11-4-2014

Design of Randomized Experiments in Networks

Dylan Walker
Chapman University, dylan@chapman.edu

Lev Muchnik
Hebrew University of Jerusalem

Recommended Citation

D. Walker and L. Muchnik, "Design of Randomized Experiments in Networks," in *Proceedings of the IEEE*, vol. 102, no. 12, pp. 1940-1951, Dec. 2014, https://doi.org/10.1109/JPROC.2014.2363674

Design of Randomized Experiments in Networks

Design of Randomized Experiments in Networks

Dylan Walker
Boston University School of Management
dtwalker@bu.edu

Lev Muchnik
Hebrew University of Jerusalem School of Business Administration
lev.muchnik@huji.ac.il

As our day-to-day activities become increasingly embedded in online and digitally enabled environments, the availability of massive scale yet highly granular data on individuals and social interaction enables new avenues of scientific discovery. The promise of big data [1], [2] seems immense – not just for its scale and scope, but perhaps more importantly because highly detailed individual-level data at scale suggests tailored policies that resist reversion to the mean in domains ranging from medicine and public health [3]–[5] to politics, web search [6], business [7], e-commerce [8], and product design [9]. Yet, the promise of big data has recently come under fire for its inability to separate correlation from causation – to derive actionable insights and yield effective policies [10], [11]. This criticism unveils the perhaps lesser known but burgeoning movement of *big experiments* that is rapidly gaining traction within both academic research and industry practice. The gold standard of causal inference through experimentation is well established in both the public and private sectors [12]–[14]. Yet, the realization that our world is highly connected and that behavioral and economic outcomes at the individual and population level *depend* upon this connectivity challenges the principles of experimental design that lie at the very heart of the scientific process.

Traditional experimental designs that randomly assign populations to control and treatment groups to measure the comparative outcome of a treatment do not account for the networked environment in which we live – the natural connections between subjects in these populations. When the impact of treatment can propagate along these connections, the traditional notions of experimental design break down. It is perhaps unsurprising that this realization has chiefly emerged from the blossoming interdisciplinary field of computational social science [15], where the focus of study is on social behaviors that are, by their nature, interactive. Yet the implications of connectivity on experimental design are far reaching and necessarily affect scientific inquiry in multiple domains including medicine, public health, media, politics, business, biology, epidemiology, sociology and many others.

However, 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 treatments that incorporate both experimental subjects and the connections between them. Done correctly, networked treatments can allow us to understand the basic dynamics of contagious phenomena that have been found to play a critical role in individual and population level outcomes, such as: the effect of dosage or multiple exposures on individuals and populations[16]–[19], the decay of spreading behavioral and economic outcomes across social distances [20], and the impact of heterogeneity in individual- and relationship characteristics on spreading [21]–[28]. In turn, such an understanding will allow us to assess and compare policies designed to promote positive contagions and contain or discourage negative contagions.

In this work, we consider several aspects of networked randomized trial design from the perspective of the experimental setting, the process being studied, and the impact of connectivity. We further address emerging methods to analyze and draw statistical inferences from networked randomized trials. Finally, we present several categories of novel networked treatment designs and discuss their potential to future research.

I. The Impact of Setting and Process on Networked Randomized Trial Design

Over the past several years, the use of randomized controlled trials in networked environments have increasingly been employed by researchers across a variety of disciplines. Though these works share in common the feature of highly interconnected environments in which they take place, they differ significantly in both intent and approach. Networked randomized controlled trials (NRCTs) can be classified along two dimensions: the *setting* in which they are conducted and the *process* they are designed to investigate. We first consider the setting in which an NRCT is conducted, which has a number of implications on aspects of experimental design, relationship to the networked environment, and on interpretability or generalizability of findings.

I.1 Setting

There are three primary settings in which NRCTs can be conducted: offline laboratories, online laboratories, and field experiments in real-world settings (often referred to as experiments "in the wild"). The main differences between these settings are: the extent to which the experimenter can control her subjects and the context, the extent to which subjects are aware that they are participating in an experiment, whether networked environments are artificially imposed or organic, the potential scale of the experiment (in terms of population size and experiment duration), the ability to run repeated experiments, the ability to recruit and maintain subject participation, and the amount and type of information on subjects that is available for posterior analysis.

Offline laboratory settings have traditionally been used in the fields of psychology, economics and sociology [29], [30]. In this setting participants are typically recruited, invited into a highly controlled physical environment and given instructions on how to participate across multiple phases of the experiment according to well-established protocols. Offline lab settings offer the advantages of strict experimenter controls (over the conditions of the environment itself, constraints on subject behavior, and the nature of subject interactions and information flow). For example, experimental subjects can be deliberately primed, exposed to controlled situations, given background and even instructed to act or interact with one another in a particular manner. The advantage of strict control, however, is often accompanied by the important tradeoff that subjects are explicitly and constantly aware of their role as experiment participants and this awareness may cause them to act, react and interact differently from their natural behavior in organic environments and in cases where they do not believe their behavior is being observed and assessed. This limitation may have important implications on the generalizability and applicability of findings to policy considerations [31]. Beyond aspects of control, subject recruitment is often limited by proximity to the lab, time availability, and effectiveness of recruitment incentives. The limitations of subject recruitment has two important implications: First, it constrains the demographics of experimental populations (and thereby the generalizability of findings) and second, the overall population size and duration of the experiment. In addition, the research questions that can be addressed by the

experiment are frequently limited by the premises at which it is set. Offline lab settings also have advantages and disadvantages with regard to the networked environment itself. In these settings, researchers may completely specify network connections between subjects and control communication or other forms of interaction along these links. However, networks imposed by researchers may be very different in structure from organic networks and artificially imposed network links may lack real social context, potentially making them a poor proxy for the real social environments they may intend to represent. These conditions facilitate investigation of well-defined situations such as a collaborative solution to the network coloring problem [32], [33], convergence to consensus through biased voting [34] or the impact of network structure on the performance of prediction markets [35], [36].

Online laboratory settings are relatively recent and primarily facilitated by the pervasiveness of online technologies and the emergence of online social network platforms and micro-labor markets (such as Amazon's Mechanical Turk [37]). These settings replicate the spirit of the offline lab in that subjects are explicitly aware that they are participating in an experiment, may be primed, given background information and requested to act or interact with one another in a particular manner. To some extent online lab settings reduce constraints on experimental scale and subject recruitment in terms of geographic proximity and duration of participation. Importantly, unlike their offline counterparts, online lab settings can leverage existing platforms to enable subject recruitment at significantly reduced costs [38], [39] and thus have the potential to enable experimentation at much larger scales. Though, like their offline counterparts, online lab settings may also suffer from concerns of generalizability arising from the makeup of micro-labor markets employed for subject recruitment [40]. In addition, these settings can also leverage the APIs of existing platforms or data sharing agreements with their operators to collect detailed information about subjects, their social network connections and to control or mediate subject interactions [41]–[44]. However, these environments necessarily sacrifice strict experimental control in terms of the conditions of the offline environment itself, constraints on subject behavior, and the nature of subjects (potentially unrecorded) offline and online actions and interactions, as well as information flow to and (in some cases) between subjects. Experiments in online lab settings also face a number of new challenges such as maintenance of subject participation (e.g., user churn) and concurrency of subject participation (i.e., experiment design may require simultaneous presence of the subjects)[1] [40]. As in the case of offline labs, the findings and inferences from experiments conducted in online lab settings may have limited applicability to real world environments because individual behavior may be affected by the knowledge that subjects are part of an experiment and are being observed. Unlike offline lab settings, online labs that leverage existing social network platforms permit experiments in real networked environment while exerting some degree of control of interactions and information flow along network links (e.g., [44]). Thus, online lab settings circumvent some limitations of their offline counterparts making them uniquely suited to address well-defined situations like the role of network in cooperation [39], [45], public goods [39] and investment games [44] as well as its impact on health behavior [41]–[43] .

