CHAPTER 15

....

NETWORKED EXPERIMENTS

....

SINAN ARAL

15.1 Embracing Interdependence

Networked experiments have been conducted since at least the 1950s, when Bavalas (1950) and Leavitt (1951) manipulated communication networks to measure their effects on group performance. However, trends toward digitization and fine-grained data collection have recently expanded the scale, scope, and complexity of networked experiments. Early experiments typically operated on hundreds of student subjects, assembled in laboratories. Today, however, we can design, conduct, and analyze networked experiments much more rapidly; and not just on samples of hundreds of people, but rather on hundreds of millions of people at a time. As more and more social interactions, behaviors, decisions, opinions, and transactions are digitized and mediated by online platforms, our ability to quickly answer nuanced causal questions about the role of social behavior in population-level outcomes such as health, voting, political mobilization, consumer demand, information sharing, product rating, and opinion aggregation is becoming unprecedented.

The importance of this new line of inquiry in the social sciences is difficult to overstate. How we relate to one another, connect with one another, coordinate, cooperate, and conflict with one another, are arguably the most fundamental data-generating processes in the social sciences. They are, in some sense, what make the social sciences *social*. In recent years, networked studies have grown to represent larger fractions of total research in traditional social sciences, like economics and sociology, as well as in other important disciplines, such as computer science and physics, which are increasingly addressing social science questions. All of these disciplines are embracing interdependence, which has become a key assumption in the most influential theoretical and empirical work in these areas.

As many scientific disciplines are simultaneously exploring the implications of interdependence in our social world, networked experiments, which help us understand our interdependence more deeply, are becoming one of the most important tools of social

science. The increasing scale and scope of modern networked experiments have created a new ability to engineer and randomize social settings to robustly estimate peer effects and the outcomes of social interactions, to explore the heterogeneity in these causal effects across subpopulations, and to unpack the nuanced behavioral mechanisms that underlie and explain these social effects. By understanding such causal behavioral mechanisms, how and why outcomes vary across individuals and how they change over time, we can develop more contextual, personalized and therefore more effective business and public policies. The ability to conduct large-scale networked experiments therefore portends a sea change in our scientific understanding of human relationships and behavior, and dramatic improvements in social policy as a result.

The precision and complexity of the experimental toolkit have also increased in recent years. As online platforms scaled to support hundreds of millions, even billions, of simultaneous users, and platform design became more open and precise, the ability to test complex dynamic hypotheses about social behavior expanded. For example, Facebook enables developers to customize application features for particular users, enabling feature and design randomization (e.g., Aral and Walker 2011a, 2012); Amazon Mechanical Turk enables the development of complex environments in which users can engage in precisely defined experimental micro-tasks (Mason and Watts 2012; Rand and Nowak 2011; Suri and Watts 2011); online labor markets and trading markets have been randomized to test market functions and market failure (Horton et al. 2010); and formal collaboration with platform developers and website administrators is enabling researchers to achieve even more comprehensive experimental control in large-scale in-vivo environments (Bakshy et al. 2012a,b; Muchnik, Aral, and Taylor 2013).

These digital tools are enabling a new era of experimental social science that has begun to reveal robust evidence of the nuanced causal determinants of human behavior. For example, recent large-scale digital experiments have revealed insights about product adoption and engagement (Aral and Walker 2011a, 2012; Bapna and Umyarov, forthcoming; Taylor, Bakshy, and Aral 2013), social commerce and advertising (Aral and Taylor 2011; Bakshy et al. 2012a; Tucker 2011), information sharing and diffusion (Bakshy et al. 2012b), herding behaviors in cultural markets (Muchnik et al. 2013; Salganik, Dodds, and Watts 2006; Tucker and Zhang 2011), health behaviors (Centola 2010, 2011), voting and political mobilization (Bond et al. 2012), performance in innovation contests (Boudreau and Lakhani 2011), the functioning of labor markets (Horton et al. 2010), coordination and cooperation (Fowler and Christakis 2010; Kearns et al. 2006; Mason and Watts 2012; Rand and Nowak 2011; Suri and Watts 2011), altruism and reciprocity (Bapna et al. 2011, Leider et al. 2009), and the growth and efficiency of two-sided matching markets (Tucker and Zhang 2010; Bapna et al. 2013). As a reference, I summarize the focus, context, and experimental procedures of recent large-scale networked experiments in Table 15.1. These studies have employed complex experimental designs that randomize social conditions at the system (e.g., Salganik et al. 2006), category (e.g., Tucker and Zhang 2010), item (e.g., Muchnik et al. 2013), group

Table 15.1 Examples of large-scale networked experiments

Study	Focus	Context	Experimental Procedure	Scale
		Product Diffusion and Engagement		
Aral and Walker (2014)	Impact of tie strength and embeddedness on social influence	Facebook app adoption	Randomize influence mediating messages at the dyad level	1.3M Users
Aral and Walker (2011a)	Viral features and social influence in product adoption	Facebook app adoption and use	Randomize viral features at the app level	1.5M Users
Aral and Walker (2012)	Social influence and susceptibility in product adoption	Facebook app adoption	Randomize influence mediating messages at the individual level	1.3M Experimental users; 12M observational users
Bapna and Umyarov (forthcoming)	Social influence in paid product adoption	Last.fm paid subscription adoption	Randomize gifted paid subscriptions at the individual level	40K Experimental users; 1.2M observational users
Taylor et al. (2013)	Downstream selection effects in online sharing behavior	Facebook offers	Randomize sharing mechanisms at platform level	1.2M Users
Hinz et al. (2011)	Seeding strategies to promote diffusion	Redeemable tokens; URLs to a viral video	Randomized initial seed recipients	120 Users/28 experiments; 1,380 students
		Social Commerce and Advertising		
Bakshy et al. (2012a)	Social advertising	Facebook social advertising	Randomize social signals in advertising	23M Users; 148K ads; 101M user–ad pairs
Tucker (2011)	Social advertising	Unnamed nonprofit	Randomize social ad text and targeting	630 Ads; 13K impressions
Aral and Taylor (2011)	Incentivized peer referral marketing	Online floral delivery site	Randomize incentive structure: selfish, generous, and fair	637 Users

Table 15.1 (*Continued*)

Study	Focus	Context	Experimental Procedure	Scale
		Information Sharing and Diffusion		
Bakshy et al. (2012b)	Social influence in information diffusion	Facebook news feed	Randomize influence mediating messages at the individual level	253M Users; 76M URLs; 1.2B user–URL pairs
		Herd Behaviors in Ratings and Cultural Markets		
Salganik et al. (2006)	Herd behavior in cultural markets	Artificial music lab	Randomize popularity information	14K Users
Tucker and Zhang (2011)	Popularity information and choice	Wedding service vendor website	Randomize availability of popularity information	3 Vendor categories; 90K click-throughs
Muchnik et al. (2013)	Social influence bias in online ratings	Ratings in an unnamed social news aggregation website	Randomize first rating at the comment level	116K Users; 101K comments; 10M user–comment impression pairs
Taylor et al. (2014)	Impact of identity on rating social content	Ratings in an unnamed social news aggregation website	Randomize availability of identity information	350K Comments; 17M impressions; 875K commenter–viewer pairs
		Health Behaviors		
Centola (2010)	Social influence in the spread of health behaviors	Artificial Health network	Randomize alters/ Network structure	1500 Users
Centola (2011)	Homophily and influence in the spread of health behaviors	Artificial health network	Randomize alters / network structure	700 Users

		Voting and Political Mobilization		
Bond et al. (2012)	Voting	Facebook voter registration campaign	Randomize social signals in voting	61M Users
		Cooperation and Coordination		
Kearns et al. (2006)	Graph coloring	Laboratory	Randomize network topology	2 Experiments; 55 users
Fowler and Christakis (2010)	Social influence in cooperation	Laboratory experiments	Random partnering in cooperation games	240 Users
Rand and Nowak (2011)	Cooperation games	Amazon Mechanical Turk Laboratory	Randomize fixed and fluid network structure	785 Users; 40 sessions
Suri and Watts (2011)	Public goods game	Amazon Mechanical Turk Laboratory	Randomize network topology	113 Experiments; 24 users/experiment
Mason and Watts (2012)	Cooperation and exploration games	Amazon Mechanical Turk Laboratory	Randomize network topology	256 Experiments; 16 users/experiment
		Reciprocity and Altruism		
Leider et al. (2009)	Altruism, directed altruism	Facebook dictator and helping games	Randomize anonymity and repeated interaction	802 Students; 2,360 students
Bapna et al. (2011)	Reciprocity games	Facebook game	Randomize anonymity	190 Users; 77 games

Table 15.1 (*Continued*)

Study	Focus	Context	Experimental Procedure	Scale
		Innovation Contest Performance		
Boudreau and Lakhani (2011)	Innovation contests	NASA TopCoder contest	Randomize incentives and sorting on competitive regime	Multiple settings
		Labor Market Outcomes		
Horton et al. (2010)	Labor markets	ODesk online labor market	Three different randomizations	Multiple settings
		Two-Sided Markets and Matching		
Tucker and Zhang (2010)	Growing two-sided markets	B2B exchange market	Randomize display of number of buyers and/or sellers	15 Categories; 3,314 listings
Bapna et al. (2012)	Effect of anonymous weak signaling on date matching	Dating website	Randomize anonymous weak signaling feature	10K Experimental users; 100K observational users

(e.g., Suri and Watts 2011), and individual (e.g., Aral and Walker 2012) levels to reveal important insights about human behavior at population scale.

