

Models, Methods and Network Topology: Experimental Design for the Study of Interference*

Jake Bowers [†] Bruce A. Desmarais [‡] Mark Frederickson [§] Nahomi Ichino [¶]
Hsuan-Wei Lee ^{||} Simi Wang ^{**†}

January 7, 2016

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 (Aronow and Samii, 2013; Bowers et al., 2013; Toulis and Kao, 2013, ex.). 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 run against standard advice in circumstances where units are presumed to be independent: information about treatment effects is *not* maximized when we assign half the units to treatment and half to control. We also show that 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.

*Authors ordered alphabetically.

[†]Associate Professor, Departments of Political Science & Statistics, University of Illinois at Urbana-Champaign, jwbowers@illinois.edu.

[‡]Associate Professor, Department of Political Science, Pennsylvania State University, bdesmarais@psu.edu.

[§]PhD Student, Departments of Political Science and Statistics, University of Illinois at Urbana-Champaign,

[¶]Assistant Professor, Department of Political Science, University of Michigan

^{||}PhD Student, Department of Applied Mathematics, University of North Carolina at Chapel Hill

^{**}PhD Student, Department of Applied Mathematics, University of North Carolina at Chapel Hill

^{††}Authors listed alphabetically.

1 Introduction

We consider the problem of designing experiments to causally identify propagation 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, observational, data (Lyons, 2011; Shalizi and 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 Schndeln, 2012; Nickerson, 2008). Aral and Walker (2011) study peer effects in Facebook application adoption through the randomization of ‘viral’ features in a Facebook application. Bond et al. (2012) conduct an experiment to study social influence in voting behavior and other forms of political activity on Facebook during the 2010 congressional election. Ichino and Schndeln (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. Bapna and Umyarov (2015) run a field experiment to determine whether peers influence each others’ adoption of premium online social network services for which they have to pay.

Recent methodological work enables scholars make statistical inferences about peer effects when the topology of a network is known (Bowers et al., 2013; Aronow and Samii, 2013; Toulis and Kao, 2013). However, there is little in the way of informative methodological research regarding the **design** of randomization schemes that account for the form(s) of peer effect(s) in which the researcher is interested.¹ In this project we consider the performance of different randomization designs using the methods of Bowers et al. (2013), Aronow and Samii (2013) and Toulis and Kao (2013) 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 without control by the experimenter. The

¹For 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 McConnell et al. (2010); Sinclair (2011); Nickerson (2008, 2011); Hudgens and Halloran (2008); Sobel (2006); Tchetgen and VanderWeele (2010); VanderWeele (2008, 2010); VanderWeele et al. (2011, 2012); VanderWeele and Tchetgen (2011); VanderWeele and Tchetgen Tchetgen (2011); Miguel and Kremer (2004); Chen et al. (2010); Ichino and Schündeln (2012)

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.²

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 followed by a single round of response measurement. We review methods that have been proposed in the literature for designing and 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 Schndeln \(2012\)](#), using the Ghanaian network of roads between voter registration stations as a realistic moderate sized graph. In the simulation study, we consider the performance of alternative experimental designs that varying 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 Recent Methodological Innovations in 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 properties of this approach. First, any test statistic, including general distributional comparisons such as the Kolmogorov-Smirnov (KS) test statistic, can be used to evaluate the differences between treatment and control. Second, the

²We direct readers to [Basse and Airolidi \(2015\)](#) 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.

approach can accommodate any model of treatment effects on a network, as the methodology does not require any form of independence assumption or the derivation of an estimator for the model parameters.

The methods developed by [Aronow and Samii \(2013\)](#) and [Toulis and Kao \(2013\)](#) compliment those proposed by [Bowers et al. \(2013\)](#) in that they propose methods for *estimating* average causal peer effects. [Aronow and Samii \(2013\)](#) develops randomization-based methods and [Toulis and Kao \(2013\)](#) develops both randomization and model-based approaches to estimating causal peer effects. In both [Aronow and Samii \(2013\)](#) and [Toulis and Kao \(2013\)](#), the target estimate is the average difference between nodes in different network/treatment exposure conditions. [Aronow and Samii \(2013\)](#) do not stipulate a constrained set of conditions, but present methods that can be applied to any partition of nodes into network/treatment exposure conditions. They present an example in which nodes in a graph are directly treated and assume that treatment can only propagate one degree, which results in four conditions: *control*, which are nodes that are not directly assigned treatment and are not tied to any treated nodes; *direct*, which are nodes that are treated and not tied to any treated nodes, *direct + indirect*, which are nodes that are directly treated and are tied to treated nodes; and *isolated direct*, which are nodes that are untreated and are tied to treated nodes. [Toulis and Kao \(2013\)](#) define k -level treatment of a unit as (1) a unit not receiving direct treatment, and (2) having exactly k directly treated neighbors. A k -level control is any vertex with at least k neighbors who did not (1) receive direct treatment and (2) is not connected to any vertices who were directly treated. These approaches assume that the researcher is interested in specific comparisons of averages and require that the researcher articulate mechanisms by which the probability of exposure to treatment may differ across units.

Both [Aronow and Samii \(2013\)](#) and [Toulis and Kao \(2013\)](#) recognize the unique challenges that arise in the context of inference regarding response to network/treatment exposure conditions. The limitations are based in the topology of the graph. Since most exposure conditions of interest in the context of interference involve the position of a node in a network, under most randomization designs (e.g., uniform random assignment to treatment), each node is not equally likely to be assigned to each exposure condition. Take the example of 2-level exposure in the framework of [Toulis and Kao \(2013\)](#). A node with only one partner would have zero probability of being assigned to the 2-level treatment group. [Aronow and Samii \(2013\)](#) do not discuss this issue at length, but imply a limitation in the derivation of their Horvitz-Thompson type estimators ([Horvitz and Thompson, 1952](#)). The estimators they define require that the analyst be able to calculate the probability $\pi_i(d_k)$, the probability that node i is in exposure condition d_k and that $0 < \pi_i(d_k) < 1$ for each node i . This means that the framework proposed by [Aronow and Samii \(2013\)](#) cannot be applied to the comparison of exposure conditions to which all nodes cannot be

assigned. [Toulis and Kao \(2013\)](#) are more explicit in their discussion of this limitation. They define a causally valid randomization design to be one in which at least one node is k -level treated and one is k -level controlled.