In contrast to offline and online labs, field experiments in real-world settings do not exert strong controls over subjects' environments, but instead assess the impact of randomized assignment directly in the natural environment of the system being studied [46]. Online field experiments in particular can provide researchers with detailed data on subject behavior (online and even offline[2]) that is not biased by

[1] See, for example, the sections on dropouts and the waiting room in [39]

[2] For example, many online platforms are location-aware.

knowledge of participation in the experiment[3] and can be conducted at extraordinarily large scales and over arbitrarily long durations. In some sense, online field experiments are a natural extension of A/B testing procedures that have become part of the standard policy for large online platforms to assess and evaluate features or the impact of platform design elements on the overall user experience [13], [47]–[49] . Because these settings facilitate experiments that can be conducted without or with limited subject knowledge, care must be taken to assess the ethical considerations of these practices and to abide by the standards of practice governing human subjects research. This concern tends to be more central to experiments addressing fundamental social science or economics research questions than in the case of routine A/B testing. Controversy surrounding recent research employing an online field experiment to study emotional contagion on Facebook emphasize these concerns [50], [51]. In addition, researchers that conduct experiments in real-world networked environments with treatment impacts that can propagate should also consider the ethical implications of treatment impact on individuals outside of the experimental population. It should be noted that despite the necessity for strong ethics, field experiments in real-world settings provide strong inferences and insights directly applicable to real-world systems and thus play a critical role in assessing the potential efficacy of important social and economic policies. Additional aspects of design of field experiments in natural settings relate to the concerns that the desired interventions should appear to be organic, in some cases not clearly detectable between subjects, and generally should not observably interfere with the normal operation of the community, platform, or online system. These concerns are important for rigorous experimental design but also to assure that experimental interventions do not adversely affect the business of firms that collaborate with researchers. Like online lab settings, online field experiments are limited by the capacity for experimenters to design interventions or otherwise control the environment. For example, it may be more difficult to expose subjects to a desired intervention and other experimental controls may be limited by the features of the online platform or system. In many cases field experiments identify effects of specific platform features, such as the impact of (in)visibility of user activity on peer interactions on an online dating site [52], the impact of social cues in word-of-mouth advertising on ad performance [53], the role of social platforms in diffusion of information [54] , the study of the mechanism of coupon sharing on Facebook [55], and the value of the content author's identity in evaluation of that content by the reader on news aggregation web sites [56]. Besides the advantage of conducting the experiment in organic settings and thus observing natural behavior, real world settings often enable research at immense scale, facilitating observation of subtle effects or heterogeneous response to interventions.

Finally, while not the focus of this work, the occurrence of *natural experiments* arising from exogenous variation in real-world systems provide yet another setting in which researchers may pursue causal inference. Unlike formal experiments which require significant investments of time and resources from experimenters, occasionally induce ethical concerns and noticeably interfere with user experience, the potentially widespread occurrence of natural experiments may permit causal inference on observational data at large scales and may be used prior to experimentation to inform experiment design. Having discussed the implications of *setting* on the design of NRCTs we now turn to a discussion of the *process* that NRCTs are designed to investigate.

[3]In many cases, experiments conducted in natural settings may notify subjects indirectly through posted policies on user research in the platform or online site's terms of use.

I.2 Classification by Process

Experiments in networked environments can also be classified by the process they are designed to investigate, including the exploration of social and economic behaviors, the underlying dynamic microscopic and macroscopic mechanisms governing these behaviors, and the resulting dynamics of outcomes at individual-, group- and population-levels. Many NRCTs focus on investigation of propagation processes such as dissemination of innovation, spread of information and behaviors or adoption of new products. Identification of factors affecting these processes is vital for informing managerial or public policies intended to promote or discourage population-level outcomes. Factors that affect process dynamics include initial conditions (such as targeting or seeding), dosage and temporal aspects (such as the extent and timing of multiple exposures), the willingness of subjects to contribute, prevent or direct the viral spread, the susceptibility of subjects to peer influence, the social network topology, and modification of the process itself (e.g., viral product design [22]) In practice, policies may need to leverage one or more of these mechanisms to achieve a desired outcome [57] .

The process under investigation is often signified by how the experimenter measures the response to their intervention. When the effect of intervention/s can propagate, the experimenter is not limited to analyzing the response behavior of directly treated subjects, but may instead focus on the response behavior of other subjects or groups in the population (such as peers of directly treated subjects or groups of locally connected treated subjects). Analysis may therefore focus on one of three aspects: the direct effect of the *treatment on the treated* (ETT) subjects, the *effect of the treatment on the co-treated* (ETC), or the *effect of the treatment on the untreated* (ETU)[4]. To avoid ambiguity, we adopt a simple definition of *treatment* that is defined for each experimental subject as the alteration of that subject's experimentally controlled experience. We leave discussion of more complex networked treatments not covered by this definition to a later section. The *effect* of treatment may be measured at the level of individual subjects, aggregated over groups of subjects, or aggregated effects of the treatment on the population at large. In addition, researchers may be interested in how the effect is *moderated* by individual attributes, local-network attributes, or the global structure of the network. In this section, we categorize existing research by process and discuss the implications on experimental design and choice of setting.

Networked experiments that study processes concerned with the effect of treatment on the treated (ETT) represent the extension of conventional non-networked experiments to networked environments. It is important to note that for many processes of interest (particularly those that involve social components) traditional experimentation may be affected by an underlying network, even when the network is not explicitly observed or recognized by experimenters. In some cases, interaction between subjects may be an unavoidable nuisance while in other cases, it may be central to the process under investigation. For example, Bapna et al. (2013) study the impact of enabling anonymous profile-viewing for users of an online dating site, an intervention that is meaningless in the absence of social interaction [52]. In another experiment Bakshy, Eckles, Yan and Rosenn vary the number and the intensity of social cues accompanying online ads to establish the degree to which they can affect the ad performance [53]. One

[4] This terminology should not be confused with the traditional terminology of the Average Treatment Effect on the Treated (ATET) and the Average Treatment Effect on the Untreated (ATEU) which pertain to analysis in non-networked environment to provide counterfactual estimates that avoid selection bias in the designation of treated populations.