The scale of modern-day experimentation also enables new levels of analysis. Smaller-scale randomized experiments are typically only sufficiently powered to estimate average treatment effects—the average effect of a policy in a population. But experimental tests of micro-level policies among millions of people allow researchers to unpack the heterogeneity of treatment effects across different populations and to explore the behavioral mechanisms that explain the treatment effects (Aral 2011). For example, the experiment conducted by Aral and Walker (2012) estimates heterogeneity in the impact of influence mediating messages on different types of consumers; Bapna and Umyarov (forthcoming) explore heterogeneity of influence across the degree distribution of users, and Muchnik et al. (2013) dig deeply into whether opinion change or selective turnout creates the social influence bias they estimate to exist in online ratings. These insights enable policies tailored for particular groups of people. Furthermore, creating a deeper understanding of the data-generating processes that explain the average behavioral effects of policy interventions will prepare policy-makers to respond to changes in these processes and to know how their interventions create dynamic changes in behavior over time.

Yet while the increasing scale, scope, and complexity of networked experiments creates tangible opportunities for scientific advancement, they also simultaneously create significant new challenges. As networked experiments become larger and more complex, the likelihood of nontrivial statistical interference, generating non-ignorable violations of the stable unit treatment value assumption (SUTVA) (Rubin 1986, 1990), must be explicitly considered. Although digital technologies enable us to observe more micro-level human interaction, they may also obscure new constraints on sampling, reducing the marginal cost of reaching individuals online while increasing the chances that we will miss large swaths of society that interact offline. Meanwhile, the digital platforms across which we observe micro-level behaviors online are constantly changing, as the algorithms that drive human behavior are altered week to week and month to month. Such changes, and vast behavioral dependencies on the rules of engagement embedded in platform technologies, must be considered as we generalize the findings of networked experiments conducted on digital substrates to the broader human condition.

The purpose of this chapter is to present a concise and accessible review of advances in modern networked experimentation and to summarize the opportunities and challenges they create for understanding our social world. The chapter is divided into two sections that discuss the *design* and *analysis* of networked experiments, respectively. I review innovations in *networked experimental design* in Section 15.2, assessing the implications of the experimental setting, sampling, randomization procedures, and treatment assignment. I then discuss the *analysis of networked experiments* in Section 15.3, with particular emphasis on modeling treatment response assumptions, inference, and estimation, including recent approaches to interference and uncertainty in dependent data. I conclude by discussing important challenges facing researchers as

```
        Design                                    Analysis

Setting  —  Sampling  —  Randomization  —  Assignment  —  Modeling  —  Inference  —  Estimation
– Context   – Random Graphs   – Peer Encouragement   – Network      – Assumptions   – Estimands   – Specification
– Network   – Node & Edge     – Mechanism            Autocorrelation – Models        – Data        – Interference
Dimensions  – Snowball         – Structural           – Sequencing                                  – Estimators
            – Random Walk      – Setting
            – Forest Fire
            – Respondent Driven

                          Sequence of Key Decisions
```
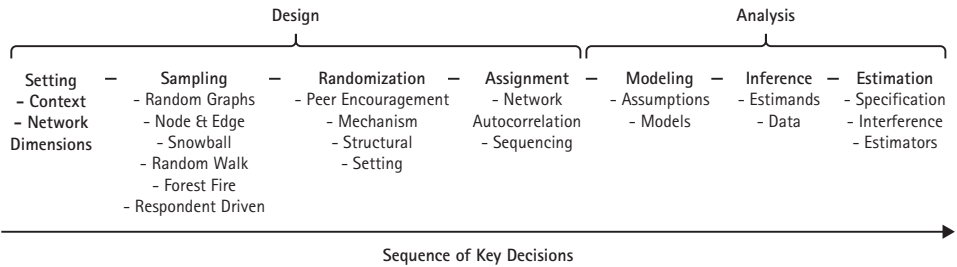
FIGURE 15.1 Key decisions in the design and analysis of networked experiments. The figure represents key decisions the researcher must make during the design and analysis of networked experiments. Though the figure presents these decisions sequentially, the process is, in a sense, iterative. Problems encountered during estimation (for example, interference), necessitate changes in one's approach to sampling, randomization, and modeling. Though there are feedback loops that connect the key decisions, this sequential representation helps frame the problems in a logical order.

we develop the future of networked experimentation, focusing on adaptive treatment assignment, novel randomization techniques, linking online treatments to offline responses and experimental validation of observational studies.

Figure 15.1 presents the sequence of key decisions in the design and analysis of networked experiments considered in Sections 15.2 and 15.3. In order to frame the discussion clearly, I use a particular paper, Aral and Walker (2011), to instantiate this sequence of key decisions. The choices made in Aral and Walker serve to concretely highlight the opportunities and challenges of networked experimentation. I have chosen this paper, to which readers may wish to refer directly while reading this chapter, not because it is the best example of networked experimentation (in fact it is an early example of large-scale, in vivo networked experimentation), but because I am personally aware of the opportunities and challenges encountered during the work.

The chapter provides a broad overview of networked experimental techniques, addressing key methodological concerns. It is intended as a review of the literature and a guide for researchers and practitioners interested in conducting experiments in networks. Given this aim, the chapter complements deeper treatments of narrower subtopics presented in this handbook in the chapters on laboratory experiments by Syngjoo Choi and Shachar Kariv, and on field experiments in developing countries by Emily Breza. The contribution of networked experimentation to our understanding of human social dynamics is likely to be large. I hope this overview, combined with the excellent contributions of my colleagues highlighted herein, provides a roadmap for extending networked experimentation in meaningful ways in the coming years.[1]

---

[1] This introductory section of this chapter relies heavily on a review of the literature I wrote with Dylan Walker in Aral and Walker (2014). Here I have expanded on and updated that review. Portions of the text as well as an earlier version of Figure 15.1 are reprinted by permission from Sinan Aral and Dylan Walker

15.2 DESIGN OF NETWORKED EXPERIMENTS

15.2.1 Setting

The settings in which networked experiments are conducted typically enable and constrain the inquiries researchers can pursue. One useful categorization divides experiments along two dimensions: the context in which the experiment takes place (in the field or in the lab) and the dimensions of networks that are examined (network structure, dyadic relationships, or individuals). Although in some sense network structure is always present, the focus of the inquiry, and thus the experimental manipulation, change simultaneously across these dimensions.

Context. Networked experiments typically take place either in a laboratory or in the field. The lab offers more experimental control but perhaps less realism and thus generalizability (depending on the research questions being asked), while the field offers less control but a more realistic setting in which to observe human behavior. There are also experiments that take place in "virtual labs," for instance in controlled laboratory environments created by researchers on the web (e.g., Salganik et al. 2006; Centola 2010, 2011; Mason and Suri 2012), as well as in "living labs," which are instrumented physical locations like domiciles, dwellings or communities in which inhabitants are digitally instrumented (e.g., Aharony et al. 2011), and in virtual worlds (Bainbridge 2007; Fiedler et al. 2011). Virtual labs, living labs, and virtual worlds have their own idiosyncrasies with regard to data collection, recruitment, sampling and so on, which are discussed in detail in the pioneering papers in these areas referenced here.

The trade-off between control and realism exists across lab settings and field settings even when the labs are virtual or the field is digitized. In a lab, the researcher can control many aspects of the environment that are difficult to control in the field. For example, complex network structure can be manipulated to study effects on information diffusion (Centola 2010), cooperation (Suri and Watts 2011) or coordination (Kearns et al. 2006). Participants are, in these cases, randomly inserted into precise, well-controlled and well-understood network structures to study how structure affects behavior. The information participants have access to can also be controlled, such that the interactions between complex network structures and information can be examined. Field studies sacrifice a measure of control for observation of more realistic networked behaviors. Subjects are observed and studied in their natural environment and through their typical modes of interaction with their real network connections. Field experiments allow us to study how real-life relationships affect our judgment and how true social influence affects our behavior or opinions. The appropriate setting depends on the

research question one is after: to understand how idealized behaviors change in response to complex changes in network structure, lab settings are quite powerful; to understand how behaviors evolve in real social structures, field experiments are perhaps more appropriate.

Aral and Walker (2011) conducted a field experiment of interactions and behaviors taking place on the live Facebook social network, to measure the effects of design features on product virality and to investigate the impact of network externalities on engagement and the likelihood of customer churn. The paper randomly enables viral product features, such as social notifications and the ability to invite friends, to test their effects on product diffusion, and measure how network externalities drive product use after adoption. I will return to this example periodically to illustrate various dimensions of networked experimentation.