In the analysis that follows, we consider the [Aronow and Samii \(2013\)](#) and [Bowers et al. \(2013\)](#) approaches to inference with experiments on networks. The methods proposed by [Toulis and Kao \(2013\)](#) are very similar to those of [Aronow and Samii \(2013\)](#), and their concepts of k -level treated and k -level controlled can be seen as special cases of the exposure conditions defined in [Aronow and Samii \(2013\)](#). Furthermore, the methods of [Aronow and Samii \(2013\)](#) have the added advantage of adjusting the treatment effect estimates (and variance estimates) for the unequal exposure condition probabilities. These Horvitz-Thompson type adjustments will correct for any associations between exposure condition probabilities and potential outcomes (e.g., higher degree nodes may exhibit higher baseline response values *and* be more likely to be indirectly exposed to treatment through propagation).

2 Design as a Function of Graph Topology

[Walker and Muchnik \(2014\)](#) reviews several applications of experiments in networked environments – including studies that are not focused on propagation – and outline several of the fundamental challenges to designing experiments on networks. They summarize the problem of design in experiments on networks succinctly (p. 1949):

“The natural connectivity of our world does not only present a challenge to the conventional paradigm of experimental design, but also reveals opportunities to leverage connectivity through the creation of novel treatment mechanisms that incorporate both experimental subjects and the connections between them.”

The practical implication of the dependence between subjects via the network is that efficient experimental designs will account for graph topology. The treatment assignment algorithms presented in [Toulis and Kao \(2013\)](#) render a clear picture of how the importance of network structure can complicate design. Considering the problem of assuring that sufficient numbers of vertices end up in the k -treated and k -controlled designs, [Toulis and Kao \(2013\)](#) present sequential randomization designs that assure that fixed numbers of vertices are assigned to the groups under comparison: for example, if a researcher desires to know the effect of the treatment on nodes having 2 directly connected neighbors in the network a design should ensure enough nodes with degree 2 assigned treatment versus not assigned treatment. Though powerful in their ability to control the distribution of vertices across exposure conditions, the complex

sequential randomization algorithms proposed by [Toulis and Kao \(2013\)](#) make closed-form calculation of the probability of exposure condition assignment intractable in most cases, which may be why they do not derive their estimators using Horvitz-Thompson adjustments such as those in [Aronow and Samii \(2013\)](#). An example of a non-sequential randomization design for which it is straightforward to derive the Horvitz-Thompson weights is one in which the probability of treatment is biased with respect to vertex degree (e.g., disproportionately treating higher degree vertices). To provide an intuitive example regarding why it might be advantageous to treat high degree vertices at a greater rate than low degree vertices; suppose the researcher is interested in comparing nodes isolated from treatment (e.g., more than two degrees from any directly treated node) to nodes that are adjacent to a treated node, but are not directly treated. Each node that is directly treated is removed from both exposure conditions of interest, so there is an incentive to treat a small proportion of nodes. However, if too few nodes are treated, there will be too few nodes in the adjacent-to-treated condition. By focusing the treatment on higher degree nodes, it takes fewer directly treated nodes to accomplish a sizable sample of adjacent-to-treated nodes. Depending upon the structure of the network and the mechanism by which treatment can propagate, there may be a considerable gain in statistical power from biasing treatment towards high degree nodes, as compared to uniform assignment to treatment.

2.1 Design is Dependent on a Model of Propagation

In the classical experimental framework, a counterfactual causal effect for a person i is defined as the situation where that person's outcome, $Y_i(Z_i = 1)$ under one treatment would differ from that same person's outcome under a different treatment $Y_i(Z_i = 0)$ (given at the same moment in time): $Y_i(Z_i) \neq Y_i(Z'_i), Z_i \neq Z'_i$. That is, a causal effect is understood in the subjunctive tense. And we can write this comparison for a single person because we presume that treatment assigned to person i will have no effect on the outcomes presented by any other person in the experiment — that there is no interference between the treatment or outcomes of person i and those of any other person. The validity of the classical experimental framework, in which a lack of interference is assumed or assured through design, and the researcher is interested in differences between control and directly treated units, addresses mechanisms for effects only indirectly. A given mechanism or theory tends to motivate a study with implications for a given comparison and clever design can rule out competing theoretical explanations for the same comparison (positive, negative, large, small, absent, etc.). The links between theoretical models and the analysis and design of experiments on social networks cannot be so loose especially if a scholar desires to learn about propagation processes. In the classical experimental setup, the groups to be compared are obvious and

determined by the direct experimental conditions. However, once the assumption of no interference is relaxed, and especially if there is interest in evaluating the effects of interference, an implicit or explicit theoretical model of interference must be drawn upon to guide the search for propagation's footprint.

Both the methods of [Aronow and Samii \(2013\)](#) and [Bowers et al. \(2013\)](#) require that the researcher draw upon a theoretical model of interference. In [Aronow and Samii \(2013\)](#), a model of interference must be used to identify the exposure conditions (e.g., adjacent to a treated unit, 2 degrees of separation from a treated unit, etc.) to be compared in the study. The model used with the methods of [Aronow and Samii \(2013\)](#) need not provide precise predictions about to which nodes the treatment will propagate and how, but the model must be complete enough to identify the groups of nodes for which different potential outcomes will be observed under a given direct treatment assignment regime. For the methods proposed by [Bowers et al. \(2013\)](#), there is great value in a precise analytic model of interference. Their approach compares the observed data in which the effect of a treatment, through a parameterized interference model, has been removed, to data generated by permuting the initial treatment and removing the effects as parameterized through the interference model. As such, it is not possible to use the methods of [Bowers et al. \(2013\)](#) without specifying a model of interference.

The need to specify a model of propagation in order to identify an efficient experimental design leads us to question where researchers might start in developing such models. There is a vibrant literature, primarily in the fields of physics and applied mathematics, on graph dynamics, that provides several excellent starting points for analytical models of propagation. These models include the susceptible-infected-recovered disease epidemic models ([Kermack, 1927](#); [Anderson, 1982](#); [Hethcote, 2000](#); [Daley, 2005](#)), the Bass Diffusion Model ([Bass and Jain, 1994](#); [Lenk and Rao, 1990](#)) and the Hopfield network [J. \(1982\)](#) and the voter model [Clifford \(1973\)](#); [Liggett \(1997\)](#); [Durrett \(1991\)](#). In the simulation study that follows, we present and use a variant of the Ising model, a model that contains several of those mentioned above as special cases ([Gallavotti, 1999](#)). The Ising model is a general formulation for stochastic binary-state dynamics. Many models of opinion dynamics are based on the classical Ising model of magnetic spins. Here, each node is in either the spin-up or spin-down state, and transitions occur according to dynamical rules that minimize the Hamiltonian of ferromagnetic interactions.