distinct class of research questions that focus on ETT addresses subjects' response to population level social signals such as conformation to peer pressure. First identified with the now classical sociologist techniques in lab or small-scale field experiments conducted in the 1950s and 1960s by Asch, Milgram and others [58], [59], these phenomena can now be examined at scale. For example, in a sequence of experiments Salganik, Dodds and Watts study the impact of popularity-based content ordering on the propensity to consume cultural products (music) [38], [60]. In these experiments, the authors randomize perceived popularity of songs to distinguish the impact of popularity on subjects' decisions to consume music from that of song quality. The rising prevalence of ranking and rating mechanisms in virtually every domain make these types of experiments both theoretically and practically important. The sheer scale of the data can permit subtle inferences that require high sensitivity and provide enough resolution to understand the moderating effect of different personal and content attributes on consumption decisions. Depending on the setting and control available to experimenters, connectivity between subjects may allow for undesirable spillover effects (interference) that contaminate or bias inferences on ETT. The potential scale and the scope of networked experiments can facilitate inferences on the ETT that emerge as a result of feedback mechanisms (due to spillover effects). For example, individuals subject to treatment can indirectly influence others by contributing their manipulated opinion to population-level social signals. Spillover effects may effect both untreated and treated subjects (through feedback) resulting in herd-like phenomena that has the potential to sway collective (population-level) behavior, potentially in undesirable ways, leading to important implications for rating, ranking and collaborative feedback systems. This effect was demonstrated at scale by Muchnik, Taylor and Aral who show that minor manipulation of the perceived scores of user-generated comments guided consequent user votes and resulted in herding, significantly affecting the content's final score [61]. Subsequent research by Godinho de Matos et al (2014) found that manipulation of rank and population level social signals for Video on Demand titles yields only a short-lived effect on herding behaviors, emphasizing that such signals may be highly context dependent [62]. From the experiment design perspective, the feedback following the randomized manipulation of the content or its ranking may mingle the treatment with endogenous processes. In fact, due to the aggregation of the collective opinion into population level social signals, all but the first impression following the treatment are conditional on the response (or non-response) of the preceding subjects. We discuss detailed strategies to address the related interference issues in section II.

Networked experiments that study processes concerned with the effect of treatment on the co-treated (ETC) include processes that involve local network externalities. Understanding such processes is central to explaining the value of network goods, products or features and necessary for modelling of propagation of knowledge, rumors and information in general. The recent emergence of pervasive online social platforms enable experimentation on ETC-related processes that can yield relevant insights of value to both platform owners and academics. For example, recent work on network bucket testing extends A/B testing procedures to assess the ETC of social product features. Additionally, many platform sponsors assess social features through beta rollouts (e.g., gmail) that allow users to invite their peers to co-adopt, making inferences on ETC of high practical importance. For this reason, studies of ETC are often conducted in real-world settings, though the available controls of offline and online lab settings make them equally suitable to study co-treatment.

Processes concerned with the effect of treatment on the untreated (ETU) are the focus of the rapidly expanding field of research into contagious phenomena across multiple disciplines. Many recent randomized experiments conducted in networks examine contagion processes in the context of diffusion

of behaviors (e.g., voting [20] or health behavior [41]), emotions [50], peer influence or product placement (seeding) [21], [63], [64] which aspire to inform polices aimed at the promotion or containment of contagions in social networks. These studies are designed to identify the impact of a variety of factors on contagious spreading. Several studies investigate the moderating effect of individual characteristics [52], dyadic properties [21], [53], [64] as well as the impact of attributes of a spreading product, norm or information [20], [22], [50] on diffusion processes on networks. Causal identification of factors that affect subject behavior can be achieved through exogenous manipulation of these factors, allowing researchers to distinguish causal impact from alternative explanations of correlated behavior such as homophily, assortative mixing and other endogenous confounds [65]–[67]. These experiments are based on selective application of treatments to focal subjects and observation of the response of their immediate or remote peers. Typical treatments include randomized gifting, variation of pricing or manipulation of product features. More sophisticated treatments focus on randomly controlling the interaction between individuals and their peers, aiming to test how peer influence is moderated by subject, peer and dyadic characteristics. For example, Aral and Walker test the moderating effects of individual and dyadic characteristics on word of mouth by issuing Facebook notifications to randomly chosen peers of experimental subjects [64]. Such networked treatments go beyond the definition of simple treatment that we have adopted here. We discuss these types of treatments in more detail in section III. Other *contagion* experiments examine the effect of local and large-scale network topology on diffusion of information and behaviors [32]–[34], [39] . The general goal of these studies is to detect the effect of network attributes (such as degree, clustering, correlation) on network diffusion processes and the effect they have on individual and collective behavior such as convergence to consensus [34], [39], a collaborative solution to network coloring problems [32], [33], and the spread of health-related behaviors [41]–[43]. Experiments on contagion processes may constrain the choice of experimental setting. For instance, exogenous manipulation of local or global social network structures requires a setting where tight control over individuals' connections is possible. Such control can be achieved in offline or online laboratory settings where the experimenter has full control over connections and/or information visibility, but may be less feasible in real-world settings where connections emerge organically and cannot be exogenously manipulated[5] [32]–[34], [39], [41], [42]. On the other hand, studies that examine the (potentially subtle) impact of individual attributes on contagions typically require rich datasets at large scales that cover the wide spectrum of attributes. Such studies are therefore best suited to settings that enable large scale experimentation where data is readily available, as is the case for real world field experiments conducted on online platforms.

II. Impact of Connectivity on Randomized Trial Design

The consequence of connectivity on inference in randomized trials is best understood by examining the Rubin Causal Model, which presents a fundamental approach to drawing causal statistical inferences from randomized experiments. A chief assumption of this approach is the Stable Unit Treatment Value Assumption (SUTVA) which demands that the observed outcome on one unit (subject) should not depend upon treatment assignments to other units (subjects) [68], [69]. When the effect of treatment can

[5] We note that online lab settings and real-world online settings that utilize platform features may effectively alter network structures exogenously by disabling certain types of interactions, rendering these settings suitable for studies aimed at inferring the impact of local or global network structures on contagion dynamics.

propagate, this assumption is violated and the standard machinery of statistical inference from randomized trials must be re-evaluated. In some cases, where propagation of treatment effects are well understood, the SUTVA can be re-established by redefining treatment to multiple treatment specifications that include indirect spillovers. However, in the highly connected environments in which we are interested (and particularly where the nature of treatment propagation is unknown), simple re-specification of treatment to rescue the SUTVA is not feasible.

In recent work, Manski (2013) has taken the first steps toward building a theory of identification in the presence of interference by extending the SUTVA from the classical assumption of non-interacting units, which he refers to as the assumption of Individualistic Treatment Response (ITR), to define multiple classes of assumptions based on the nature of interaction (or lack of interaction) between units. For example, he defines the assumption of Constant Treatment Response (CTR) as the case when each individual in the experimental population has some reference group (of other units or subjects) for which his or her outcome remains constant when treatment varies beyond his or her reference group. He further relates these assumptions to models of endogeneous interactions through systems of simultaneous equations that connect treatment and outcomes of all individuals in the population to the outcome of any particular individual. These considerations lead to restrictions on when inference from observed outcomes can be point-identified and, importantly, how this relates to treatment designation [70].