Network Dimensions. Inquiry can focus on the role of structure, the functioning of dyadic relationships or the implications of changes in setting or technology on the networked social behaviors of individuals or groups. Choices made among these network dimensions frame experimental studies and the inferences and generalizations that can be drawn from them. Such choices also have implications for downstream choices regarding network sampling, randomization procedures, treatment assignment, modeling, measurement and inference. Analyses of the role of structure require sampling a set of individuals representative of the population. Analyses of dyadic peer influence may require samples of representative ego networks from larger graphs that maximize statistical precision and minimize interference between units. Analyses of changes in technology; for instance, the degree to which network externalities exist for a given technology, may require comprehensive samples of larger groups of interacting contacts to analyze the user experience under realistic network externality conditions, such as uniform access to the technology.[2]

For example, Aral and Walker (2011) are interested in estimating (i) dyadic peer influence conveyed through different interpersonal mechanisms (in their case, differences in peer influence conveyed by one to one personal invitations and one to many broadcast notifications), (ii) the overall impact of viral features on social contagion in product adoption processes, and (iii) the effect of network externalities on the persistence of behaviors inspired by social influence over time. Thus the work contemplates both dyadic and network-level dimensions of social contagion, which affect subsequent sampling, randomization, inference, and estimation decisions.[3]

---

[2] For a more comprehensive treatment of empirical investigations of network effects, I refer the reader to the excellent chapter on "Recent Developments in the Empirics of the Effects of Networks" by Bernard Fortin and Vincent Boucher in this handbook.

[3] It is important to note here that a networked experiment may also be interested in subsequent dynamic changes in network structure itself, a topic I do not cover in this chapter due to the dearth of prior research in this area. Although models of network dynamics consider link formation and disillusion, few, if any, experimental studies examine determinants of network evolution (for a notable exception, I refer the reader to work by Sharique Hasan, who has exploited natural experiments to study network evolution [Hasan and Bagde, forthcoming] and is building a living laboratory in India

15.2.2 Sampling

Given a setting, the network on which the experiment will be conducted must be generated or sampled from a graph or graph distribution. In cases where the network can be controlled, one may wish to initialize particular network structures based on theory. For example, a researcher may be interested in comparing coordination outcomes in cohesive or diverse networks (Suri and Watts 2011; Shore et al. 2014), or the propagation of simple or complex contagions across weak ties or wide bridges (Centola 2010). Alternatively, the researcher may generate synthetic graphs that display certain properties, such as power law degree distributions, particular path lengths, or clustering characteristics, using a model of network formation. Random graph models, which provide a family of network generation models that produce graphs with precisely solvable average properties (Newman 2002), trace their origins to Erdos and Reyni (1959, 1960, 1961), whose random graphs underpin today's modern graph generation procedures. Unfortunately, the original models also produced some unrealistic social network properties, like Poisson degree distributions and a lack of clustering. More modern generalized random graph models incorporate arbitrary non-Poisson degree distributions, clustering and more realistic power-law degree distributions (e.g., Aiello et al. 2001; Newman 2002).

In field experiments, naturally occurring network structure is typically taken as given, though not necessarily fixed, and randomization occurs through, for example, the perturbation of behaviors or information channels across the observed network. In this setting, individuals are sampled together with their peers from real graphs to study how changes in behavior affect social outcomes in the naturally occurring network. The goals of sampling can vary from estimating node or edge attributes, collecting subgraphs that are representative of local ego networks, or collecting a sample of network paths that reliably reproduces likely diffusion events. Several sampling methodologies can produce the network on which the experiment is conducted. For example, *node sampling* and *edge sampling*, which choose nodes and edges independently and uniformly at random from the population network, typically do not preserve graph properties, such as the degree distribution, density, connectivity, and clustering of the network (Stumpf et al. 2005; Lee et al. 2006).

Topology-based sampling is better equipped to preserve representative graph properties and structures. For example, *complete snowball sampling* (e.g., Aral et al. 2009), which uses breadth-first search from randomly selected seed nodes, can accurately preserve network properties within the snowball, but suffers from boundary bias as nodes at the edge of the snowball sample will be missing neighbors (Lee et al. 2006). Snowball sampling can also miss the connectivity of the global graph because, although clusters of representative subgraphs around random seed nodes are complete, the

to study network evolution through experimental means in a startup accelerator). Experimental studies of network evolution represent a fertile area for future research, one in which important insights are likely to be found.

connections between such subgraphs are typically under-sampled. Complete snowball sampling therefore best suits studies of micro-level social behavior (e.g., studies of average local peer effects or micro-level, interpersonal diffusion dynamics) rather than global diffusion or cascading behavior.

Random walk sampling, which begins by selecting a seed node uniformly at random and simulating a random walk along the network, more accurately captures global graph properties and community structure, but makes micro-level, dyadic social processes difficult to characterize. *Random jump sampling* mirrors the random walk, but jumps to any random node instead of returning to the seed node and therefore suffers from a lack of reliable reproduction of symmetric and/or complete local network structures. *Forest fire sampling* (Leskovec, Kleinberg, and Faloutsos 2005) conducts a partial-breadth first search, which samples some proportion of the edges of a randomly sampled "ambassador" node (and its corresponding dyadic node counterparts) with some probability, continuing the process until the "fire" burns itself out. These walk-based sampling methods replicate global graph properties and community structure more accurately than snowball sampling (because they are not constrained to local networks), but they do not sample complete local networks around seed nodes. *Respondent-driven sampling* (Heckathorne 1997; Salganik and Heckathorne 2004) relies on respondents to recruit their peers and a model of the recruitment process to weight the sample to compensate for nonrandom recruitment patterns. While such methods are well suited to recovering statistically valid samples of difficult-to-reach groups (e.g., IV drug users, homeless youth), they can also produce biased samples of local structure.

These sampling techniques make realistic testing of local social interactions more difficult to generalize because complete social environments are not observed. Structural determinants of social behavior are also difficult to analyze. For example, the debate about the roles and relative importance of wide bridges (Centola and Macy 2007) or thick bridges (Aral and Van Alstyne 2011) for the spread of complex contagions becomes more difficult to resolve in the absence of complete local structure. The point of this discussion is not to provide a comprehensive treatment of network sampling (e.g., Leskovec, Kleinberg, and Faloutsos 2005; Leskovec and Faloutsos 2006; Ahmed et al. 2014). Rather, the aim is to point out that sampling or graph generating procedures, which produce the graphs on which experiments are conducted, can enable and constrain the inferences that are possible as well as the degree to which those inferences are generalizable to larger populations or sets of graph structures.

In Aral and Walker (2011), we aimed to generalize our findings on social influence to users of Facebook and online social interactions more broadly. We therefore went to great lengths to ensure our experimental sample was representative of the Facebook population. We employed an ad-targeting agency to target advertisements about the application on which we conducted the experiment such that the adopting population was representative of the Facebook population. Distributions of individual characteristics of the resultant sample were then compared to published statistics about the overall Facebook population to ensure the sample was representative. After

approximately 10,000 initial adopters were recruited, a complete one-step snowball sample was collected using the Facebook API to create a population of approximately 1.4 million friends of initial adopters. All edges originating from the initial adopters as well as ties between friends of initial adopters were sampled. In addition, any diffusion event was captured comprehensively, meaning if an invitation was sent from an initial adopter to a friend, who then adopted the application and invited a friend of a friend to adopt, and so on, every adoption in the chain prompted data collection about the adopter, their networks, and their characteristics to ensure complete observation of all diffusion events regardless of their depth. This example links the sampling strategy to the goals of the experiment. The representative sample of initial adopters allowed us to measure the impact of including viral invitation and broadcast notification features on the application's diffusion process, the dyadic peer influence conveyed through different interpersonal mechanisms and the effect of network externalities on the prolonged use of the application for a representative sample of Facebook users. Experimental results based on this sampling strategy can therefore be generalized to Facebook users as a whole.

15.2.3 Randomization

In some sense, the core of networked experimental design is the randomization procedure. Randomization creates the exogenous variation that identifies the causal effects of interest. If we aim to understand the impact of network structure on social behavior, exogenous variation should be introduced at the network level. If instead we aim to understand peer effects, exogenous variation should be introduced at the individual level. At least four categories of randomization procedures, aimed at different research questions, have gained traction in the last several years.

Peer encouragement designs randomly encourage particular behaviors in a set of nodes to analyze the effects of those behaviors on the nodes' peers. They enable studies of peer effects or social influence in a variety of settings and across a number of behaviors. For example, Bapna and Umyarov (forthcoming) gift premium subscriptions to users of a social music streaming website to estimate peer influence in the decision to adopt a premium service. Hinz et al. (2011) randomize initial seed recipients to test the impact of seeding strategies on the success of viral marketing campaigns in three different contexts. As with traditional encouragement designs, peer encouragement allows researchers to study social effects emanating from the adoption of products, services, or behaviors whose adoption cannot be guaranteed by policy instruments in real world settings.

Mechanisms designs change the mechanisms through which individuals interact. The focus could be on how different communication channels transmit information or the way behaviors spread in social networks through peer influence. For example, Bakshy et al. (2012b) randomize exposure to signals about friends' information sharing on Facebook. They find those who are exposed are significantly more likely to spread

information. Aral and Walker (2011) randomize the channels of communication through which users can promote applications on Facebook and find active personalized messaging is more effective per message and generates more engagement, while passive broadcasts generate faster growth through more messages. Aral and Walker (2012, 2014) randomize the delivery of broadcast notifications to peers of selected Facebook users to estimate individual and dyadic correlates of social influence. Their randomization, at the ego-network level, also addresses known degree and homophily biases in independent random assignment designs (Thomas and Finegold 2013), a topic I cover in more detail in Section 15.3.