3 Simulation Study

In what follows we conduct a simulation study in which we evaluate the statistical power of the methods proposed by [Aronow and Samii \(2013\)](#) and [Bowers et al. \(2013\)](#). The objective of this simulation study

is to demonstrate that the statistical power of these procedures depends upon design parameters that are intuitively meaningful in the context of propagation and that power is maximized at design parameters that differ considerably from those commonly used in experiments without interference – random uniform division of the sample into half control and half treatment. It is outside the scope of the current study to compare the merits of the methods proposed by [Aronow and Samii \(2013\)](#) with those proposed by [Bowers et al. \(2013\)](#): further, we see the two approaches as complementary in the same way that hypothesis testing and estimation complement one another. We hope that our simulation will illustrate the importance and feasibility of simulation analysis for parameterizing designs for studies of propagation in networks. Throughout this simulation we use the Ghana voter registration station network from [Ichino and Schndeln \(2012\)](#) as our example network. This network is derived from the roads connecting the registration stations. Two registration stations are considered to be tied if they are within 20km of each other on the road network. This results in a network with 868 vertices, and a density of 2.2%. The network is depicted in [Figure 3](#).

3.1 Simulation Parameters and Definitions

We denote the treatment assigned to vertex i as $Z_i \in \{0, 1\}$ and the vector of treatment assignments to all vertices as $\mathbf{Z} = \{Z_1, \dots, Z_n\}$. The fixed potential outcome of vertex i that would be shown under treatment vector \mathbf{Z} is denoted $Y_i(\mathbf{Z})$. Under the sharp null of no treatment effect, all potential outcomes are equal; $Y_i(\mathbf{Z}) = Y_i(\mathbf{Z}') = Y_i(Z_i = 1, \mathbf{Z}_{-i}) = Y_i(Z_i = 0, \mathbf{Z}'_{-i}) \quad \forall \mathbf{Z} \neq \mathbf{Z}'$. $Y_{it}(D_{it})$ is the potential outcome of vertex i at time t under exposure condition D_{it} . The exposure condition indicates a mapping of the graph topology and \mathbf{Z} into the categories of exposure defined by the researcher (e.g., directly treated and not tied to any directly treated vertices, not directly treated and adjacent to at least one directly treated node).

The model we use for treatment propagation in our simulations is a variant of the Ising Model. The initial treatment assignment of each vertex is drawn independently from $Z_i \sim \text{Bernoulli}(\alpha)$. The infection probability at each iteration of propagation is

$$\frac{1}{1 + \exp\left(\frac{2}{F}(k_i - 2m_i)\right)}$$

, where k is number of (directly adjacent) neighbors, m is the number of previously exposed neighbors ($0 \leq m_i \leq M$), F is a “temperature” parameter that governs the extent to which the propensity to be infected depends on the infection rate among i ’s neighbors. We run the model for just two two time

periods ($t \in \{0, 1\}$). The Ising model controls actual infection after an experimenter assigns Z_i at $t = 0$.

We specify the potential outcomes of the vertices in our simulation according to the following scheme depending upon the infection status at a given time point. We denote this by $Y(Z_{i0}, Z_{i1})$, in which Z_{i0} indicates i 's initial treatment status and Z_{i1} indicates whether treatment propagates to i at time 1. We generate a baseline (pre-treatment) response $Y(0, 0) \sim U(0, 1)$. Our simple treatment effect model changes the baseline in the same, multiplicative, way regardless of the time or manner of ‘infection’ (directly assigned by researcher or propagated from a neighbor), and $Y(1, 0) = Y(0, 1) = \lambda Y(0, 0)$. We consider values of $\lambda \in \{0.26, 0.63\}$, which correspond to approximately one and two standard deviation shifts in the mean of a standard uniform random variable, and simulate 1,000 treatment propagation and outcome sets at each combination of $F \in \{0, 10, \dots, 100\}$ and $\alpha \in \{0.05, 0.10, \dots, 0.50\}$.

3.2 Application of Inference Methods

[Aronow and Samii \(2013\)](#) method requires that we define exposure conditions (which include assignment to treatment and also the probability of exposure to a treatment via the network). We define the exposure conditions with respect to what a researcher would be able to observe from the experimental design, assuming the network were observed, and the response. Importantly, in defining the exposure conditions of interest, we assume that the researcher does not exactly know the set of vertices to which the treatment has propagated. We see this as more realistic than a situation in which the researcher knows exactly where the treatment has propagated.

We define the following three distinct exposure conditions, differentiating between those vertices treated initially (d_1), those vertices that are untreated initially and are adjacent to at least one treated vertex ($d_{(0,1)}$), and those vertices that are untreated initially and are not adjacent to any treated vertices ($d_{(0,0)}$).

- $d_1 \equiv D_i(Z_i = 1, 0 \leq m_{it} \leq M)$
- $d_{(0,0)} \equiv D_i(Z_i = 0, m_{it} = 0)$
- $d_{(0,1)} \equiv D_i(Z_i = 0, m_{it} \geq 1)$

Figure 1 gives a visual example of a subgraph drawn from the Ghana road network. In the study of the [Aronow and Samii \(2013\)](#) methods, we focus on identifying the difference between the $d_{(0,1)}$ and $d_{(0,0)}$ conditions.