Practical strategies to account for connectivity in randomized trials are currently an active topic of research and fall into two general categories: *inference* strategies and *design* strategies [64]. The former strategies address interference after an experiment has been conducted, during the inference or analysis phase, while the latter strategies address the potential for interference prior to experimentation by modifying aspects of the design of randomized trials, such as treatment assignment procedures, to minimize interference.

To clarify our discussion of these strategies, we introduce some terminology to describe treatment and exposure to treatment. For the purposes of simplicity we assume for now that experimental treatments apply directly to individuals (or units) in the population and leave complex treatment types that may include simultaneous experimental controls on individuals, their peers and the nature of their interaction/s, for subsequent discussion. We also assume, for simplicity of discussion, that treatments are temporally static, assigned prior to the experimental period and consist of only one kind of treatment (i.e., *treatment* or *control*; though these definitions may easily be extended to the case of multiple treatment types). We define direct treatment as the alteration of each individual's experimentally controlled experience throughout the course of the experiment, as specified by the *direct treatment* vector, T_i^{dir}, where i indexes experimental subjects. This follows the conventional usage of the term *treatment* in traditional RCTs and its assignment is directly controlled by the experimenter. In contrast to direct treatment, we define indirect treatment as the experience induced on peers of directly treated users (through their direct connection or through one or more pathways of multiple connections in the network) as a consequence of direct treatment, as specified by the *indirect treatment* vector, T_{ik}^{ind}. Unlike direct treatment, indirect treatment is not exogenously assigned, but arises instead from both direct assignment, the (often endogeneous) network itself, and the (often endogeneous and unknown) dynamics of propagation of the *impact* of direct treatment. The subscript k is included to enumerate the multiple types of indirect treatment that arise through exposure even given our assumption of one kind of direct treatment. For example, one type of indirect treatment may be defined as having one and only

one treated neighbor (regardless of treatment assignments at larger network distances from the subject); another type may be defined as having two treated neighbors who are not connected to one another (regardless of treatment assignments at larger network distances from the subject). As these examples suggest, the multiplicity of indirect treatment types depends on assumptions about exposure and propagation. As a consequence, indirect treatment may also be time dependent. For completeness, we also define the total effective treatment as the combination of both direct and indirect treatment, T_{ik}^{tot}.

II.1 Inference Strategies

Inference strategies attempt to remove or reduce bias and/or variance from estimates that identify the impact of treatment in connected settings and typically assume a specified vector of treatments. Inference strategies are distinguished by the type of estimation strategy, from the fundamental estimate of the average impact of treatment (ATE) to more sophisticated modeling techniques.

In some recent work, researchers have developed methodologies to partially account for statistical interference in NRCTs (Networked Randomized Controlled Trials) through modified average treatment effect (ATE) estimators with reduced bias [71], [72]. In these methodologies, an exposure model is assumed and employed to enumerate multiple total treatment types, T_{ik}^{tot}. A modified Horvitz-Thompson or Hajek estimator is then constructed that accounts for the bias introduced by the propensity to receive any of the total effective treatment types. The multiplicity of the total effective treatments is determined by assumptions of the exposure model. For example, for an exposure model that assumes propagation to fall to zero beyond one network link, all individuals with no treated neighbors will have a total effective treatment equal to their direct treatment, regardless of the treatment status of peers at network distance greater than one. However, it is clear that for arbitrary exposure models (where the propagation of the treatment effect may not fall sharply with increasing network distance), estimating the causal impact of the treatment becomes severely limited, as the number of potential indirect treatments (K) becomes increasingly large relative to the size of the experimental population. Coppock and Sircar (2013) summarize this difficulty succinctly:

> "The basic difficulty inherent in design of experiments facing interference between units is that it reduces power. If units are exposed to complex spillovers, the outcomes revealed by those units are not useful for the estimation of any quantities of interest."[73]

Nonetheless, modified estimator approaches may be particularly fruitful when strong assumptions of limited propagation apply or when the experimenter can exert strict control over propagation. However, when knowledge of the propagation (and hence exposure) is unknown, practitioners must turn to empirical evidence to first adjudicate between multiple potential exposure models. It is important to note that the statistical interference methods discussed above are not designed to discriminate between different exposure models. This highlights a critical challenge in analysis of networked experimentation in novel contexts: researchers must simultaneously estimate both the *treatment impact* and the *nature of exposure dynamics*.

Other inference strategies go beyond modification of estimates of the average treatment effect (ATE) impact and incorporate constraints on inference in more sophisticated approaches to model treatment impact. Modeling in NRCTs has three primary advantages over ATE estimation. First, use of models that

incorporate interactions of characteristics or attributes with both direct and indirect treatment allow inferences surrounding the heterogeneity of treatment impact. Such inference can be used to understand and predict how different subpopulations would respond to treatment[6]. This is particularly important from the standpoint of personalized policy development. While true assessment of the efficacy of personalized policies should be verified by evaluating interventions specifically designed to affect targeted subpopulations, inferences on heterogeneous treatment impact can act as a guide to develop personalized policies by identifying subpopulations (from the wide range of possibilities) for which treatment impacts significantly differ. Second, modeling permits identification of moderators of treatment impact *ceteris paribus*, allowing researchers to partially disentangle the treatment impact of correlated characteristics, provided there is significant diversity in subject populations. Third, modeling strategies allow researchers to employ tools such as censoring, stratification, and matching to estimate the impact of indirect treatment on individuals who have received different exposures relative to those in appropriate reference groups that have not. In models that employ duration analysis, censoring techniques can be used to censor outcomes of users only after they are exposed to complex indirect exposures. This technique allows researchers to reduce bias in estimates of treatment impact while both retaining the maximal amount of outcome data in their analysis and correctly parameterizing their ignorance regarding what might have happened had complex indirect exposure not occurred. For example, Aral and Walker (2011) employ censoring in hazard models to exclude subject outcomes from analysis only after they have been indirectly exposed to multiple treated peers (with potentially different treatments) [22]. Unlike the modifications to ATE estimators discussed above, censoring techniques do not require complete specification of an exposure model, but instead assert limiting assumptions regarding exposure in exchange for both a loss of statistical power for censored observations and the inability to estimate the impact of some complex exposures. This is an important tradeoff. Stratification (in non-duration modeling) and dynamic risk group assignment (in duration modeling) further allow researchers to partition subjects according to different indirect exposures they may have received and separately estimate the impact of these indirect exposure on subject outcomes. Stratification on indirect exposure types are also subject to assumptions regarding the nature of exposure (as indirect exposure types must be specified prior to stratification), but also do not require a complete specification of the exposure model. Instead, complex exposure types excluded from any strata or risk group are effectively censored (in duration models) or truncated (in non-duration models), sharing the associated tradeoffs with censoring discussed above. Importantly, because indirect exposure is endogenously determined (by the natural connectivity of the network and in some NRCTs potentially also by endogenous propagation of the treatment), concerns of generalizability apply. Specifically, populations receiving different types of indirect exposure to treatment may be fundamentally different (in terms of observable and unobservable/latent characteristics) from the representative population at large. Researchers employing these techniques should therefore take care in generalizing inferences on the impact of complex indirect exposures to situations that would not arise organically (such as policies that exogenously determine complex indirect exposures). Matching techniques generally may be employed to establish appropriate reference groups and specifically to control for the propensity to receiving a particular type of indirect exposure to treatment and to balance the makeup of direct or indirectly treated populations relative to controls. Matching on propensity to receive a type of indirect treatment is comparable to the modifications to ATE estimators discussed above, but matching techniques can be generalized to