Structural designs randomize network structure, or individuals' positions within it, to understand the impact of structure on social behavior. For example, Kearns et al. (2006) study how network structure affects individuals coordinating to solve the graph coloring problem. By randomizing structure, they find that networks generated by preferential attachment make solving the coloring problem more difficult than do networks based on cyclical structures or "small worlds." Centola (2010) studies how network structure affects the spread of health behaviors in an online social network. By randomizing the degree of local clustering, he finds that diffusion is more likely across clustered-lattice networks than random networks. Mason and Watts (2012), Suri and Watts (2012), and Rand and Nowak (2011) all randomize network topology to study cooperation and public goods games played by Amazon Mechanical Turk users.

Setting designs change the settings in which social behavior takes place. Particular behaviors are not encouraged and interaction mechanisms are not necessarily altered. Instead, the environment in which the interaction takes place is altered and studied. For example, Aral and Taylor (2011) randomize the incentive structure in which social behaviors take place, to examine how incentives affect people's social behavior. Bapna et al. (forthcoming) and Taylor et al. (2014) randomize the presence of identifying information to study the effect of anonymity on social behavior in the context of online dating and online ratings respectively. Muchnik et al. (2013) study how the presence of popularity information affects the ratings of friends and enemies on each others' social content. Table 15.2 summarizes these four types of randomization procedure and their applications and gives examples of each.

Aral and Walker (2011) implemented a mechanism design experiment that randomly enabled and disabled certain features on a Facebook application to measure their impact on viral diffusion. As people downloaded the application, they are randomly assigned to one of three treatment groups: baseline, passive-broadcast, and active-personalized. Users assigned to the baseline treatment group received a version of the application in which both notifications and invitations were disabled. In the passive-broadcast treatment group (passive), only notifications were enabled. In the active-personalized treatment group (active), both notifications and invitations were enabled. There were no other differences between the baseline, passive, and active applications. In this sense, the randomization takes place at the user level and the environmental condition that is randomized is the state of the viral features that are enabled or disabled on the application.

Table 15.2 Common Networked Randomization Procedures

Design	Procedure	Application	Examples
Peer encouragement designs	Randomize encouragement of behavior in focal node; measurement of response in peers	Measurement of peer effects, social influence, or social contagion	Randomly encourage product adoption (Bapna and Umayurov [forthcoming]); randomly encourage thankful status updates on facebook (Eckles)
Mechanism designs	Randomize characteristics of channels of interaction between nodes; measure responses of ego and peers	Measurement of impact of interaction channels on diffusion, contagion, cooperation	Randomize viral features of a product (Aral and Walker 2011); randomize message delivery (Bakshy et al. 2012; Aral and Walker 2012)
Structural designs	Randomize properties of network structure; measure responses of ego and peers	Measurement of impact of network structure on diffusion, contagion, cooperation, team performance	Randomize network structure in behavior diffusion (Centola 2010); in cooperation games (Suri and Watts; Rand et al.)
Setting designs	Randomize environment in which social behavior takes place (e.g., availability of contextual information or incentive structures); measure responses of ego and peers	Measurement of social environment effects on interactions, transactions, diffusion, cooperation	Randomize incentives to propagate a product (Aral and Taylor 2011); randomize incentives and sorting (Boudreau and Lakhani 2011)

15.2.4 Treatment Assignment

Treatment assignments map nodes and/or edges to treatment conditions. Depending on the study design, assignments may be independent or correlated across nodes and edges in the network. Typical experimental treatment assignments independently assign units to treatment and control conditions. While this is certainly possible in network settings, treatment assignment with network autocorrelation may produce some desirable properties of the units chosen. Furthermore, in considering network autocorrelated assignments in which the assignment of some units is a function of the assignment of others, the sequencing of treatment assignments becomes critical.

Network Autocorrelation. Network autocorrelation in treatment assignment may be used to study group interventions; for example, in studies of inducement to peer pressure among sets of connected nodes (Mani et al. 2013) or in studies of the adoption of products and services that exhibit network externalities (Ugander et al. 2013). In the latter case, realistic parameter estimates of network effects may require assignment of entire network neighborhoods to encouragements to adopt a product because, in a world with universal access to the product, network externalities operate on clustered adoption with all peers of an adopter able to adopt the product. Blocked treatment assignment is also essential to design-based approaches to statistical interference (e.g., Ugander et al. 2013; Eckles et al. 2014), which I discuss in detail in Section 15.3. In these designs, clusters of treated nodes are optimally "surrounded" by peers with the same treatment assignment to reduce bias and error in estimation. In peer encouragement designs it may be useful to treat a set of seed nodes from a network and to block treatment of peers of those nodes (Airoldi et al. 2013). Such a setup can facilitate measurement of peer effects while minimizing interference. This example highlights another key design choice: the temporal sequencing of treatments.

Sequencing. The sequencing of treatment assignments is a critical aspect of randomization procedures that enable and constrain inference. Sequencing dilemmas are highlighted by viral marketing experiments, which utilize response driven sampling.[4] As networked experiments typically study cascading behaviors, products, or services, how to treat units sampled through the spread of social behavior is a fundamental decision. For example, Aral and Walker (2011) experimented with multiple "viral product design" strategies to estimate their impacts on peer to peer sharing of an online application. They randomly assigned subjects to different versions of the application with different viral messaging features and measured referrals as the different versions diffused through the Facebook network. An important choice in this context is how to assign sequentially referred users to treatments. If, for example, a user with version A refers a second user, should the referee be assigned to treatment A or

---

[4] While respondent-driven sampling asks respondents to name new respondents, I use the term response-driven sampling here to refer to new units sampled as a consequence of responding to a peer behavior induced by the experiment; for example, adopting a product through a peer referral designed by the experimenter.

randomly assigned to either A or B? As diffusion dynamics may depend on the consistency of peer experiences (e.g., due to network effects or consumer expectations), consistent treatment assignment—assigning all referees to the same treatment status as their referrers—enables estimation of the global treatment impact of implementing version A. Estimands can reflect what the world would look like if everyone received version A. However, consistent treatment assignment reduces the power of the experiment in estimating differences in micro-level peer effects—the effects of A or B on sharing—because fewer units are randomly assigned to treatment. Aral and Walker (2011) randomly assign response-driven samples, rather than maintaining consistent treatment assignment throughout the referral process, and discuss this trade-off in experimental design at some length. In essence, the decision maps directly to the goal of the research. As the goal of the research is to robustly estimate micro-level peer influence effects rather than global diffusion dynamics, random assignment of referred peers increases the statistical power of such estimation.

Sequential treatment assignment also plays a significant role in downstream inference. As I discuss in more detail in Section 15.3, two-stage clustered treatment assignment strategies can reduce selection bias due to latent homophily and the chances of selecting high-degree nodes in the second stage (Thomas and Finegold 2013). As peers of randomly selected users are likely to be of higher degree and similar in characteristics to their randomly selected peers, randomizing treatment assignment in the second stage can help alleviate these biases (Aral and Walker 2012). Sequential treatment assignment can also address other important statistical concerns, such as interference, that arise in interdependent data. As Airoldi et al. (2013) propose, sequential assignment can help alleviate such concerns by making treatment assignments a function of the previous treatment assignment of peers.

Finally, as we look toward the future of networked experimentation, adaptive sequential treatment assignment, in which treatment assignments are a function of prior treatment responses, could help us maximize precision, minimize interference, cost, and harm, or investigate treatment effect heterogeneity or particular hypotheses. For example, if the result of a particular hypothesis is becoming obvious, fewer units could be allocated toward testing that hypothesis as the experiment progresses. I discuss the potential of adaptive treatment assignment in Section 15.3.

15.3 ANALYSIS OF NETWORKED EXPERIMENTS

15.3.1 Modeling Treatment Response

The potential outcomes approach, formally the Rubin Causal Model or the Neyman-Rubin Causal Model, attempts to tackle the fundamental problem of causal inference: We wish to estimate differences in units' outcomes under different treatments, but are never able

to observe a given unit under multiple treatments at the same time (Neyman 1923; Rubin 1974). As only one potential outcome can be observed for any individual, the differences between potential outcomes (causal effects) are impossible to observe, but they can be inferred under some basic assumptions, which I discuss in more detail in Section 15.3.3. A key assumption of the potential outcomes approach is that there is no interference between units, meaning "the observation on one unit [is] unaffected by the particular assignment of treatments to the other units." (Cox 1958: 19) If this stable unit treatment value assumption (SUTVA) (Rubin 1986, 1990) does not hold, then rather than having two potential outcomes under treatment and control, experimental units have many potential outcomes that depend on other units' treatments (Rosenbaum 2007).

Analysis of networked experiments therefore begins with a theoretical approach to treatment responses and assumptions about their interdependence.[5] There exists some data-generating process that translates treatment assignments and network structure into the observed outcomes of networked nodes. The assumptions we make about this data-generating process—whether it is assumed to be known or unknown, or in equilibrium or not—will guide our inference and our interpretations of experimental results.