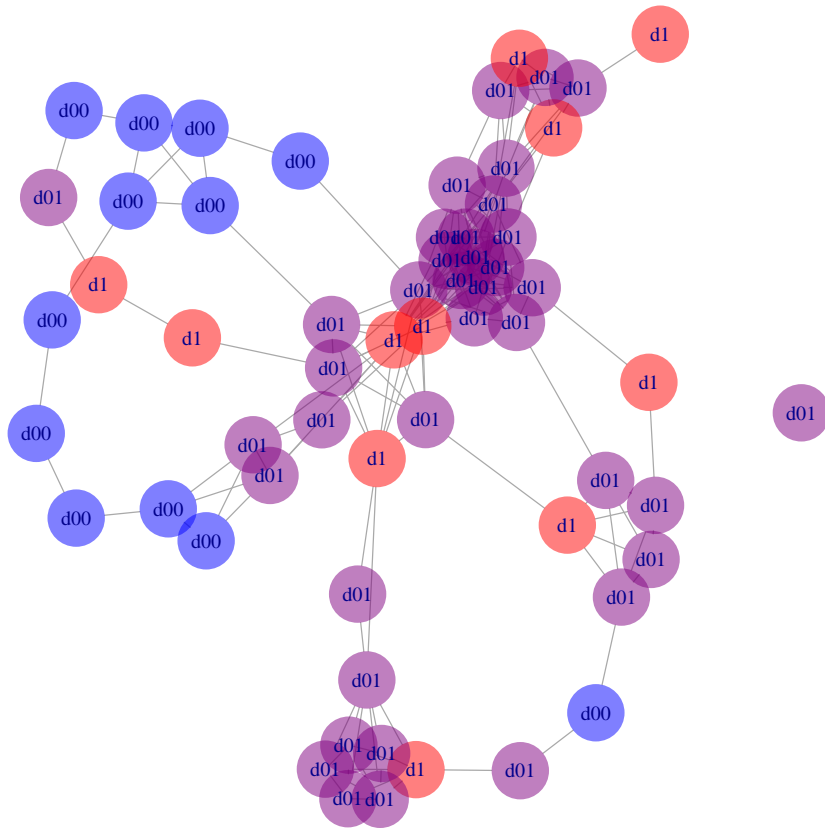


Figure 1: Illustration of exposure conditions. One draw from the Ising propagations with $\text{pr}(Z_i = 1) \sim \text{Bernoulli}(.15)$ for $t \in \{0, 1\}$ and Temperature= 10.

The [Aronow and Samii \(2013\)](#) estimand is the difference between vertices in two exposure conditions. First, let

$$\hat{\mu}(d_k) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(D_i = d_k) \frac{Y_i(d_k)}{\pi_i(d_k)},$$

be the estimator of the mean potential outcome among vertices in exposure condition d_k , where $\pi_i(d_k)$ is the probability that vertex i ends up in condition d_k . Then the estimator of the difference between potential outcomes in the two exposure conditions is

$$\hat{\tau}(d_{(0,1)}, d_{(0,0)}) = \hat{\mu}(d_{(0,1)}) - \hat{\mu}(d_{(0,0)}).$$

Following [Horvitz and Thompson \(1952\)](#), [Aronow and Samii \(2013\)](#) show that $\text{Var}(\hat{\tau}(d_{(0,1)}, d_{(0,0)})) \leq 1/N^2 (\text{Var}(\hat{\mu}(d_{(0,1)})) + \text{Var}(\hat{\mu}(d_{(0,0)})))$. In the interest of brevity, we do not reproduce the full variance estimators, but refer readers to [Aronow and Samii \(2013\)](#). A power analysis gauges the probability that a true hypothesis would be rejected against different alternatives. In the results that follow, we set the truth as $\tau = 0$ and thus assess power to reject $H_0 : \tau = 0$ at the 0.05 significance level.

We test hypotheses about effects following [Bowers et al. \(2013\)](#) using the Anderson-Darling k -sample test statistic [Scholz and Stephens \(1987\)](#) to compare the outcome distributions of vertices in the three different exposure conditions.³ Following the norms of assessing power against truth, we test hypotheses generated by the correct Ising model record the amount of false rejections as the parameter values of λ and F move away from their true values. The values of the parameters λ and F are manipulated to test the null of no effects (i.e., $\lambda = 1, F = \infty$) in order to assess power. In the case of the methods proposed by [Bowers et al. \(2013\)](#), we assess both the power against the null of no effects across all three exposure conditions and the null of no difference between $d_{(0,1)}$ and $d_{(0,0)}$.

3.3 Results

The results from the [Aronow and Samii \(2013\)](#) tests are reported in [Figure 2](#). Three distinct sets of results are presented: (a) those with a multiplicative effect and stochastic propagation (i.e., $Y(1,0) = Y(0,1) = \lambda Y(0,0)$), (b) those with an additive effect (i.e., $Y(1,0) = Y(0,1) = \lambda + Y(0,0)$), and (c) those with an additive effect and certain propagation (i.e., $Y(0,1) = Y(d_{(0,1)})$). We note three characteristics of our results. First, the tests exhibit fairly low power overall, hitting a maximum of approximately 0.8,

³Randomization distributions are simulated using a development version of the `RITtools` package for **R** ([Bowers et al., 2014](#)).

but sitting below and often far below 0.5. Second, the most powerful design is that in which $\alpha = 0.05$, the smallest proportion assigned to initial treatment. Third, as indicated by the relationship of power with the x -axis of the plots in panels (a) and (b), the larger the sample of vertices in the $d_{(0,1)}$ condition that are actually exposed to treatment in period one, as governed by the temperature parameter, the more powerful the tests.