[6] Recent work on analyzing heterogeneous treatment effects with dependent data provide a variety of bootstrap methods to properly handle uncertainty [85].

simultaneously account for both propensity to be exposed and endogeneous variations in the makeup of subpopulations that receive different types of direct and indirect exposure. It is important to note that relative to fundamental ATE approaches, modeling approaches may often involve strong assumptions regarding the mechanism of treatment response. For this reason researchers must establish that these assumptions are theoretically grounded and reasonable through empirical validation and ensure that the robustness of inferences to model specification are thoroughly explored.

II.2 Design Strategies

In contrast to inference strategies, design strategies alter aspects of the design of the experiment itself in order to constrain the manner of interference between subjects. Typically design strategies involve rearranging assignment of treatment to subjects in a manner that incorporates information on network connectivity. Existing design strategies fall into two categories: treatment clustering strategies and treatment separating strategies. Treatment clustering strategies seek to closely approximate the counterfactual conditions in which the entire network is exposed to either treatment or control by assigning subjects in well-defined local sub-networks the same treatment. In contrast, treatment separating strategies seek to assign treatments to experimental subjects that are well-separated from one another in network distance in order to minimize interference. Existing design strategies to deal with interference also differ by whether they are appropriate for making unbiased inferences on the effect of treatment on the treated (ETT), co-treated (ETC), or untreated (ETU) members of the populations and the extent to which they are suitable for empirically inferring (rather than assuming) exposure dynamics.

Treatment clustering strategies relate treatment designation to the natural structure of the network in terms of clusters, components and communities[7]. These strategies stem from attempts to extend A/B testing to networked environments where treatment is oriented around enabling new features or products in social network platforms that exhibit strong local network externalities. For example, the evaluation of a new social messaging feature would be inaccurate if the feature was not simultaneously available to individuals and their direct network peers with whom they would typically communicate. Treatment clustering strategies use (a variety of) algorithms to assign the same treatment to clusters of well-connected nodes [17], [74]. Ugander et al. (2013) use the terminology “network exposed” to describe the condition under which an individual and some sufficient number or fraction of his or her peers have received the same (direct) experimental treatment. They show that, using their technique of *graph cluster randomization*, an efficient dynamic program can used to exactly calculate the probability that each individual in the network is network exposed. When an exposure model is specified, these probabilities can be used in modified ATE estimators to reduce bias. Moreover, they also show that under the right conditions, graph cluster randomization can significantly reduce ATE estimator variance [75]. Airoldi et al. (2013) also consider a Simple Sequential Randomization algorithm that clusters direct treatments in local networks as well as an Insulated Neighbor Randomization algorithm that relaxes treatment conditions to partial neighborhoods to yield a higher probability of valid causal estimates where treatments are matched with counterfactual controls [76]. In subsequent work, Eckles et al. (2014) point out that many

[7] Clusters are subgraphs into which an overall network is partitioned according to some clustering rule. Components are sets of nodes that are connected to one another via network paths of any length. Communities in networks are defined as sets of nodes which are well connected to one another and relatively sparsely connected to other nodes in the population.

tractable exposure models do not realistically account for the role of peer effects in mediating exposure. Instead, they consider dynamic outcome generating processes in discrete time for which a subject's response at time t depends upon their own direct treatment, as well as the direct treatment and behavior of their peers at time $t-1$. Outcome generating processes go beyond specification of exposure alone and specify a mechanism by which responses are induced by direct and indirect treatment. They employ graph cluster randomization on several artificial models of networks assuming fractional neighborhood treatment response (FNTR) in which a subject is assigned the treatment condition of a specified fraction of their peers. Subjects without a sufficient fraction of peers assigned to treatment or control conditions are excluded from analysis. Using simulations of outcome generating processes on artificial network models, they show that in the presence of peer effects that mediate exposure, graph clustering randomization can reduce bias in modified ATE estimators with comparably small increases in estimator variance when the network itself exhibits sufficient clustering [72]. Thomas and Finegold (2013) employ a form of indirect treatment clustering. They consider random treatment assignment and use a pseudo-randomized trial (where mock treatment designation does not alter user experience whatsoever) to demonstrate that simple t-tests on the impact of indirect treatment (ETU) spuriously bias p-values towards zero. They implicitly assume exposure does not extend beyond a network distance of one and consider permuting direct treatment assignment so that all peers of directly treated subjects have the same unequivocal indirect treatment designation. They show that clustering of indirect treatments restores uniform p-value distributions for t-tests on the impact of indirect treatment, as would be expected given the mock nature of the treatment [77]. As the above discussion should make clear, treatment clustering strategies can reduce bias and variance in inferences on ETC. However, these strategies necessarily reduce heterogeneity in types of indirect exposure, making them less suitable for inferences on the effect of the treatment on the untreated (ETU), including the ability to empirically evaluate the dynamics of contagious phenomena, such as how multiple indirect exposures add together or how exposure decays over social distance. In some cases, *indirect* treatment clustering strategies may be appropriate for inferring the effect of treatment on the untreated (ETU), when exposure does not extend beyond a network distance of one. Importantly, treatment clustering strategies may yield unbalanced assignment of nodes to treatment conditions in terms of individual-level or network characteristics of subjects (such as degree). Specifically, Ugander et al. (2013) point out that subjects with high network degree are less likely to be assigned to extreme definitions of co-treatment (e.g., an effective treatment where most or all of a subject's peers have the same treatment) [75]. Likewise, Thomas and Finegold (2013), who primarily focus on the impact of indirect treatment, discuss concerns of selection bias for indirectly treated subjects in terms of bias in the distributions of individual characteristics (that may arise from e.g., homophily), and network characteristics (such as degree), that arise as a consequence of designating treatment either randomly or with treatment clustering strategies. While reweighting designation of direct treatment can alleviate selection bias on indirectly treated subpopulations, it necessarily induces selection bias in the directly treated subpopulations, as they point out [77]. One promising approach to address concerns of balance is presented in the recent work on by Nishimura and Ugander on graph partitioning [78].

