock et al. [2007]. Where we model $p_{i,j}$ using a logistic regression where the probability of an edge depends on euclidean distance in latent space:

$$\text{log-odds}(p_{i,j}) = \beta X_{i,j} - ||z_i - z_j||^2 \tag{1}$$

Where z_i and z_j are node i and node j's respective positions in latent space, and $X_{i,j}$ is some vector valued edge covariates. Here, it is assumed that the existence of an edge is independent of other edges. Handcock jointly estimates the parameter β and the latent z_i's using MCMC, with prior:

$$z_i \sim \sum_{g=1}^{G} \lambda_g \text{MVN}_d(\mu_g, \sigma_g I_g)$$

Where G is the possible number of latent social clusters. Since we can take $G = 1$, the latent variable model provides a flexible and well studied approach to estimating the probability of edges between nodes in a variety of experimental settings. We test both possible approaches to estimating edge probabilities in the following section.

Given the edge probabilities $p_{i,j}$, we can specify the the edge potentials in H. In doing so, we make the assumption that a higher edge probability leads to greater social influence, and thus likelihood that two adjacent nodes have the same value, i.e. $P(X_i = X_j)$ is monotonically increacing in $p_{i,j}$. To acheive this, we define the edge potentials as:

$$\phi_{i,j}(X_i, X_j) = 1 - p_{i,j} \mathbf{1}\{X_i \neq X_j\}$$

We found that this specification resulted in a plausible empirical results across a number of different social networks, but this is another place where a researcher may choose to incorporate prior beliefs in specifying how social influence propogates in a specific network.

3.2 Estimating Causal Effects

With a distribution over the exposure probabilities for each node, we can turn to the problem of estimating the causal effects of interest. First, we are interested in estimating the average treatment effect, ATE. A naïve estimate of ATE in our setting would be a simple difference of means between the treatment and control groups. However, as discussed, in the presence of spillover effects, this estimate is that of both ATE, and whatever spillover effects may be present in the experiment. Instead, we propose the OLS regression:

$$y_i = \alpha + \rho Z_i + \gamma(1 - Z_i)\pi_i + \varepsilon_i \tag{2}$$

Where ρ is the parameter of interest. Since $\pi_i = 1$ if $Z_i = 1$, we add the interaction term $(1 - Z_i)$, so that the causal effect of treatment assigment is only contained in ρ. Unfortunatley, the coefficient γ does not have much causal interpretation, since π_i is merely the probability that unit i is socially exposed to the treatment. In theory, the above regression model, and our approach in general, is better suited to the heterogeneous spillover effects case, since potential outcomes are linear in π_i. In this case, we could interperent $\gamma \times \pi_i$ to be the partial effect of exposure for unit i.

In the case of homogeneous spillover effects, what we really desire is to be able to classify each non-treated unit into either the socially exposed ($C_i = 1$), or control group ($C_i = 0$), as doing so would allow for perfect identification of average spillover effects. There are a number of possible approaches. One might be assigning units with exposure probability sufficiently close to zero to the control group, and having a similar condition for assigment to the expsoure group. Another may be to explicitly learn a classifier. If treatment assigment were confounded by latent or observed covariates, this may be possible without taking the outcome of interest into account, since the treatment and exposure groups would have greater similarity than the controls. However, the setting of this paper assumes that treatment assigment is unconfounded. Under the assumption that the joint distribution of outcomes and exposures, $P(y, \pi)$ is a mixture of two Gaussians—the control and exposure groups—we propose fitting the Gaussian Mixture Model:

$$P(y, \pi | \mu, \Sigma, w) = \sum_{k=1}^{2} w_k \mathcal{N}(y, \pi; \mu_k, \Sigma_k) \tag{3}$$

And then estimating the average spillover effect as the difference in means: $\bar{y}^{k=2} - \bar{y}^{k=1}$, where $k \in \{1, 2\}$ indexes the control and expsoure groups, respectively.