For example, Manski (2013) provides precise theoretical assumptions about treatment response that vary the degree to which a unit's response depends on other units' treatment assignments. Manski's (2013) constant treatment response assumption considers a function that maps treatment assignment vectors to effective treatments such that a unit's response depends on its own assignment as well as the assignments of other units. If node responses are assumed to depend only on their own assignment in what Manski (2013) calls an *individualistic treatment response* assumption, the function simply maps units' own assignments to their outcomes. In this case, there is no interference and SUTVA holds. If, on the other hand, node responses depend on their own assignment and the assignments of a reference group with which they interact, then the function maps sub-vectors of treatments in the reference group to node outcomes. One example of such an assumption is the *neighborhood treatment response* assumption of Aronow and Samii (2012), which posits that effective treatments depend on the unit's assignment and the assignment of their immediate network neighbors. However, other assumptions about effective treatments are also possible. For example, when estimating the effectiveness of child vaccinations, a plausible assumption is that a child's illness is a function of their own vaccination status and the number or proportion of children vaccinated in their school, but not on the specific identities of those vaccinated. Manski (2013) refers to this as an *anonymous treatment response* assumption.

The "exposure mapping" Aronow and Samii (2012) present in their simulation of Add Health data models units' responses as a function of their own treatment and

---

[5] I return to a detailed treatment of recent approaches to interference in networked experiments in Section 15.3.3.

the number of their treated peers. Treatment responses may not only be a function of peer treatment assignments, but also of peer behavior. As Eckles et al. (2014) point out, under such assumptions, peer behavior may fully mediate the effects of peers' treatment assignments. Rather than comprehensively reviewing various treatment response assumptions, I present these as examples of the types of assumptions one can make. We can then use these assumptions to inform our models of treatment response, which we use to estimate and make inferences about networked behavior.

One of the most widely used models of treatment response is the *linear-in-means model*, which specifies a linear additive contribution of treatment assignment and peer effects to the mean treatment response, such that behavior is a function of a direct effect of treatment and the mean of peers' behaviors (which may mediate peers' treatment responses):

$$Y_i = \alpha + \beta X_i + \gamma \, \frac{\sum_{j=1,j=i}^{k} Y_j}{k_i} + \varepsilon_i \qquad (15.1)$$

where β represents the direct effect of a unit's treatment assignment (Xi), and γ the indirect effects of the mean of peer behaviors (where j indexes i's peers and ki represents i's degree). A similar model for binary outcomes is described in Brock and Durlauf (2001). The linear-in-means model has become a staple in the estimation of peer effects, in part because of its simplicity, generality, and applicability to a wide variety of social effects of interest in economics and sociology. Identification of the linear-in-means model has received a good deal of attention in the economics literature (e.g., Manski 1993; Graham and Hahn 2005; Lee 2007; Bramoulle et al. 2009) and has been used to estimate peer effects in a variety of settings. In more recent work, Dieye et al. (2014) propose a structural extension to the linear-in-means model that accounts for social network effects in treatment response and allows for a contextual effect of treatment alongside the endogenous peer effect.

However, the linear-in-means model is not without its limitations. For instance, in [1], γ captures the slope of the average peer effect of i's neighbors, ignoring the possibility of dyadic peer effects that vary with the characteristics of i and j. Such a model is not well suited to situations in which peer effects are not simply linear in means. As Burke and Sass (2013) show for educational outcomes, average peer effects may be small while peer effects conditioned on the individual, the peer, or the dyad may be quite large. If high-ability students are influenced positively by other high-ability students but negatively by low-ability students, then a linear-in-means model will miss important heterogeneity in the data. When linear-in-means models of social effects in educational outcomes are compared to models with interactions by own achievement, large peer effects are detected by the conditional model but missed by the linear-in-means model (Burke and Sass 2013).[6] Similar extensions to the linear-in-means approach have been used in the study of peer influence in product adoption (Aral and Walker 2012), premium subscriptions (Bapna and Umyarov, forthcoming), online ratings (Muchnik

[6] Thanks to Michaela Kerrissey for referring me to Burke and Sass (2013).

et al. 2013), and in the role of identity in opinion formation (Taylor et al. 2014). It is also plausible that something other than the second moment of the distribution of peer behaviors is relevant to a unit's outcome. For example, there may be threshold effects (Granovetter 1978; Watts 2002), quadratic reinforcement effects (Centola and Macy 2007), effects from the variance of peer behaviors, or other more complex dependencies.

15.3.2 Inference

Estimands. In randomized experiments, the randomization distribution generated by the experimental design forms the basis of inference: one infers individuals' outcomes under treatments they were randomly denied from observations of individuals' outcomes under treatments they were randomly assigned. For simplicity I denote the randomization distribution of treatments under no intervention as η and under an arbitrary intervention as θ. Since any individual cannot be observed under multiple treatments, causal estimands are typically defined as averages of potential outcomes (Rosenbaum 2007; Hudgens and Halloran 2008).

Many different quantities of interest form the basis of relevant estimands in networked experiments. We may seek to understand individual-, group-, or population-level interventions and outcomes. In addition, we may be interested in experimental effects on the treated, peers of the treated, or the groups or populations in which treatment is studied. The estimands implied by these research goals vary and are being extended and re-conceptualized in some of the most recent literature. Precisely defining the experimental quantities of interest can help researchers draw bounded generalizations from experimental results. A large body of literature addresses the decomposition of treatment effects into direct, indirect and total effects across the units in an experiment (Halloran and Struchiner 1991, 1995). Using this decomposition, Hudgens and Halloran (2008) define several estimands useful for conceptualizing the quantities of interest in experimental studies with interdependence across groups. I describe how these estimands can be related to quantities of interest in networked experiments and illustrate their application in a peer encouragement design example. For simplicity, consider a basic peer encouragement design where treated nodes are given an encouragement to adopt a product and the researcher is interested in effects of this encouragement on product adoption outcomes (Y). In this context we can think about direct, indirect, and total effects as follows:

Direct causal effects, the effect of the treatment on the treated, can be formalized as the difference in individual average potential outcomes under treatment and no treatment conditions given the randomization distribution θ (i.e., in the case where an intervention is initiated): $\bar{D}_i(\theta) \equiv \bar{Y}_i(0;\theta) - \bar{Y}_i(1;\theta)$ where $\bar{D}_i(\theta)$ represents the average direct effect under the randomization distribution, θ, $\bar{Y}_i(0;\theta)$ represents the average potential outcome under control and $\bar{Y}_i(1;\theta)$ the average potential outcome under

treatment. In the toy example, this represents the direct effect of the encouragement on product adoption outcomes among treated nodes in the network.

Indirect causal effects, the effect of the treatment on others, can generally be defined as the difference in individual average potential outcomes without treatment when a reference group[7] to which the individual is related in some way is treated and untreated: $\bar{I}_i(\theta, \eta) = \bar{Y}_i(\theta; \eta) - \bar{Y}_i(0; \theta)$. In our peer encouragement example this could represent the indirect effect of treatment on untreated peers of the treated, where a "peer" is a node with any arbitrary relationship to the treated node, including being connected, directly or indirectly, to the treated node in a social network or being a member of a group to which the treated node belongs (e.g., a classroom or a neighborhood).[8]

Total causal effects, the sum of the direct and indirect causal effects, can then be defined as the difference in individual average causal effects for an individual under treatment when an intervention is initiated (θ) and under no treatment when an intervention is not initiated (η): $\bar{T}_i(\theta, \eta) \equiv \bar{Y}_i(0; \eta) - \bar{Y}_i(1; \theta)$. In our example, the potential outcomes of interest are those of an encouraged node when the encouragement is introduced randomly to members of the node's reference group and those of a node that is not encouraged when no encouragement is introduced at all.

These estimands formalize the quantities of interest that form the basis for estimating the effects of networked social behavior. For example, in the peer encouragement example, we may be interested in estimating the micro-level peer influence effect of an encouragement to adopt a product on the peers of the encouraged. In other words, we may want to measure how peer influence affects the likelihood of adoption. Alternatively, we may be interested in the total effect on adoption in a reference group (e.g., a consumer population) of a program to encourage adoption by some members of the group. If peer spillovers exist, measuring the direct effect of treatment will underestimate the total effect of the intervention. Returning to our example, Aral and Walker (2011) are interested in the direct and indirect effects of viral features, but less so in isolating the total causal effect. The decision to randomly assign referee adopters to the baseline, active-personalized, and passive-broadcast conditions, rather than maintaining consistent treatment assignments throughout the diffusion process, enables robust estimation of direct and indirect effects and reflects a decision to forgo estimation of the total causal effects of the design decision. These trade-offs are inevitable in any networked experiment.

Data. It is worth mentioning briefly that the granularity of the data we can now collect through large networked experiments is changing the nature of what is estimable.

---

[7] A reference group is simply a group to which the individual belongs such that an intervention in that reference group could plausibly affect their potential outcomes. This could be a school, classroom, or social network to which the individual belongs.