Treatment separating strategies attempt to reduce interference between subjects by constraining direct treatment assignment to subjects that are well separated from one another. Coppock and Sircar (2013)

define the SUTVA degree (λ) as the network distance beyond which spillover does not occur[8]. In this methodology, well-defined direct and indirect treatment types on which the experimenter would like to make inferences are specified in advance and all other (complex) exposures to treatment are minimized through a two-stage random direct treatment assignment algorithm that incorporates the assumption of the SUTVA degree. Modifications of the direct treatment assignment algorithm can be performed to permit inferences on the dynamics of contagious phenomena such as how indirect exposures to treatment add together or decay over social network distance. Analysis procedures may also employ modified estimators or other modeling techniques that adjust for the propensity to receive an indirect treatment [73]. Consequently, treatment separating strategies are ideal for estimating the effect of the treatment on the untreated (ETU). Because this strategy primarily seeks to separate treated subjects from one another in network distance, it is less appropriate for inferring the effect of the treatment on the co-treated (ETC) when a substantial number of co-treatments amongst directly connected individuals is desired. It is important to note that the assumption of a SUTVA degree excludes cases when maximal spillover distance can depend on the number of indirect exposures. For example, in complex contagion scenarios, a subject may be more likely to be affected by multiple directly treated peers at distance $\lambda + 1$ than by a single peer at the same distance. Post-hoc inferences on how exposure adds and decays over social network distance (within the SUTVA distance) obtained from treatment separating approaches can be examined to evaluate whether this is a concern. Practitioners may wish to modify the direct treatment assignment algorithm to reduce multiple exposures at the cost of reducing treated population sizes and statistical power. Just as treatment clustering schemes may induce selection bias in individual-level or network-level characteristics of directly or indirectly treated populations as a consequence of clustering, treatment separating schemes may also induce a similar selection bias. The algorithmic removal of subjects within a distance λ from treated and indirectly treated subjects from consideration to receive a direct treatment could impact the balance of treated and indirectly treated subpopulations in terms of individual and network-level characteristics. The presence of homophily on individual-level characteristics in a variety of real-world networks emphasizes this concern. As such, care should be taken to ensure that directly and indirectly treated subpopulations are balanced with respect to one another and any reference groups. When this is not the case, the two stage random direct treatment assignment algorithm can be modified to reassert balance.

Interestingly, both treatment clustering and treatment separating strategies require assumptions about exposure distance introduced through the choice of the cluster size in the former case, or through the specification of SUTVA degree in the latter. When empirical evidence is unavailable to inform these decisions, practitioners may employ combinatorial designs to vary cluster sizes in treatment clustering strategies or to empirically infer decay of exposure across social distance in treatment separating strategies. In many circumstances, quasi-experiments that apply matching to observational data may act as a useful guide to inform experimental design surrounding exposure assumptions and the requisite statistical power necessary to infer significant effects [63], [66], [79]. In addition, both treatment clustering and treatment separating strategies assume that the network structure is known. While unbiased sampling can be achieved through a variety of means, e.g., [80], [81], it may not always be feasible. When only partial information on network structure is available, adaption of the strategies

[8] The definition of SUTVA distance is closely related to the concept of r-nets in metric spaces, which is discussed by Ugander et al. (2013)[75].

presented here in combination with network sampling techniques may be required. This is another avenue for potential future research.

III. Networked Treatments

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. Where simple treatments are defined as those that are applied to and alter an individual subject's experience, networked treatments involve interventions that may alter how connected subjects interact with one another, encourage or incentivize a subject to promote or influence the actions of one or more peers in a particular way, affect shared experiences and interactions between groups of subjects, or even encourage the formation of new connections between subjects. Such networked treatments are in part made possible by the emergence of online social networking platforms and other digital social environments that permit firm-mediation of social interactions to both platform owners and to other researchers through APIs [49], allowing for varying degrees of experimental control along the *channel* of social interaction [55], [82]. Networked treatments also enable experiments that can act as important test beds for emerging social policies aimed at producing or altering population-level change. Categories of networked treatments include *peer-oriented incentive schemes, communication altering schemes, subject-grouping schemes,* and *network topology manipulation schemes. Peer-oriented incentive schemes* reward subjects when their peers take a particular action such as purchasing a product or service [83], making certain choices [34], [39], spreading a particular piece of content or message (such as encouragement to have a flu shot or get an HIV test), or encouraging referral chains [84] that yield desired outcomes (such as a solution to a crowd-sourced problem). *Communication-altering schemes* may send automated referrals from a subject to their peer [22], randomize the target of automated messages from a subject to randomly chosen subsets of their peers [21], or even block or moderate information exchange between subjects [50], [83]. *Subject-grouping schemes* may randomly designate experimental subjects to social environments (such as pairing participants with online health buddies [41]–[43] or designating subjects to online study groups in MOOCs) contingent upon subject or environmental characteristics. *Network topology manipulation schemes* are designed to test the implications of network topology for social computation processes such as collaborative problem solving of competitive games [32]–[34], [39], [45]. Depending on type, instantiation and context, networked treatments may either remain susceptible to or circumvent interference effects. Future research should evaluate when and to what extent emerging design and inference strategies to address interference can be extended to networked treatments or whether new strategies are required.

IV. Conclusion/Discussion

The increasing prevalence of networked environments and the natural connectivity of our world presents both challenges to existing design and analysis methods for randomized trials and opportunities to conduct novel experiments involving networked treatments. It is likely that large-scale experimentation in social networks will lead to significant advances in the social sciences, just as conventional randomized controlled trials advanced medicine in the second half of the 20th century. However, just as the widening

use of RCTs in medicine, psychology and other domains necessitated the development of specialized methodologies and analysis techniques, the emergence of NRCTs introduces a number of new challenges, issues and concerns. While we have systematically reviewed emerging approaches to address these topics, the study of the implications of setting, process and connectivity on design and analysis of networked randomized trials is still very much in its infancy. Well-designed networked treatments and other novel approaches to the mechanism of randomization [64] may circumvent many of the issues discussed here. Future research employing networked treatment designs should thoroughly consider issues of inference in the presence of interference. More generally, practitioners conducting NRCTs should evaluate the suitability of the design and analysis strategies outlined here to their particular context. The dual challenge of estimating both the impact of experimental interventions that can propagate and the dynamics of propagation itself may call for the development of concurrent design strategies that allow for simultaneous empirical inferences on the former and the latter. The development of analysis techniques that can discriminate between multiple models of propagation or outcome generating processes is also an important avenue for future research.

Acknowledgements

We would like to thank Johan Ugander and the anonymous reviewer for their insightful comments and discussion.

Bibliography

[1] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, 1 edition. Eamon Dolan/Houghton Mifflin Harcourt, 2013, p. 256.

[2] F. Provost and T. Fawcett, *Data Science for Business*. O'Reilly Media, 2013, p. 414.

[3] M. A. Hamburg and F. S. Collins, "The Path to Personalized Medicine," *N Engl J Med*, vol. 363, pp. 301–304, 2010.

[4] T. B. Murdoch and A. S. Detsky, "The inevitable application of big data to health care.," *JAMA*, vol. 309, no. 13, pp. 1351–2, Apr. 2013.

[5] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care.," *Nat. Rev. Genet.*, vol. 13, no. 6, pp. 395–405, Jun. 2012.

[6] P. Brusilovsky, A. Kobsa, and W. Nejdl, *The adaptive web: methods and strategies of web personalization*. Springer, 2007, p. 766.

[7] H. Chen and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Q.*, vol. 36, no. 4, pp. 1165–1188, 2012.