4 Simulations

In order to test the performance of our proposed methods, we simulate random experiments on two real world social networks. The first is a group of 55 eigth-graders, with edges between students who were surveyed on which other students they would like to sit next to in class. The second data set is a social network between 61 employees of the Aarhus Computer Science Department, where edges represent colleagues who ate lunch together in a given week. We simlulate the outcome of interest drawing from a normal distribution $y_i \sim \mathcal{N}(\mu, \sigma^2)$, and then consider a dilated effects scenario, where spillover is half of the average treatment effect. For each graph, we consider three possible data generating processes for the spillover effects. In the first, all neighbors of treated units are socially influenced. In the second, social influence is propogated across each edge (i, j) with probability $p_{i,j}$ according to the latent variable model (1), i.e. socially influenced neighbors of treated units can influence their neighbors etc. until all possible influence has propogated across the network. The third is the same as the second process except $p_{i,j} = 0.5$ for all edges in the network. We also test two different spesifications for the edge potentials in each scenario. In the first, the edge potentials accord with the latent variable model (1), and in the second we assume that $p_{i,j} \sim \text{Beta}(5, 5)$. The following table report the point estimates of the OLS regression (2) and corresponding standard errors following 1,000 simulated treatment assigments on each of the possible combinations of data generating processes, edge potentials, and social networks.

Table 1: Simulation Results (True parameter: $\rho = 3$)

		Eighth Graders		Aarhaus CS	
DGP	$\phi_{i,j}$	$\bar{y}^1 - \bar{y}^0$	ρ	$\bar{y}^1 - \bar{y}^0$	ρ
Neighbors	Latent	2.67 (0.22)	3.03 (0.30)	2.31	3.03 (0.43)
	Beta	2.66 (0.22)	3.14 (0.38)	2.32	3.63 (0.63)
Latent	Latent	2.73 (0.22)	2.93 (0.33)	2.55	2.93 (0.43)
	Beta	2.72 (0.22)	2.95 (0.35)	2.54	3.08 (0.61)
$p_{i,j} = 0.5$	Latent	2.55 (0.22)	2.67 (0.38)	2.31	2.48 (0.46)
	Beta	2.56 (0.22)	2.81 (0.40)	2.32	2.64 (0.67)

In general, the point estimates for ρ are certainly an improvement over the naïve estimate $\bar{y}^1 - \bar{y}^0$. However, performance was quite varied across the different data generating processes. In particular, the model was not well suited to capture the effects of the third DGP ($p_{i,j} = 0.5$), even in the case of the Beta prior on the edge potentials, which had quite high Standard Errors for the Aarhaus Computer Science dataset. Still, in the first two data generating processes, which are quite different from one another, our model is able to reliably detect the presence, and extent of spillover effects in the data, isolating the average treatment effect as desired. Next, we test performance when spillover effects are heterogeneous, which is the natural setting for our model. In the following table, we report results from an identical simulation as above, instead now spillover effects are propogated across each edge $(i - j)$ proportional to $p_{i,j}$ according to the latent variable model (1). We do this by conducting a breadth first traversal over the graph where the starting queue is the treatment group.

Table 2: Heterogeneous Spillover Effects Simulation Results (True parameter: $\rho = 3$)

	Eighth Graders			Aarhaus CS		
$\phi_{i,j}$	$\bar{y}^1 - \bar{y}^0$	ρ	γ	$\bar{y}^1 - \bar{y}^0$	ρ	γ
Latent	2.74 (0.21)	2.95 (0.25)	1.40 (0.39)	2.31	3.03 (0.43)	
Beta	2.75 (0.21)	2.99 (0.28)	0.84 (0.40)	2.32	3.63 (0.63)	

As expected,

5 Conclusion

This paper proposes a probabilistic graphical model based framework for causal inference under interference in social networks.

References

Donald B. Rubin. Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference*, 25(3):279–292, 1990. ISSN 0378-3758. doi: https://doi.org/10.1016/0378-3758(90)90077-8. URL https://www.sciencedirect.com/science/article/pii/0378375890900778.

Jason J. Jones Robert M. Bond, Christopher J. Fariss. A 61-million-person experiment in social influence and political mobilization. *Nature*, (489):295–298, 2012. doi: https://doi.org/10.1038/nature11421. URL https://www.nature.com/articles/nature11421.

Panos Toulis and Edward Kao. Estimation of causal peer influence effects. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1489–1497, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL https://proceedings.mlr.press/v28/toulis13.html.

Peter M. Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912 – 1947, 2017. doi: 10.1214/16-AOAS1005. URL https://doi.org/10.1214/16-AOAS1005.

Nir Friedman Daphne Koller. *Probabilistic Graphical Models: Principles and Techniques*. Massachusetts Institute of Technology, 2009.

Mark S. Handcock, Adrian E. Raftery, and Jeremy M. Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2): 301–354, 2007. doi: https://doi.org/10.1111/j.1467-985X.2007.00471.x. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2007.00471.x.