[8] Indirect effects can exist as treatment effects on the untreated or reinforcement effects from treated nodes on other treated nodes. Assuming there is no complex interaction between treatments of distinct nodes, the indirect effects can be considered marginal effects holding own treatment status constant. Of course, interaction effects could exist and could subsequently be modeled.

Today's networked experiments, especially those conducted online, are collecting data on micro-level, individual behaviors and opinions in unprecedented detail and time stamped to the second. Our observations of human behavior have moved from static snapshots, to longitudinal panels, to continuous data streams (Domingos et al. 2000). Online platforms that record continuous streams of behavioral data are also themselves evolving over time. As a consequence, new machine learning methods are being devised to deal with key issues in the analysis of streaming data, including concept drift (Wang et al. 2003), clustering (Aggarwal et al. 2003), and querying (Babu and Widom 2001). Fine-grained data collection and new techniques for mining massive data streams are enabling new estimation strategies that incorporate subtle behaviors and time. As new data enable new models, future work should build theory around how granular, continuous time data can be incorporated into experimental analysis.

15.3.3 Estimation

Model Specification. The most straightforward approach to analyzing networked experiments is to estimate Average Treatment Effects (ATE): the difference in means between the treatment and control outcomes, as described in the estimands above. However, occasionally it may be fruitful to specify more sophisticated models that allow us to explore more of the heterogeneity in the data; for example, how treatment effects vary across subpopulations, what moderates the treatment effects, or the underlying behavioral processes that explain the outcomes of the experiment. Specifying a more sophisticated model allows more complex investigation, but at the cost of having to make more restrictive assumptions about treatment responses. For example, Aral and Walker (2011) incorporate time into their analysis by fitting hazard functions that estimate the rate at which individuals react to a treatment, rather than simply whether or not they react. They then validate their more sophisticated models with simple tests of differences in means that require fewer assumptions.

Interference. If the treatment response of one individual depends on the treatments of other individuals in the experiment, the treatments are said to "interfere" with one another. Interference is possible in a variety of settings including, for example, the effects of vaccination programs (Halloran and Struchiner 1991), housing vouchers (Sobel 2006), and educational interventions (Rosenthal and Jacobsen 1968). A substantial literature addresses contexts in which interference may exist between individuals in the same group, but not across groups (Sobel 2006; Rosenbaum 2007; Hudgens and Halloren 2008; Middleton and Aronow 2011; Tchetgen Tchetgen and VanderWeele 2012). This dependence structure, known as partial interference (Sobel 2006), is common in multi-level, group randomized, or cluster randomized studies, in which clusters represent well-defined groups, such as households (Tchetgen Tchetgen and VanderWeele 2012), neighborhoods (Ali et al. 2005), classrooms (Sobel 2006), or land plots (Kempton 1997).

Networked interference is in some sense a substantial departure from partial interference, as the pattern of interdependence is more complex and perhaps more uncertain in that it is not cleanly contained within well-defined groups. In another sense, however, recent approaches to interference in networks extend approaches to partial interference, modeling interference as a decreasing function of social distance and using the observed graph structure to define groups (or subgraphs) within which interference is more or less likely. Two sets of such strategies for addressing networked interference have emerged in recent years, which I refer to as *inference strategies* and *design strategies*.

Inference strategies correct for interference during analysis. For example, Aronow and Samii (2012) define exposure models that map assignment vectors and units' characteristics to exposure values or likelihoods of being subject to a given exposure. They then derive unbiased average treatment effect (ATE) estimators and variance estimators based on the assumptions of their exposure mapping. Airoldi et al. (2013) take a similar approach, deriving causal estimands based on the adjacency weighted matrix of exposure to treatment through social influence in a linear additive model. In Aronow and Samii's work, the basic approach is to model direct and indirect exposure and to base inference on the generalized probabilities of exposure implied by the model. For example, they consider an exposure model that assumes individuals' outcomes depend only on their own assignments and the assignments of their immediate neighbors. In their analysis of Add Health data, they assume a subject's outcomes depend only on their own assignment and a count of their treated neighbors. In both cases, each node's probability of being subject to a given exposure is precisely defined given a well defined experimental design and known network structure.

One important challenge with this approach is the difficulty of validating the exposure models. Aronow and Samii (2012: 19) acknowledge this criticism and make a convincing counterargument that "there is no escaping specification of exposure mappings for causal analysis." After all, analysis that does not specify a model of indirect exposure de facto embraces an assumption of no interference. However, it is possible to approach the problem by making less restrictive exposure assumptions and foregoing point identification to establish bounds on bias reduction. For example, Eckles et al. (2014) establish bounds on bias reduction achieved with estimands derived solely from successively more restrictive fractional neighborhood treatment response assumptions and an assumption of monotonicity in direct and indirect treatment responses (see their Theorem 2.2).[9] Using Manski's (2013) notion of effective treatments, they develop estimands that consider nodes as effectively treated if they satisfy successively more restrictive fractional neighborhood treatment response assumptions, where a node is considered treated if at least a fraction λ of its neighborhood is treated. For example, considering nodes as effectively treated if all of their neighbors are treated is more

---

[9] This is less restrictive than Aronow and Samii's (2014) generalized probability of exposure ($\pi_i(d_k)$) which precisely defines the probability that node i is subject to exposure d_k.

restrictive than including nodes as effectively treated if half of their peers are treated.[10] Assuming treatment responses are monotone, estimands with more restrictive assumptions on effectively treated nodes achieve bias reduction greater than or equal to estimands with less restrictive assumptions. Aral and Walker (2011) take a similar approach to interference, censoring peers any time after they have more than one treated neighbor, which defines a node as effectively treated at time t if and only if one of their peers was treated at time t.[11] These restrictive assumptions on effective treatment in experiments with independent assignment achieve bias reduction at the expense of increased variance (see Eckles et al. 2014, Figure 5). Though as I discuss below, when combined with design strategies like graph cluster randomization (Ugander et al. 2013), they can reduce bias and variance simultaneously under certain conditions.

Design strategies define treatment assignments over a network to minimize interference. More specifically, they use information about network structure to select treated nodes according to some interference minimizing protocol. Two types of design strategy take contrasting approaches to interference minimization (Walker and Muchnik 2014):

(1) Treatment clustering strategies randomize connected clusters of nodes, with high-edge density within clusters and low-edge density across clusters, into treatment and control groups to simulate global treatment and control conditions in which the entire network would either be treated or untreated. For example, Ugander et al. (2013) begin by defining an exposure condition called network exposure, in which a node is "networked exposed" if their treatment response under a given randomization distribution is the same as their treatment response under a randomization distribution in which the entire network is treated. They then investigate several conditions, analogous to fractional neighborhood treatments, which constitute network exposure. The main design element of the work is a "graph cluster randomization" procedure that partitions the graph into clusters across which treatments are randomized. The idea is to randomize treatment at the cluster level such that nodes are maximally surrounded by other nodes with the same treatment status. Although the protocol can admit any arbitrary clustering procedure and graph structure, Ugander et al. (2013) show that for any graph satisfying their restricted growth condition, specific graph cluster randomization protocols produce unbiased estimates with variance on the order of $1/n$ that is linear in the degrees of the graph. This is an important result because naïve clustering can produce estimators with variances that increase exponentially in the degrees. As Eckles et al. (2014) show, graph cluster randomization produces dramatic decreases in root mean squared error (RMSE) compared to independent assignment when peer effects are large (see their Figure 5). Graph cluster randomization is useful

---

[10] As Eckles et al. (2014) note, a fractional neighborhood treatment response assumption where $1 > \lambda > 0$ is less restrictive than Aronow and Samii's (2014) neighborhood treatment response assumption ($\lambda = 1$) and more restrictive than Manski's (2013) independent treatment response assumption ($\lambda = 0$).

[11] They then control for variation in degree by including a degree parameter in their model. A more comprehensive description of their approach to interference can be found in the paper's appendix (see Aral and Walker 2011).

when we want to estimate the effect of a global treatment. For example, if we are interested in changes in the degree to which local network externalities exist for a given technology, we may need to understand the effects of clustered adoption rather than of isolated adoption as the value to a new adopter is theoretically a function of the number of their peers who adopt the product.

One limitation to standard graph cluster randomization, which assigns all nodes in a cluster to the same treatment, is that it can become impossible to observe some nodes with a particular number of treated peers. This can complicate, for example, estimation of threshold effects in behavioral contagion studies because all nodes may not have a positive probability of being assigned to all threshold conditions (e.g., having 1, 2, 3, 4, or more treated peers). Eckles et al. (2014) therefore extend standard graph cluster randomization to allow for some nodes to be assigned to a different treatment than the rest of their cluster—a procedure they term "*hole punching*." Hole punching adds node-level randomness to the clustered treatment assignment by specifying independent switching variables that flip the assignment of some nodes in a cluster with a given probability. The result is that some fraction of the nodes in a cluster with a given cluster-level treatment assignment are assigned to the opposite condition, allowing for measurement of threshold effects in contagion or comparison of direct effects and peer effects.