[8] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Analysis of Recommendation Algorithms for E-Commerce,” in *EC ’00 Proceedings of the 2nd ACM conference on Electronic commerce*, 2000, pp. 158–167.

[9] J. Lampel and H. Mintzberg, “Customizing customization,” *MIT Sloan Manag. Rev.*, vol. 38, no. 1, pp. 21–30, 1996.

[10] G. Marcus and E. Davis, “Eight (No, Nine!) Problems With Big Data,” *The New York Times*, New York, New York, USA, 06-Apr-2014.

[11] D. Boyd and K. Crawford, “Six provocations for big data,” *SSRN Electron. J.*, 2011.

[12] “FDA Clinical Trials Guidance Documents.” 2014.

[13] R. Kohavi, “Online controlled experiments,” in *Proceedings of the 1st workshop on User engagement optimization - UEO ’13*, 2013, pp. 15–16.

[14] H. O. Stolberg, G. Norman, and I. Trop, “Randomized Controlled Trials.” American Roentgen Ray Society, 23-Nov-2012.

[15] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, “Computational Social Science,” *Science (80-.).*, vol. 323, no. 5915, pp. 721–723, 2009.

[16] D. Centola and M. Macy, “Complex Contagions and the Weakness of Long Ties,” *Am. J. Sociol.*, vol. 113, no. 3, pp. 702–734, 2007.

[17] L. Backstrom and J. Kleinberg, “Network bucket testing,” in *Proceedings of the 20th international conference on World wide web - WWW ’11*, 2011, p. 615.

[18] D. J. Watts and P. S. Dodds, “Influentials, Networks, and Public Opinion Formation,” *J. Consum. Res.*, vol. 34, no. 4, pp. 441–458, Dec. 2007.

[19] P. Dodds and D. Watts, “Universal Behavior in a Generalized Model of Contagion,” *Phys. Rev. Lett.*, vol. 92, no. 21, 2004.

[20] R. Bond, C. Fariss, J. Jones, A. Kramer, C. Marlow, J. Settleb, and J. Fowlera, “A 61-million-person experiment in social influence and political mobilization,” *Nature*, vol. 489, no. 7415, pp. 295–298, 2012.

[21] S. Aral and D. Walker, “Identifying Influential and Susceptible Members of Social Networks,” *Science (80-.).*, vol. 337, no. 6092, pp. 337–341, Jul. 2012.

[22] S. Aral and D. Walker, “Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks,” *Manage. Sci.*, vol. 57, no. 9, pp. 1623–1639, Sep. 2011.

[23] D. Godes, “Firm-Created Word-of-Mouth Communication: Evidence from a Field Test,” *Mark. Sci.*, vol. 28, no. 4, pp. 721–739, 2009.

[24] J. Goldenberg, S. Han, D. R. Lehmann, and J. W. Hong, “The Role of Hubs in the Adoption Process,” *J. Mark.*, vol. 73, no. 2, pp. 1–13, 2009.

[25] R. Iyengar, C. Van den Bulte, and T. W. Valente, “Opinion Leadership and Social Contagion in New Product Diffusion,” 2010.

[26] C. Van den Bulte and S. Wuyts, *Social Networks and Marketing*. Cambridge, MA: Marketing Science Institute, 2007.

[27] A. Goyal, “Social Influence and its Applications,” THE UNIVERSITY OF BRITISH COLUMBIA, 2013.

[28] E. Bakshy and J. Hofman, “Everyone’s an influencer: quantifying influence on twitter,” in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 65–74.

[29] M. Webster and J. Sell, *Laboratory Experiments in the Social Sciences*. Academic Press, 2007, p. 576.

[30] A. Falk and J. Heckman, “Lab Experiments Are a Major Source of Knowledge in the Social Sciences,” *Science (80-.).*, vol. 326, no. 5952, pp. 535–538, 2009.

[31] S. Levitt and J. List, “Field experiments in economics: the past, the present, and the future,” *Eur. Econ. Rev.*, vol. 53, no. 1, pp. 1–18, 2009.

[32] M. Kearns, S. Suri, and N. Montfort, “An Experimental Study of the Coloring Problem on Human Subject Networks,” *Science (80-.).*, vol. 313, no. 5788, pp. 824–827, Aug. 2006.

[33] S. Judd, M. Kearns, and Y. Vorobeychik, “Behavioral dynamics and influence in networked coloring and consensus.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 34, pp. 14978–82, Aug. 2010.

[34] M. Kearns, S. Judd, J. Tan, and J. Wortman, “Behavioral experiments on biased voting in networks.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 5, pp. 1347–52, Feb. 2009.

[35] L. Qiu, H. Rui, and A. B. Whinston, “The Impact of Social Network Structures on Prediction Market Accuracy in the Presence of Insider Information,” *J. Manag. Inf. Syst.*, vol. 31, no. 3, pp. 145–172, 2014.

[36] L. Qiu, H. Rui, and A. Whinston, “Effects of Social Networks on Prediction Markets: Examination in a Controlled Experiment,” *J. Manag. Inf. Syst.*, vol. 30, no. 4, pp. 235–268, 2014.

[37] G. Paolacci, J. Chandler, and P. Ipeirotis, “Running experiments on Amazon Mechanical Turk,” *Judgm. Decis. …*, vol. 5, no. 5, pp. 411– 419, 2010.

[38] M. J. Salganik, P. S. Dodds, and D. J. Watts, “Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market,” *Science (80-.).*, vol. 311, no. 5762 , pp. 854–856, 2006.

[39] S. Suri and D. J. Watts, “Cooperation and contagion in web-based, networked public goods experiments.,” *PLoS One*, vol. 6, no. 3, p. e16836+, Mar. 2011.

[40] W. Mason and S. Suri, “Conducting behavioral research on Amazon’s Mechanical Turk,” *Behav. Res. Methods*, vol. 44, no. 1, pp. 1–23, 2012.

[41] D. Centola, “The spread of behavior in an online social network experiment.,” *Science*, vol. 329, no. 5996, pp. 1194–7, Sep. 2010.

[42] D. Centola, “An experimental study of homophily in the adoption of health behavior.,” *Science*, vol. 334, no. 6060, pp. 1269–72, Dec. 2011.

[43] D. Centola and A. van de Rijt, “Choosing your network: Social preferences in an online health community.,” *Soc. Sci. Med.*, May 2014.

[44] R. Bapna, A. Gupta, S. Rice, and A. Sundararajan, “Trust, Reciprocity and the Strength of Social Ties: An Online Social Network based Field Experiment,” *Work. Pap.*, Jul. 2011.

[45] J. Wang, S. Suri, and D. J. Watts, “Cooperation and assortativity with dynamic partner updating.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 36, pp. 14363–8, Sep. 2012.

[46] G. Harrison and J. List, “Field experiments,” *J. Econ. Lit.*, vol. 42, no. 4, pp. 1009–1055, 2004.

[47] D. Edwards, *I’m feeling lucky: The confessions of Google employee number 59*, 1 edition. Houghton Mifflin Harcourt, 2011, p. 432.