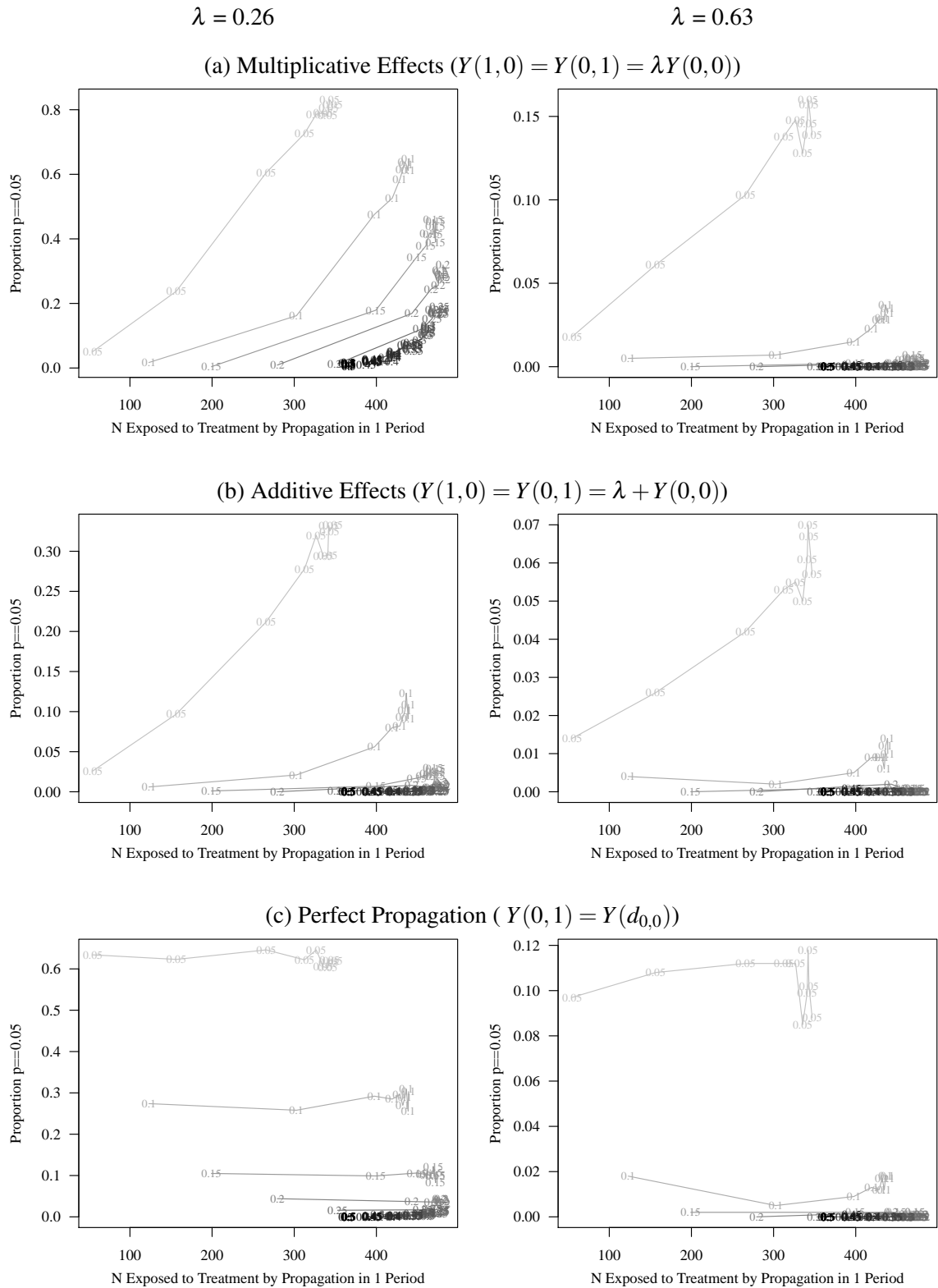
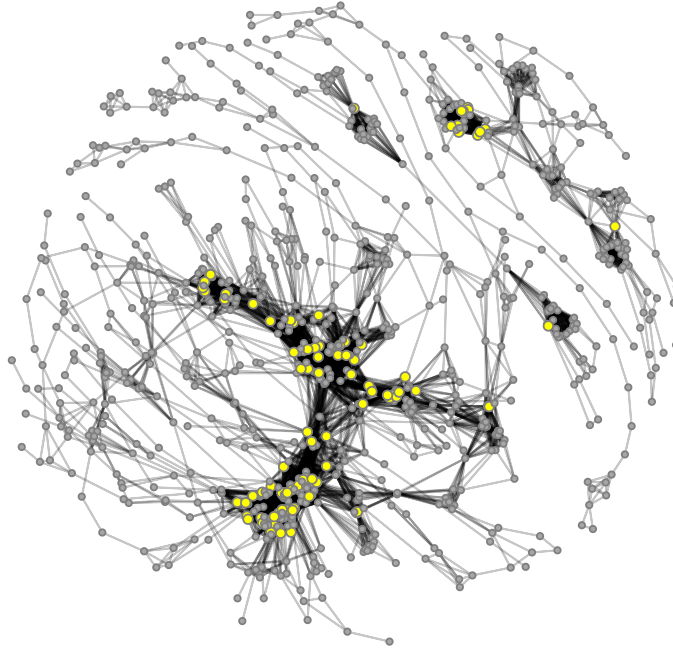


Figure 2: Power Results: [Aronow and Samii \(2013\)](#) test. Lines are labeled by the proportion assigned to treatment. High power depends on low proportions assigned to treatment. The network contains a total of 868 nodes.

If a researcher wants to estimate the network exposure weighted average causal effect developed by [Aronow and Samii \(2013\)](#) these kinds of results raise the question about whether the best approach to randomization of treatment assignment is simple uniform assignment without any blocking or use of

information about the fixed network. To investigate this we plot the correlation between m and Z_{i_0} (i.e., the correlation between vertex degree and the initial treatment assignment). The networks depicted in Figure 3 represent examples of treatment assignments biased in favor of high degree vertices (a) and low degree vertices (b). The degree-treatment correlation results are given in Figure 4. We can see that at each α value there is a strong positive relationship between the degree-treatment correlation and statistical power. This indicates that for the Ghana network structure and the propagation/effects models we have specified, designs that bias treatment assignment towards higher degree nodes would exhibit greater statistical power. This finding generalizes what we know about the use of prognostic background covariates to increase power in randomized experiments to the situation where network degree can be thought of as a moderator of treatment effects.

(a) Treating High Centrality Nodes



(b) Treating Low Centrality Nodes

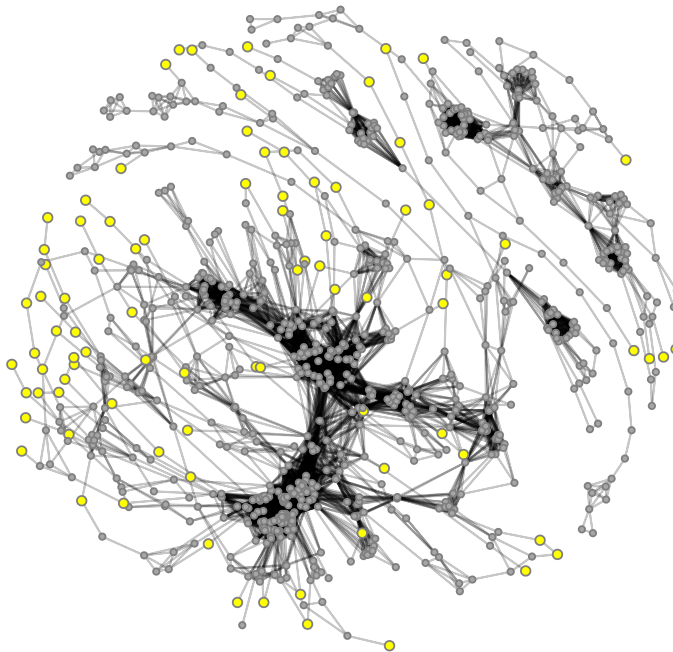


Figure 3: Examples of treatment assignment to either high or low degree nodes in the Ghana 2008 Voter Registration Fraud Experiment network. Vertices are registration stations and yellow nodes are assigned to initial treatment.