(2) Treatment separating strategies assign treatments that maximize network distance between treated nodes in order to minimize interference between treatments. In contrast to Ugander et al. (2013), Coppock and Sircar (2013) develop a protocol that partitions the graph into non-interfering clusters according to assumptions about effective treatments described by what they call the "SUTVA degree" of a graph. The SUTVA degree corresponds to the maximum path length between two nodes over which interference is assumed to occur. A graph partition based on SUTVA degree produces a sample of nodes and edges that are relatively isolated from one another, enabling estimates of direct treatment and indirect exposure effects that minimize interference. Treatment separating strategies similar to Coppock and Sircar (2013) are not well suited to studying the effects of a global treatment or groups of commonly treated nodes, for instance in the case of estimating implications of changing network externality conditions. But they are well suited to studying micro-level peer effects. Airoldi et al. (2013) propose a different treatment separating randomization design called "insulated neighbors randomization" (INR), which assigns nodes to a non-exposure or k-level exposure status at random, such that the treatment assignments of the nodes already assigned are maintained and a percentage of shared neighbors between treated nodes are assigned to control. The purpose is to increase the causal information produced by the randomization at the expense of bias. Thomas and Finegold (2013) suggest a related block-randomized design in which a sample of nodes are selected to receive a treatment and a subset of their peers are randomly chosen to be denied access to (or awareness of) the treatment. They propose this design to deal with selection bias on degree and homophily between selected nodes and their peers, rather than interference per se. The peers of randomly selected nodes will be of higher degree and will exhibit observed and latent homophily with the selected

nodes. Block randomization can ameliorate this selection bias by randomizing at the level of the ego network. Aral and Walker (2012) achieve a similar goal through a mechanism randomization design that sends notifications of ego's behaviors only to randomly selected peers of the treated.

Eckles et al. (2014) observe larger bias and error reductions through design strategies than through inference strategies, though they admittedly explore specific design and inference strategies under a particular set of network and social effect conditions. Though the inference strategies they examine reduce bias, this bias reduction comes at a significant cost to precision. Combining design and inference approaches to networked interference also marginally improve both bias reduction and precision, though the effects on precision come mainly from assignment clustering rather than from neighborhood response assumptions or weighting.

Estimators (and their Variance). Since network exposure probabilities vary across nodes, estimators must adjust for different proportions of units in different exposure conditions. The Horvitz-Thompson (HT) estimator can produce unbiased estimates the differences in means of the ATE estimand in stratified samples, using inverse probability weighting to adjust for the varying proportions (Horvitz and Thompson 1952). However, heavy-tailed degree distributions in networks ensure high variability in the likelihood of network exposure, generating very large or small numbers of units with high weights, increasing the variance of the HT estimator, at times dramatically. Several adjustments have been suggested to reduce the variance of the HT estimator in the context of network experiments. The Hajek (1971) refinement, which uses the sum of the weights in the denominator, achieves efficiency gains at the cost of bias, though increases in bias are typically small relative to estimator's variability. Bootstrap methods have been evaluated to deal with a lack of coverage of standard difference in means estimators (Thomas and Finegold 2013) and the multiway dependence structure common in online settings (Bakshy and Eckles 2013). Ugander et al.'s (2013) graph cluster randomization further reduces the variance of both the HT and Hajek estimators by increasing the expected number of the individuals who are network exposed to a treatment or control condition and limiting the number of individuals outside of selected clusters who are network exposed (see Eckles et al. 2014, Figure 5).

15.4 THE FUTURE OF NETWORKED EXPERIMENTATION

15.4.1 Adaptive Treatment Assignment

An area of innovation that may be relevant to the future of networked experimental design is the extension of sequential treatment assignment to more adaptive dynamic

treatment assignment protocols. Such protocols might consider the sample charac-
teristics or responses of the initially treated to inform downstream assignments to
maximize precision, minimize interference or cost or harm, or to investigate treatment
effect heterogeneity or a particular hypothesis. Sequential treatment assignment, in
which a node's assignment is a function of network structure and prior assignments,
has been proposed as an approach to minimize interference (Airoldi et al. 2013).
However, one could imagine adaptive assignments that use response information
obtained during early estimation to adjust the proportions or types of nodes assigned
to a given treatment. Adaptive treatment assignment has been used in medical trials to
expose fewer patients to avoidable harm than strategies with fixed assignments (Hu and
Rosenberger 2006). Though these adaptive cases are difficult to optimize in situations in
which the number of experimental units is large, heuristic approaches and near-optimal
strategies have been shown to be valid in non-networked settings, even when the size
of study population is unknown or unbounded (Press 2009).

Adaptive treatment assignment is essentially a bandit problem in which the researcher
wishes to allocate treatments among experimental units, whose properties and responses
are only partially known at the time of allocation, but which may become better
understood as time passes during the experiment (Berry and Fristedt 1985; Berry
2006; Press 2009). Bandit problems exhibit the classic trade-off between the cost of
gathering information and the benefit of exploiting information already gathered (e.g.,
exploration vs. exploitation). I can imagine a line of research that develops adaptive
treatment assignment procedures in network contexts, deriving optimal solutions
to relevant allocation problems, incorporating network information and inference
into the approaches. Networked information and the regularity of assortativity and
other networked characteristics may prove useful in cases analogous to contextual
bandit problems because social structure provides so much information about the
relationships between experimental units and their relative context.

15.4.2 Novel Randomization Techniques

The flexibility of modern application programming interfaces (APIs) is enabling
new forms of randomization in networked experiments. As more comprehensive
observability and new communication and messaging features imbue our online social
networks with new functionality, new ways of creating exogenous variation at different
network levels are constantly emerging. Several examples demonstrate the power of
modern online environments to create useful variation with which to identify causal
effects in social behavior.

Aral and Walker (2012) utilize the Facebook API to randomize the peer recipients of
automated notifications of a node's behaviors. For example, when a user rated a movie
or linked to a celebrity, packets of notifications informing their peers of their behaviors
were automatically generated and randomly delivered to their Facebook friends. This

randomization enables them to identify peer influence, susceptibility to influence, and individual and dyadic correlates of influence and susceptibility at the individual level. Bakshy et al. (2012b) use a similar design by directly manipulating Facebook's algorithms, randomly blocking information delivered to users' news feeds in order to identify causal information diffusion dynamics in networks. These types of micro-level randomization enable quite nuanced investigations of social behavior.

Bakshy et al. (2012b) and Aral and Walker (2012, 2014) examined correlates of social influence by randomizing messages between individuals in the network. However, these designs cannot estimate the causal effect of receiving a message from a particular individual with particular characteristics. Taylor et al. (2015) select ad exposures where viewers could potentially be exposed to social cues involving two different peers and then randomly choose which peer is shown, providing exogenous variation in the dyadic characteristics of friends displayed in the social cue.[12] This exogenous variation enables causal estimates of the influence of particular social cues (i.e., the age, gender, similarity, or tie strength associated with a particular peer) on behavior. These types of randomization will form the basis of personalized social advertising in the near future.

Finally, anonymity designs provide another novel randomization technique that can help identify subtle social effects. Bapna et al. (2014) enable randomly selected users of an online dating site to browse other users' profiles without leaving a digital signal of their interest (the default setting on the website is to provide viewing information to the user whose profile is being viewed). This enables causal estimates of the effects of anonymity on both browsing behaviors and the matches that are eventually formed. Interestingly, anonymous browsers are more likely to view other users who are of different races. Taylor et al. (2014) use a similar anonymity design to suppress the identity of users posting content to a social news aggregation website. They find strong identity effects on opinion formation as measured through ratings. These examples highlight the new designs that are possible today. In the future, nuanced randomization procedures such as these are likely to improve our understanding of subtle social effects at scale.

15.4.3 Linking Online Treatments to Offline Responses

Although digitization is enabling advances in networked experimentation, there remains a danger in relying too heavily on digital substrates to explore human behavior. Not only are digital samples biased toward those who are more active online, potentially missing large swaths of society, but limiting inquiry to digital behaviors constrains the theoretical reach of experimental work. It will therefore be important to devise strategies for linking online treatments to meaningful offline behaviors.

---

[12] This paper is based on Chapter 3 of Sean Taylor's PhD thesis (Taylor 2014).

For example, Bond et al. (2012) go to great lengths to validate the results of their Facebook voting experiment in public voting records. They find that users who received a social message were more likely to vote than users who received no message or an informational message with no social cue, raising doubts about the efficacy of information-only appeals to vote (Aral 2012). Similarly, in an ongoing collaboration with the Praekelt Foundation in South Africa, we are connecting data from experiments conducted on a mobile messaging platform with data on HIV testing conducting in physical clinics. We designed scratch cards with unique identifiers that are given to people who test in a facility and mention the program. They unlock their experimental incentive by entering the unique identifier on the card into their phone. When this identifier is entered, we can connect cell phone referral data to data on verified HIV tests. To maintain anonymity, we only record users' data as de-identified strings and never inquire about individual HIV test results, only whether a test was taken.

These examples provide evidence of the importance of linking online and offline behaviors in digital networked experiments. Validating the effects of digital treatments on physical behaviors or the digital measurement of physical behaviors will help establish and maintain the external validity of studies conducted on digital substrates that wish to generalize findings to a broader set of people and subjects.