[48] R. Kohavi, A. Deng, B. Frasca, R. Longbotham, T. Walker, and Y. Xu, “Trustworthy online controlled experiments,” in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’12*, 2012, p. 786.

[49] E. Bakshy, D. Eckles, and M. Bernstein, “Designing and deploying online field experiments,” in *Proceedings of the 23rd international conference on World wide web*, 2014, pp. 283–292.

[50] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, “Experimental evidence of massive-scale emotional contagion through social networks.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 24, pp. 8788–90, Jun. 2014.

[51] S. Harriman and J. Patel, “The ethics and editorial challenges of internet-based research,” *BMC Med.*, vol. 12, no. 1, p. 124, Jul. 2014.

[52] R. Bapna, J. Ramaprasad, G. Shmueli, and A. Umyarov, “One-Way Mirrors and Weak-Signaling in Online Dating: A Randomized Field Experiment,” *Work. Pap.*, 2013.

[53] E. Bakshy, D. Eckles, R. Yan, and I. Rosenn, “Social influence in social advertising: evidence from field experiments,” in *Proceedings of the 13th ACM Conference on Electronic Commerce*, 2012, pp. 146–161.

[54] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, “The role of social networks in information diffusion,” in *WWW ’12 Proceedings of the 21st international conference on World Wide Web*, 2012, pp. 519–528.

[55] S. Taylor, E. Bakshy, and S. Aral, “Selection Effects in Online Sharing: Consequences for Peer Adoption,” in *Proceedings of the fourteenth ACM conference on Electronic commerce.*, 2013, pp. 821–836.

[56] S. Taylor, L. Muchnik, and S. Aral, “Identity and Opinion: A Randomized Experimen,” *(Under Rev.*, 2014.

[57] S. Aral, L. Muchnik, and A. Sundararajan, “Engineering Social Contagions: Optimal Network Seeding and Incentive Strategies,” *Netw. Sci.*, vol. 1, no. 2, pp. 125–153, Feb. 2013.

[58] S. E. Asch, “Effects of group pressure upon the modification and distortion of judgment,” in *Groups, leadership and men*, Pittsburgh, PA: Carnegie Press, 1951, pp. 177–190.

[59] S. Milgram, L. Bickman, and L. Berkowitz, “Note on the drawing power of crowds of different size,” *J. Pers. Soc. Psychol.*, vol. 13, no. 2, pp. 79–82, 1969.

[60] M. Salganik and D. Watts, “Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market,” *Soc. Psychol. Q.*, 2008.

[61] L. Muchnik, S. Aral, and S. Taylor, “Social Influence Bias: A Randomized Experiment,” *Science (80-.).*, vol. 341, no. 6146, pp. 647–651, 2013.

[62] M. Godinho de Matos, P. Ferreira, M. D. Smith, and R. Telang, “Culling the Herding: Using Real World Randomized Experiments to Measure Social Bias with Known Costly Goods,” *SSRN Work. Pap.*, May 2014.

[63] R. Bapna and A. Umyarov, “Are Paid Subscriptions on Music Social Networks Contagious? A Randomized Field Experiment,” *SOBACO Work. Pap. Carlson Sch. Manag. Univ. Minnesota, Minneapolis, MN.*, 2012.

[64] S. Aral and D. Walker, “Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment,” *Manage. Sci.*, vol. 60, no. 6, pp. 1352–1370, Jun. 2014.

[65] C. F. Manski, “Identification of Endogenous Social Effects: The Reflection Problem,” *Rev. Econ. Stud.*, vol. 60, no. 3, pp. 531–542, 1993.

[66] S. Aral, L. Muchnik, and A. Sundararajan, “Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks,” *PNAS*, vol. 106, no. 51, pp. 21544–21549, 2009.

[67] C. R. Shalizi and A. C. Thomas, “Homophily and Contagion Are Generically Confounded in Observational Social Network Studies,” *Sociol. Methods Res.*, vol. 42, no. 2, pp. 211–239, 2011.

[68] D. R. Cox, *Planning of Experiments*. New York, New York, USA: Wiley-Interscience, 1958, p. 320.

[69] D. Rubin, “Bayesian inference for causal effects: The role of randomization,” *Ann. Stat.*, vol. 6, no. 1, pp. 34–58, 1978.

[70] C. F. Manski, “Identification of treatment response with social interactions,” *Econom. J.*, vol. 16, no. 1, pp. S1–S23, Feb. 2013.

[71] P. M. Aronow and C. Samii, “Estimating Average Causal Effects Under General Interference,” in *Summer Meeting of the Society for Political Methodology, University of North Carolina, Chapel Hill*, 2012, pp. 19–21.

[72] D. Eckles, B. Karrer, and J. Ugander, “Design and analysis of experiments in networks: Reducing bias from interference,” *Work. Pap.*, p. 29, 2014.

[73] A. Coppock and N. Sircar, “Design of Field Experiments under Unknown Interference Structures,” *Work. Pap.*, 2013.

[74] L. Katzir, E. Liberty, and O. Somekh, “Framework and algorithms for network bucket testing,” in *Proceedings of the 21st international conference on World Wide Web - WWW ’12*, 2012, p. 1029.

[75] J. Ugander, B. Karrer, L. Backstrom, and J. Kleinberg, “Graph Cluster Randomization: Network Exposure to Multiple Universes,” in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 329–337.

[76] E. Airoldi, P. Toulis, E. Kao, and D. B. Rubin, “Estimation of causal peer influence effects,” in *Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, JMLR: W&CP. Vol. 28*, 2013.

[77] A. C. Thomas and M. Finegold, “Protocols for Randomized Experiments to Identify Network Contagion,” *Work. Pap.*, p. 13, 2013.

[78] J. Nishimura and J. Ugander, “Restreaming graph partitioning: simple versatile algorithms for advanced balancing,” in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’13*, 2013, pp. 1106–1114.

[79] D. Eckles and E. Bakshy, “Bias and high-dimensional adjustment in observational studies of peer effects,” *Work. Pap.*, 2014.

[80] M. J. Salganik and D. D. Heckathorn, “Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling,” *Sociol. Methodol.*, vol. 34, no. 1, pp. 193–240, Dec. 2004.

[81] M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou, “A Walk in Facebook: Uniform Sampling of Users in Online Social Networks,” *Work. Pap.*, May 2009.

[82] D. Godes, D. Mayzlin, Y. Chen, S. Das, C. Dellarocas, B. Pfeiffer, B. Libai, S. Sen, M. Shi, and P. Verlegh, “The Firm’s Management of Social Interactions,” *Mark. Lett.*, vol. 16, no. 3/4, 2005.

[83] S. Aral and S. Taylor, “Viral Incentive Systems: A Randomized Field Experiment,” *Work. Pap. Stern Sch. Business, NYU*, 2014.

[84] G. Pickard, W. Pan, I. Rahwan, M. Cebrian, R. Crane, A. Madan, and A. Pentland, “Time-critical social mobilization.,” *Science*, vol. 334, no. 6055, pp. 509–12, Oct. 2011.

[85] E. Bakshy and D. Eckles, “Uncertainty in online experiments with dependent data: An evaluation of bootstrap methods,” in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1303–1311.