The results from the model-testing methods of [Bowers et al. \(2013\)](#) in Figures 5–7 differ in their implications from the results based on estimation. First, the power in these tests is much higher overall. This makes some sense: we are using the correct propagation model and only varying parameters, a good test should eventually reject values of parameters that are distant from the truth. The methods focused

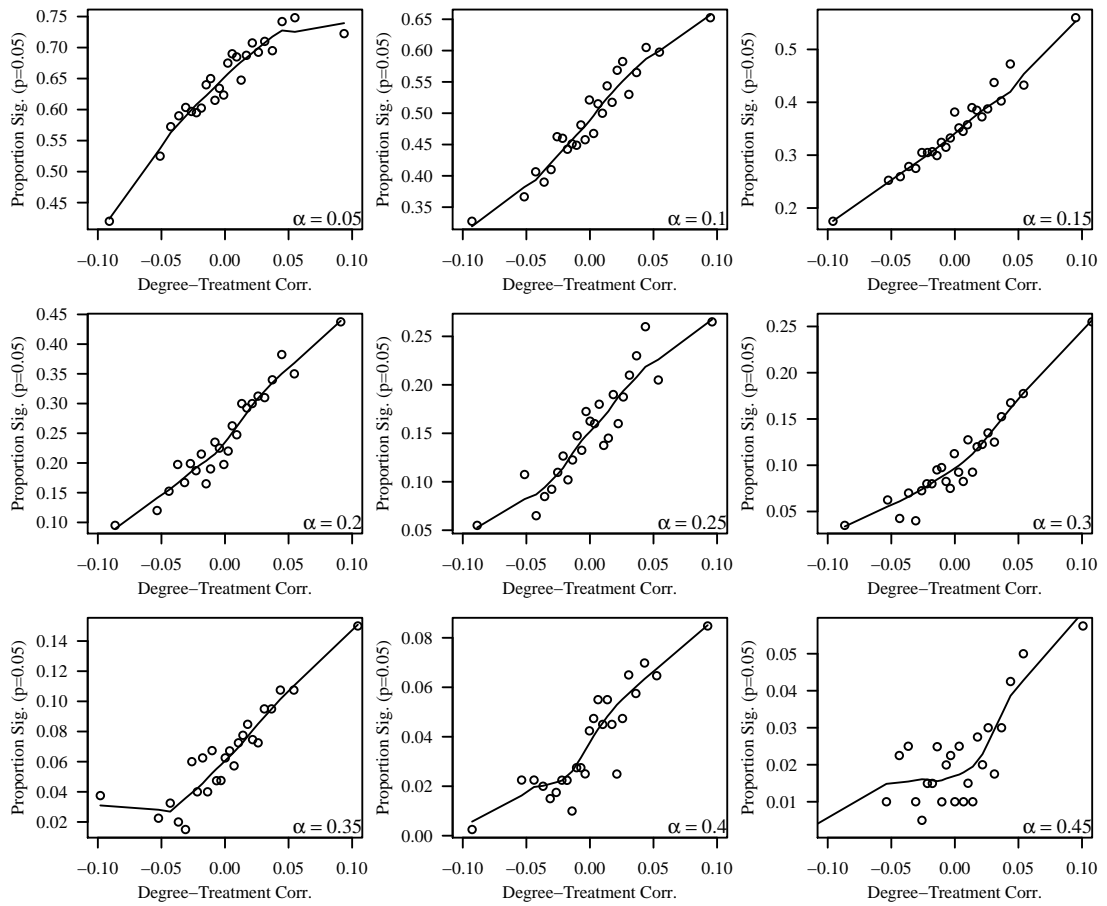


Figure 4: When node degree is correlated with probability of treatment assignment, the power to detect indirect effects increases.

on differences of averages do not include much information about the propagation model except as a weight arising indirectly from what we can observe, so those procedures have less information to use in statistical inference in this simulation study. Second, when we focus on the test of the sharp null of no effects when all three exposure conditions are included, we see that the low values of α exhibit low power. However, when the test compares only $d_{(0,1)}$ against $d_{(0,0)}$, which can be considered a test for propagation effects (i.e., excluding directly treated vertices), the lower α designs perform better. This tradeoff between the power to detect any effect versus the power to detect propagated effects replicate the findings from [Bowers et al. \(2013\)](#): in order to learn about propagation, one should assign relatively few nodes to treatment, in order to test an overall null of no effects, then more power arises from more directly assigned to treatment nodes. There is also a positive association between the number of vertices in the $d_{(0,1)}$ condition that are exposed to treatment by period one and the power of the tests. This last result arises from the fact that, when few in the $d_{(0,1)}$ condition are exposed to treatment, the $d_{(0,1)}$ outcomes are, in large part, equivalent to the $d_{(0,0)}$ outcomes.

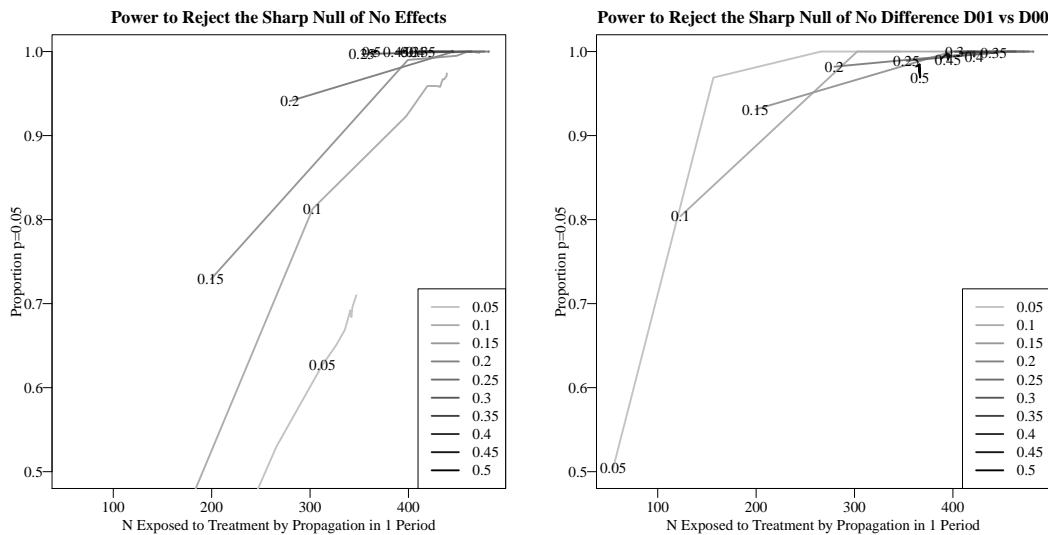


Figure 5: Power with multiplicative effects model and true Ising propagation model following [Bowers et al. \(2013\)](#).

