Social Networks 54 (2018) 196–208

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Models, methods and network topology: Experimental design for the study of interference

Jake Bowers[a,*], Bruce A. Desmarais[b], Mark Frederickson[a], Nahomi Ichino[c], Hsuan-Wei Lee[d], Simi Wang[e,1]

[a] Departments of Political Science and Statistics, University of Illinois at Urbana-Champaign, United States
[b] Department of Political Science, Pennsylvania State University, United States
[c] Department of Political Science, University of Michigan, United States
[d] Department of Sociology, University of Nebraska-Lincoln, United States
[e] Department of Applied Mathematics, University of North Carolina at Chapel Hill, United States

ARTICLE INFO

Article history:
Available online 14 March 2018

Keywords:
Experimental design
Spillover
Interference
Networks
Causal inference
Statistical power

ABSTRACT

How should a network experiment be designed to achieve high statistical power? Experimental treatments on networks may spread. Randomizing assignment of treatment to nodes enhances learning about the counterfactual causal effects of a social network experiment and also requires new methodology (ex. Aronow and Samii, 2017a; Bowers et al., 2013; Toulis and Kao, 2013). In this paper we show that the way in which a treatment propagates across a social network affects the statistical power of an experimental design. As such, prior information regarding treatment propagation should be incorporated into the experimental design. Our findings justify reconsideration of standard practice in circumstances where units are presumed to be independent even in simple experiments: information about treatment effects is *not* maximized when we assign half the units to treatment and half to control. We also present an example in which statistical power depends on the extent to which the network degree of nodes is correlated with treatment assignment probability. We recommend that researchers think carefully about the underlying treatment propagation model motivating their study in designing an experiment on a network.

Published by Elsevier B.V.

1. Introduction

We consider the problem of designing experiments to causally identify propagation on networks. In a simple experiment on *independent* units with complete randomization to two treatment arms, it is often assumed that one should assign half of the experimental pool to treatment and half to control (Gerber and Green, 2012).[2] When treatment given to one unit may affect another unit, however, we show (in a simulation study using a realistic network and realistic model of network treatment propagation) that it may be better to assign *less* than half of the pool to treatment from the perspective of statistical efficiency. The intuition is simple: if treatment spreads rapidly across a network, then comparisons of outcomes between treated and control units will become very small or even vanish as the control units to which the treatment spread will act just like treated units. Thus, one might field a very effective experiment, perhaps an experiment in which controls race to get access to the treatment or treated units spread the information or other active ingredient far and wide, but be unable to detect effects if everyone in the whole network reveals the same outcome whether or not they were assigned to treatment. The simulations that we show here confirm this intuition, but also reveal a trade-off between ability to detect the direct effects of treatment assignment on the units initially assigned to treatment and the ability to detect the indirect or network mediated effects of the treatment as it propagates to control units. One point that we emphasize in this paper is that the way in which a treatment propagates matters a great deal as we think about how to design experiments on networks.

In fields across the social and physical sciences, there is considerable and growing interest in understanding how features propagate over the vertices (i.e., nodes) in a graph (i.e., network) via the graph topology. Furthermore, precise questions about causal peer, spillover and propagation effects are becoming more common. Recent theoretical developments highlight the barriers to the identification of causal peer/contagion effects in networks with non-randomized, or observational, data (Lyons, 2011; Shalizi and

* Corresponding author.
E-mail addresses: jwbowers@illinois.edu (J. Bowers), bdesmarais@psu.edu (B.A. Desmarais).
[1] Authors listed alphabetically.
[2] Technically speaking, the 50/50 treatment allocation is optimal for precision when randomization is complete at the unit-level and outcomes have equal variance in both treated and control groups.

https://doi.org/10.1016/j.socnet.2018.01.010
0378-8733/Published by Elsevier B.V.

Thomas, 2011). Several recent papers have employed randomized experimental designs to facilitate the identification of causal peer effects (Aral and Walker, 2011; Ostrovsky and Schwarz, 2011; Bapna and Umyarov, 2015; Bond et al., 2012; Ichino and Schündeln, 2012; Nickerson, 2008). For example, Ichino and Schündeln (2012) conduct a field experiment during a national election in Ghana to gauge how voter registration responds to the placement of election monitors at registration workstations — an effect that is hypothesized to spread geographically through the road network.

Recent methodological work enables scholars to make statistical inferences about peer effects or global average effects when the topology of a network is known (Bowers et al., 2013; Aronow and Samii, 2017a; Eckles et al., 2017; Toulis and Kao, 2013).[3] As the ability to pose questions of spillover has increased, researchers have begun to address how well these methods work, particularly with respect to statistical efficiency. Eckles et al. (2017) show that a graph cluster randomization design — where groups of nodes are randomized to treatment together — reduces bias in estimates of global average treatment effects with relatively little cost in terms of statistical power. Baird et al. (2017) derive the efficiency calculations for estimates of average spillover effects for randomization designs in which isolated groups of nodes are randomized first to a saturation proportion — the proportion of units within the group to be randomized to treatment — and then within group randomization proceeds according to the first level randomization. Hirano and Hahn (2010) derive efficiency calculations regarding cluster-wise and within-cluster treatment proportions for estimates of direct and indirect effects in two-level cluster randomization designs. These approaches answer important questions about particular designs; however, there is still a need to address how to *design randomization schemes* to increase the statistical power to detect specific forms of network mediated peer effects.

