Network Exposure to Multiple Universes

Johan Ugander[1], Brian Karrer[2], Lars Backstrom[2], Jon Kleinberg[1]

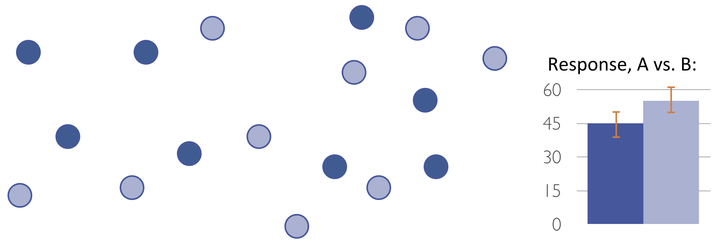
(1) Cornell University (2) facebook

Abstract

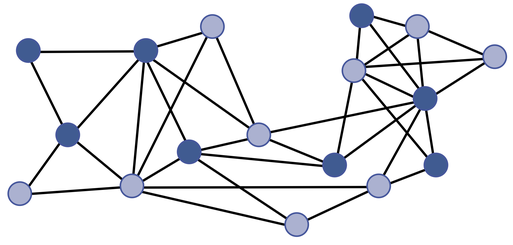
A/B tests are a standard approach for evaluating the effect of a product or feature on a set of individuals – to estimate the so-called unit-level average treatment effect. Here we propose a methodology for computing the unit-level average treatment effect of A/B tests under social interference. We characterize formal conditions under which users are considered to be 'network exposed' to an experiment, along with a graph randomization scheme based on vertex clustering that yields non-vanishing probabilities of network exposure for users. We provide algorithms for efficiently computing these exposure probabilities, and show how a Horvitz-Thompson estimator can provide an unbiased effect estimate. We develop an understanding of the variance of this estimator through sufficient conditions on its asymptotic smallness, and construct an example in which our randomization scheme requires a non-trivial optimal vertex cluster size of 3 vertices.

A/B testing with and without interference

Assuming no interference, exposure condition is simply treatment condition:



With interference, exposure can be a complex function of everyone's treatment:



Average treatment effect between universes

Our **underlying goal** is to compute the average treatment effect between the universe where 'everyone is treated', the treatment universe, and the universe where 'no one is treated', the control universe. Some **notation**:

Let $\vec{z} \in \{0,1\}^n$ be the treatment assignment vector, where $z_i = 1$ means that user i is subject to a treatment intervention and $z_i = 0$ means they are in the control. Let $Y_i(\vec{z}) \in \mathbb{R}$ be the potential outcome of user i under the treatment assignment vector $\vec{z}$. The fundamental quantity we are interested in is the average treatment effect, τ, between the two diametrically opposite universes $\vec{z} = \vec{0}$ and $\vec{z} = \vec{1}$,

$$\tau(\vec{z} = \vec{1}, \vec{z} = \vec{0}) = \frac{1}{n}\sum_{i=1}^{n}\Big(Y_i(\vec{z} = \vec{1}) - Y_i(\vec{z} = \vec{0})\Big).$$

We define the set d_i^k of $\vec{z}$'s that map a user i into a particular potential outcome (with outcomes indexed by k) for that user as an 'exposure condition'. We focus on the sets d_i^1 and d_i^0 that we define to include $\vec{z} = \vec{1}$ and $\vec{z} = \vec{0}$ respectively.

Exposure models

The **treatment condition** of an individual decides whether or not they are subject to an intervention, e.g. `treatment' or `control'. In most experiments, we have explicit control over how to randomize treatment conditions.

Meanwhile, the **exposure condition** of an individual determines how they experience the intervention in full conjunction with how the world experiences the intervention.

Without SUTVA: 2^n possible exposure conditions for each user! Simplifying assumptions:

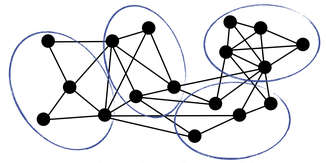
Full neighborhood exposure: *A vertex experiences full neighborhood exposure to a treatment condition if they and all their neighbors receive that treatment condition.*

Absolute k-neighborhood exposure: *A vertex of degree r, where $r \geq k$, experiences absolute k-neighborhood exposure to a treatment condition if they and $\geq k$ of their neighbors receive that treatment condition.*

Fractional q-neighborhood exposure: *A vertex of degree r experiences fractional q-neighborhood exposure to a treatment condition if they and $\geq qr$ of their neighbors receive that treatment condition.*

Clustered randomization

Perform graph clustering, and randomize treatment IID at the level of clusters (while still seeking unit-level response). Procedure is algorithm agnostic. Given a clustering, we can compute probabilities of fractional neighborhood exposure using a simple **dynamic program**:



Proposition 1 *The probability that vertex i is treated and $\geq k$ neighboring vertices are treated under independent cluster randomization is given by* $\Pr[Z \in d_i^1] = pf(\ell{-}1, k{-}w_{i\ell}; p, \vec{w})$ *where*

$$\begin{aligned} f(0, T; p, \vec{w}_i) &= 1 - p\mathbf{1}[T < w_{i0}], \\ f(j, T; p, \vec{w}_i) &= pf(j-1, T - w_{ij}; p, \vec{w}_i) + (1-p)f(j-1, T; p, \vec{w}_i). \end{aligned}$$

The probability that vertex i is in control and $\geq k$ neighboring vertices are in control under independent cluster randomization is given by $\Pr[Z \in d_i^0] = (1-p)[1 - f(\ell{-}1, r_i{-}k{+}1; p, \vec{w})]$.

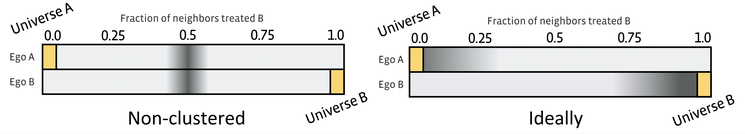
We can now use these probabilities in a **Horvitz-Thompson estimator of the ATE** [1,2]:

$$\hat{\tau}(Z) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i(Z)\mathbf{1}[Z \in d_0^i]}{\Pr(Z \in d_0^i)} - \frac{Y_i(Z)\mathbf{1}[Z \in d_1^i]}{\Pr(Z \in d_1^i)}\right).$$

Exposure probability distribution

IID cluster-level randomization induces a probability distribution across the space of exposure conditions for each vertex. For non-clustered assignment, the probability of having many friends co-treated is **very** low. Clustering aims to increase this.

Below are sketches of probability distributions over the {0,1}x[0,1] space of exposures. For fractional q-neighborhood exposure, exposure condition is captured by yellow box:



Small clusters for small variance

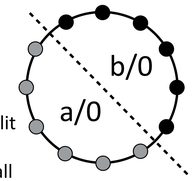
The main goal of clustering treatment is to reduce the variance of the ATE estimator. Not all clustering strategies are equal. As part of our analysis of the Horvitz-Thompson ATE estimator variance, we identify the following **sufficient conditions** for asymptotic smallness of the variance (proof in paper):

Proposition 2 *Assume the potential outcomes are all $O(1)$ in n. If G is sparse with degrees bounded by $O(1)$ and the size of each cluster is $O(1)$, then the variance of the Horvitz-Thompson ATE for full, k-neighborhood, and q-fractional neighborhood exposure under clustered graph randomization is $O(1/n)$.*

This proposition says that we require the clustering algorithm to produce small clusters, and having any O(n) sized clusters may undermine the variance.

Homophily example with non-trivial optimal clustering

Let G be a cycle with 2n vertices. Consider the full neighborhood exposure model. We are interested in the average treatment effect between being (A) treated and both neighbors treated, and (B) not treated and neither neighbor treated.



Consider a polarized fixed response model where the vertices are split by a balanced division of the cycle: one half responds Y = a to the network exposure condition, the other responds Y = b. Meanwhile, all respond Y = 0 to the network control exposure condition. Importantly, the orientation of the partition is unknown to the experiment designer.

The Horvitz-Thompson ATE variance [2] for this example reduces to:

$$Var[\hat{\tau}(Z)] = \frac{a^2+b^2}{4n^2}\left[\sum_{i=1}^{n}\frac{1-\pi_i}{\pi_i} + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j}\right] + \frac{ab}{4n^2}\sum_{i=1}^{n}\sum_{j=i}^{n}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j}.$$

Let c be the number of continuous nodes to cluster together. The variance computes to:

$$\begin{aligned} c = 1: &\quad Var[\hat{\tau}(Z)] = (15/4)(a^2+b^2)\frac{1}{n} + O(1/n^2) \\ c = 2: &\quad Var[\hat{\tau}(Z)] = (10/4)(a^2+b^2)\frac{1}{n} + O(1/n^2) \\ c \geq 3: &\quad Var[\hat{\tau}(Z)] = \left(\frac{c}{4} + \frac{2}{c} + 1\right)(a^2+b^2)\frac{1}{n} + O(1/n^2), \quad c = O(1) \text{ in } n \end{aligned}$$

Thus, the variance is minimized by clusters of size **c = 3, an internal optima**.

References

[1] P. Aronow and C. Samii. Estimating average causal effects under general interference. *Working Paper*, September 2012.
[2] D.G. Horvitz and D.J. Thompson. A generalization of sampling without replacement from a finite universe. *J Amer Statist Assoc*, 47(260):663–685, 1952.

Paper: http://bit.ly/nips12-ukbk Poster: http://bit.ly/nips12-ukbk2