3.4 Summary

We have shown that learning about a simple model of propagation of treatment through a network (via comparisons of nodes possibly indirectly exposed to the treatment to nodes not exposed to treatment) can be enhanced when relatively few of the nodes in a network are assigned treatment. And our analysis acts an example for those desiring to evaluate their own models of propagation and networks and experimental

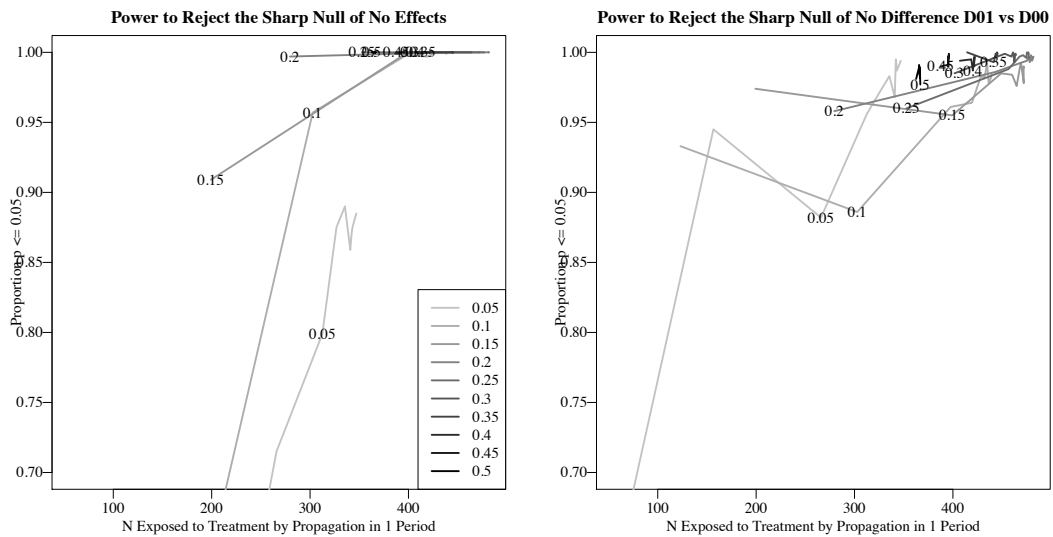


Figure 6: Power with additive effects model and true Ising propagation model following [Bowers et al. \(2013\)](#).

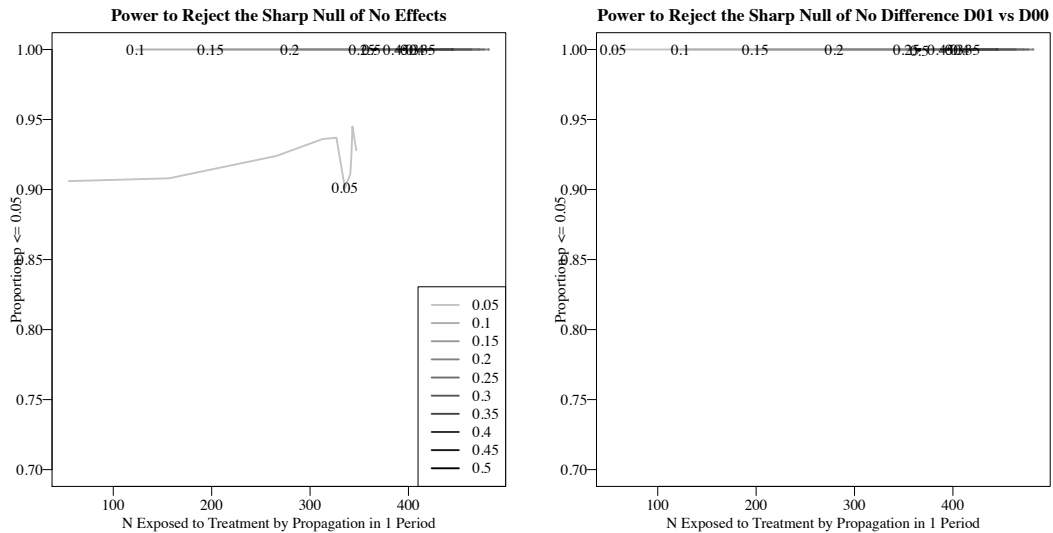


Figure 7: Power with additive effects and certain propagation and true Ising propagation model following [Bowers et al. \(2013\)](#).

designs before going into the field. We have not shown how the [Bowers et al. \(2013\)](#) approach performs when an incorrect model is assessed. We did not do this because we are focusing on design — and power analysis requires that we create a truth against which to compare alternatives. [Bowers et al. \(2013\)](#) show that, in the analysis stage, hypothesis tests may have very low or no power if the model being assessed has no bearing on the underlying mechanism, but we are not certain what kind of design advice would follow from such findings — merely increasing the size of a fixed network may be impossible. We have also not considered how design affects statistical power when there is uncertainty regarding the network

structure. To incorporate uncertainty regarding the network a stochastic model for the network would need to be integrated into the analytical procedure(s) (e.g., [Desmarais and Cranmer, 2012](#)).

4 Conclusion

We describe the challenges in experimental design that arise when the researcher is interested in studying the process of propagation on a network with the objective of drawing causal inference. The experimental designs that work most effectively in experiments in which there is no interference are unlikely to be directly transferable to experimental research on propagation. We review two recently developed frameworks for statistical inference regarding interference in networks. One commonality we draw from these two frameworks is that theoretical analytic models of propagation play a key role in their application, which means that theory features more prominently in the statistical analysis of experimental data generated for the study of propagation than in the classical, non-interference, experimental framework.