In this project we consider the performance of different randomization designs using the methods of Bowers et al. (2013) and Aronow and Samii (2017a) under different models of propagation. Each of the methods we consider depends upon a typology of exposure conditions based on the treatment status of each node and the topology of the graph. For example, a node could be treated directly by an experimenter, isolated from treatment (i.e., several hops away from any treated nodes) or exposed to the treatment at one degree of separation by virtue of the network relationship — without control by the experimenter. The performance of randomized experimental designs on networks depends on (1) the exposure conditions of theoretical interest (say, direct treatment versus indirect treatment; or more generally some propagation flow parameter), (2) the topology of the network, (3) the ways in which the propagation model affects nodes in each exposure condition, and (4) the exposure condition distribution as determined by the randomization design.[4]

To anchor our interest in interference, consider Coppock's (2014) recent replication of Butler et al. (2011). Butler et al. (2011) run a field experiment that is focused on a special session of the New Mexico legislature that was called to consider a specific budgetary question. The field experiment was designed to test the influence of providing information about constituents' preferences on legislators' votes. Constituents across the state were first surveyed on the budget question on which their legislators would be voting. Butler and Nickerson sent district-specific results to randomly selected members of the legislature. They found that providing information about constituents' preferences shifted legislators' votes in the direction of those preferences. Coppock (2014, pp. 159–160) notes that,

> "The estimates of responsiveness recovered by Butler et al. (2011) rely on an assumption of non-interference (Cox 1958; Rubin, 1980): Legislators respond only to their own treatment status and not to the treatment status of others. This assumption requires that legislators not share treatment information with one another, which is at odds with the observation by Kingdon (1973, p. 6) that legislatures are information-sharing networks."

In replicating Butler et al. (2011), Coppock (2014) specifies a model for the propagation of effects that spread through a network between legislators defined by ideological similarity. Accounting for the fact that the treatment assigned to one legislator had effects on other legislators, Coppock (2014) estimates that the experiment shifted nearly twice as many votes in the legislature as was originally estimated by Butler et al. (2011).[5]

In what follows, we study the problem of causal inference given treatment propagation in the context of a fixed graph topology and a single round of randomized treatment and by a single round of response measurement. We review methods that have been proposed in the literature for analyzing single-round (pre versus post), fixed graph experimental data; and also review the substantive experimental applications that have used such designs. We then conduct a simulation study motivated by the registration monitor randomization in Ichino and Schündeln (2012), using the Ghanaian network of roads between voter registration stations as a realistic moderate sized graph.[6] In the simulation study, we consider the performance of alternative experimental designs that vary the treatment probability: the number of nodes assigned to initial treatment, who is treated: the association between treatment probability and node degree (i.e., a node's number of ties), and how they are treated: different parameterizations of the propagation model.

1.1. Statistical inference for propagated causal effects

We consider two general approaches to statistical inference about causal effects when those effects may propagate through a network. The flexible approach developed by Bowers et al. (2013) is a hypothesis testing framework designed to evaluate whether differences between the treatment and control groups are more effectively characterized by one model of treatment effects, which can include propagation effects, than another model. Bowers et al. (2013) focus on a natural sharp null model of no treatment effects (i.e., stochastic equivalence across all experimental conditions). The null distribution is derived exactly or generated approximately through repeated computations of the test statistic using permutations in which the treatment vector is re-randomized according to the experimental design, and the hypothesized effects of the propagation model are removed. There are two highly appealing

[3] For now, we set to the side the work on identifying how much of a total average effect can be attributed to mechanisms other than direct treatment assignment — for example, the work on spillovers and indirect effects (Sinclair et al., 2012; Sinclair, 2011; Nickerson, 2008, 2011; Hudgens and Halloran, 2008; Sobel, 2006; Tchetgen and VanderWeele, 2010; VanderWeele, 2008, 2010, 2011; VanderWeele et al., 2011, 2012; VanderWeele and Tchetgen, 2011; Miguel and Kremer, 2004; Chen et al., 2010; Ichino and Schündeln, 2012).

[4] We direct readers to Basse and Airoldi (2015) for a methodological investigation similar to ours. They consider the problem of designing a randomized experiment to minimize estimation error when outcomes are correlated on a network. Their focus is, however, on estimating the direct effects of treatment, not on identifying indirect or propagation effects.

[5] Coppock (2016) later shows that the test statistic and research design was underpowered to detect this effect.

[6] Ichino and Schündeln (2012) did not use the road network in their paper, but instead focused on estimating average spillover effects within radii of 5 km and 10 km following the multi-level experimental design of Sinclair et al. (2012). We use the road network to provide us with a realistic network for use in our simulations studying the power of different randomization allocation plans.