We present a simulation study to (1) illustrate how simulation can be a useful guide in identifying design parameters for experimental studies of interference and (2) study the properties of the two frameworks for statistical inference presented in the front end of the paper. Three broad findings from the simulation study are notable. First, statistical power depends upon design parameters and, for example, the optimal proportion assigned to initial treatment may be much lower than the conventional 0.5 that is found in the non-interference experimental literature. Second, the relationship between design parameters and power depends upon the framework for statistical inference. Third, our results regarding the positive relationship between the degree-treatment correlation and power indicates that randomization designs that take into account graph topology are likely to exhibit substantial power gains over uniform randomization designs. Taken together, the results from the simulation study underscore the importance of considering design parameters for experimental studies of propagation, as the standards of the classical experimental framework are unlikely to apply.

References

- Anderson, R. M. e. (1982). *Population Dynamics of Infectious Diseases: Theory and Applications*. Chapman and Hall.
- Aral, S. and Walker, D. (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639.

- Aronow, P. M. and Samii, C. (2013). Estimating average causal effects under interference between units. *arXiv preprint arXiv:1305.6156*.
- Bapna, R. and Umyarov, A. (2015). Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. *Management Science*, 61(8):1902–1920.
- Bass, Frank M.; Krishnan, T. V. and Jain, D. C. (1994). Why the bass model fits without decision variables. *Marketing Science*, 13(3):204–223.
- Basse, G. W. and Airoidi, E. M. (2015). Optimal design of experiments in the presence of network-correlated outcomes. *arXiv preprint arXiv:1507.00803*.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298.
- Bowers, J., Fredrickson, M., and Hansen, B. (2014). *RIttools: Randomization Inference Tools*. R package version 0.1-12.
- Bowers, J., Fredrickson, M. M., and Panagopoulos, C. (2013). Reasoning about interference between units: A general framework. *Political Analysis*, 21(1):97–124.
- Chen, J., Humphreys, M., and Modi, V. (2010). Technology diffusion and social networks: Evidence from a field experiment in uganda.
- Clifford, P. A. W. S. (1973). A model for spatial conflict. *Biometrika*, 60:581–588.
- Daley, D. J. s Gani, J. (2005). *Epidemic Modeling: An Introduction*. NY: Cambridge University Press.
- Desmarais, B. A. and Cranmer, S. J. (2012). Statistical inference for valued-edge networks: the generalized exponential random graph model. *PloS one*, 7(1):e30136.
- Durrett, Richard; Kesten, H. (1991). *Random walks, Brownian motion, and interacting particle systems*. Springer.
- Gallavotti, G. (1999). *Statistical mechanics: A short treatise*. Springer, New York.
- Hethcote, H. W. (2000). The mathematics of infectious diseases. *Society for Industrial and Applied Mathematics*, 42:599–653.

- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.
- Hudgens, M. and Halloran, M. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842.
- Ichino, N. and Schündeln, M. (2012). Deterring or displacing electoral irregularities? spillover effects of observers in a randomized field experiment in ghana. *Journal of Politics*, 74(1):292–307.
- Ichino, N. and Schndeln, M. (2012). Deterring or displacing electoral irregularities? spillover effects of observers in a randomized field experiment in ghana. *The Journal of Politics*, 74:292–307.
- J., H. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558.
- Kermack, W. O.; McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Mathematical, Physical and Engineering Sciences*, 115(772):700.
- Lenk, P. J. and Rao, A. (1990). New models from old: Forecasting product adoption by hierarchical bayes procedure. *Marketing Science*, 9(1):42–53.
- Liggett, T. M. (1997). Stochastic models of interacting systems. *The Annals of Probability*, 25(1):1–2.
- Lyons, R. (2011). The spread of evidence-poor medicine via flawed social-network analysis. *Statistics, Politics, and Policy*, 2(1).
- McConnell, M., Sinclair, B., and Green, D. (2010). Detecting social networks: design and analysis of multilevel experiments. In *third annual center for experimental social science and New York University experimental political science conference*.
- Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.
- Nickerson, D. (2008). Is voting contagious? evidence from two field experiments. *American Political Science Review*, 102(01):49–57.
- Nickerson, D. (2011). Social networks and political context. *Cambridge Handbook of Experimental Political Science*, page 273.

- Ostrovsky, M. and Schwarz, M. (2011). Reserve prices in internet advertising auctions: a field experiment. In *Proceedings of the 12th ACM conference on Electronic commerce*, EC '11, pages 59–60, New York, NY, USA. ACM.
- Scholz, F. W. and Stephens, M. A. (1987). K-sample anderson-darling tests. *Journal of the American Statistical Association*, 82(399):918–924.
- Shalizi, C. R. and Thomas, A. C. (2011). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociol Methods Res*, 40(2):211–239.
- Sinclair, B. (2011). Design and analysis of experiments in multilevel populations. In *Cambridge Handbook of Experimental Political Science*, page 906. Cambridge University Press.
- Sobel, M. (2006). What do randomized studies of housing mobility demonstrate? *Journal of the American Statistical Association*, 101(476):1398–1407.
- Tchetgen, E. and VanderWeele, T. (2010). On causal inference in the presence of interference. *Statistical Methods in Medical Research*.
- Toulis, P. and Kao, E. (2013). Estimation of causal peer influence effects. In *Proceedings of the 30th International Conference on Machine Learning*, ICML '13.
- VanderWeele, T. (2008). Ignorability and stability assumptions in neighborhood effects research. *Statistics in medicine*, 27(11):1934–1943.
- VanderWeele, T. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Methods & Research*, 38(4):515.
- VanderWeele, T. and Tchetgen, E. (2011). Bounding the infectiousness effect in vaccine trials. *Epidemiology*, 22(5):686–693.
- VanderWeele, T., Tchetgen, E., and Halloran, M. (2011). Components of the indirect effect in vaccine trials: identification of contagion and infectiousness effects. *COBRA Preprint Series*, page 85.
- VanderWeele, T. and Tchetgen Tchetgen, E. (2011). Effect partitioning under interference in two-stage randomized vaccine trials. *Statistics & probability letters*, 81(7):861–869.
- VanderWeele, T., Vandenbroucke, J., Tchetgen Tchetgen, E., and Robins, J. (2012). A mapping between interactions and interference: implications for vaccine trials. *Epidemiology*, 23(2):285.

Walker, D. and Muchnik, L. (2014). Design of randomized experiments in networks. *Proceedings of the IEEE*, 102(12):1940–1951.