15.4.4 Experimental Validation of Observational Methods

An important use of networked experiments is in benchmarking the results of observational studies against provably causal estimates. For example, Aral et al. (2009) adapted high dimensional propensity score matching (HDPSM) to dynamic network settings to identify peer influence in product adoption decisions. They analyzed a global instant messaging network of 27.4 million Yahoo users, combined with comprehensive daily product adoption data on a personalized news and weather application and users' longitudinal behavioral, demographic, and geographic characteristics. They found that previous methods overestimated peer influence by 300%–700%, and that homophily explained greater than 50% of the perceived behavioral contagion. Shalizi and Thomas (2012) subsequently argued that, theoretically speaking, homophily and contagion are "generically confounded." They claimed that latent unobserved (or unobservable) homophily will always thwart the identification of peer effects. Given the expense and difficulty of running experiments and the availability of observational data, it would be useful to develop reliable, nonexperimental causal inference strategies in networks. But how can we know how close our observational estimates are to provably causal estimates of social effects? How can we measure the error and bias of observational methods that potentially miss unobserved latent effects?

Eckles and Bakshy (2014) demonstrated one approach to the experimental validation of observational methods by conducting a "constructed observational study" evaluating the effectiveness of HDPSM in identifying peer effects in networks. Their study uses the

results of a randomized experiment of peer influence in information diffusion among 35 million Facebook users (Bakshy et al. 2012b) to benchmark HDPSM methods applied to identify the same effects. In essence, they treat the results of the randomized experiment as a gold standard and measure how well, or how poorly, HDPSM does in identifying peer effects, compared to the experimental estimates, using the same data.

Three notable results emerge from their study: First, HDPSM methods can achieve a 70% bias reduction in estimates of peer effects compared to experimental benchmarks. This provides grounds for optimism in that observational techniques can effectively estimate peer effects and other causal effects in networks. Whether estimates that achieve 70% bias reduction are useful to policy-makers depends on the policy being evaluated. In many cases, a reliable estimate of the direction and order of the effects can help guide policy. In other cases, however, more reliable estimates may be necessary. Second, contextual variables used in the first-stage matching procedure are critical to bias reduction. Eckles and Bakshy (2014) find models that use a limited number of demographic features to create the matched sample only achieve approximately an 8% bias reduction. Contextual features relevant to the behavior being modeled improve matches and thus the reliability of matching estimates. Finally, the overestimates of social contagion by naïve methods found in Aral et al. (2009) and Eckles and Bakshy (2014) are quite similar. Both studies find naïve models overestimate social contagion by approximately 300%.[13] These benchmarks help us understand the errors produced by observational methods.

Constructed observational studies are one example of how networked experiments can be used to validate observational approaches to the identification of social effects. Using experimental results as a benchmark for true causal estimates can help us understand when and to what degree observational methods produce bias.[14]

15.5 CONCLUSION

Networked experiments are rapidly becoming one of the most precise tools available for studying social behavior. As more and more social behaviors are digitized and mediated by online platforms, our ability to quickly answer nuanced causal questions about the role of social behavior in population-level outcomes is becoming unprecedented. Interdependent behaviors potentially represent the most fundamental data-generating processes in the social sciences, as they are in some sense what make the social sciences social.

---

[13] Aral et al. (2009) find a 700% overestimate in the early weeks of the product's life cycle, immediately after the product launch. Their estimates then decrease to a consistent 300% overestimate.

[14] Of course, experiments are not necessarily a silver bullet. There exist many known potential biases in experimental settings as well. However, under certain assumptions and with robust execution, experiments provide the most reliable causal estimates of networked behavior.

Yet, while the increasing scale, scope, and complexity of networked experiments creates tangible opportunities for scientific advancement, they also simultaneously create significant new challenges. This chapter presented a review of advances in modern networked experimentation, summarizing the opportunities and challenges they create for understanding our social world. Fundamental innovations in networked experimental design are changing the implications of the experimental setting, sampling, randomization procedures, and approaches to treatment assignment. At the same time, the analysis of networked experiments is witnessing a rapid evolution, from modeling treatment response assumptions, measurement and inference to recent approaches to interference, and uncertainty in dependent data. These changes require networks researchers to be aware of the most modern approaches to experiments in interdependent data. At the same time, new challenges and opportunities are shaping the future of networked experimentation, particularly in adaptive treatment assignment strategies, novel randomization techniques, linking online treatments to offline responses and experimental validation of observational methods.

There is much work to do as we usher in this new era of networked experimentation and I cannot think of a more important set of social scientific endeavors. Scalable networked experimentation represents an application of the scientific method to questions about our social world, at unprecedented scale and precision. It behooves us to tackle such scientific challenges with vigor as we develop these new tools in the pursuit of human progress.

References

Aggarwal, C. C., Han, J., Wang, J., and Yu, P. S. (2003). "A framework for clustering evolving data streams." *Proceedings of the 29th International Conference on Very Large Data Bases* (29).

Aharony, N., Pan, W., Ip, C., Khayal, I., and Pentland, A. (2011). "Social fMRI: Investigating and shaping social mechanisms in the real world." *Pervasive and Mobile Computing* 7, 643–659.

Ahmed, N., Neville, J., and Kompella, R. (2014). "Network sampling: From static to streaming graphs." *ACM Transactions on Knowledge Discovery from Data*, 8.

Aiello, W., Chung, F., and Lu, L. (2001). "A random graph model for power law graphs." *Experimental Mathematics* 10(1), 53–66.

Airoldi, E., Toulis, P., Kao, E., and Rubin, D. B. (2013). "Estimation of causal peer influence effects." *Journal of Machine Learning Research* 28(3), 1489–1497.

Ali, M., Emch, M., von Seidlein, L., Yunus, M., Sack, D. A., Rao, M., Holmgren, J., and Clemens, J. D. (2005). "Herd immunity conferred by killed oral cholera vaccines in Bangladesh: A reanalysis." *Lancet* 366, 44–49.

Aral, S. (2011). "Commentary—identifying social influence: A comment on opinion leadership and social contagion in new product diffusion." *Marketing Science* 30(2), 217–223.

Aral, S. (2012). "Social science: Poked to vote." *Nature* 489(7415), 212–214.

Aral, S. and Walker, D. (2011a). "Creating social contagion through viral product design: A randomized trial of peer influence in networks." *Management Science* 57(9), 1623–1639.

Aral, S. and Walker, D. (2011b). "Identifying social influence in networks using randomized experiments." *IEEE Intelligent Systems* 26(5), 91–96.

Aral, S. and Walker, D. 2012. "Identifying influential and susceptible members of social networks," *Science* 337(6092), 337–341.

Aral, S. and Van Alstyne, M. W. (2011). "The diversity-bandwidth tradeoff." *American Journal of Sociology* 117(1), 90–171.

Aral, S. and Taylor, S. (2011). "Viral incentive systems: A randomized field experiment." Working Paper, Stern School of Business, NYU.

Aral, S., Muchnik, L., and Sundararajan, A. (2009). "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks." *Proceedings of the National Academy of Sciences* 106(51), 21544–21549.

Aral S., Muchnik L., and Sundararajan, A. (2013). "Engineering social contagions: Optimal network seeding and incentive strategies." *Network Science* 1(2), 1–29.

Aronow, P. M. and Samii, C. (2012). "Estimating average causal effects under general interference." Summer Meeting of the Society for Political Methodology, University of North Carolina, Chapel Hill, July.

Bakshy, E. and Eckles, D. (2013). "Uncertainty in online experiments with dependent data: An evaluation of bootstrap methods." *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Bakshy, E., Eckles, D., Yan, R., and Rosenn, I. (2012a). "Social influence in social advertising: Evidence from field experiments." *Proceedings of the 13th ACM Conference on Electronic Commerce*, 146–161. ACM, New York.

Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. (2012b). "The role of social networks in information diffusion." *Proceedings of 21st International Conference on World Wide Web*, 519–528. ACM, New York.

Bapna, R., Gupta, A., Rice, S., and Sundararajan, A. 2011. "Trust, reciprocity and the strength of social ties: An online social network based field experiment." Working paper, University of Minnesota, Minneapolis.

Bapna, R., Ramaprasad, J., Shmueli, G., and Umyarov, A. (2013). "One-way mirrors and weak-signaling in online dating: A randomized field experiment." Working paper, University of Minnesota, Minneapolis.

Bapna, R., Umyarov, A. Forthcoming. "Do your online friends make you pay? A randomized field experiment in an online music social network." Forthcoming in *Management Science*.

Bavelas, A. (1950). "Communication patterns in task-oriented groups." *The Journal of the Acoustical Society of America* 22(6), 725–730.

Babu, S. and Jennifer, W. (2001). "Continuous queries over data stream." *ACM Sigmod Record* 30(3), 109–120.

Berry, D. A. (2006). "Bayesian clinical trials." *National Review of Drug Discovery* 5, 27–36.

Berry, D. A. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments*. London, UK: Chapman & Hall.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., 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.

Boudreau, Kevin, J., and Lakhani, Karim R. (2011). "High incentives, sorting on skills—or just a taste for competition? Field experimental evidence from an algorithm design contest." No. 11-107. Harvard Business School Working Paper.

Bramoullé, Y., Habiba, D., and Fortin, B. (2009). "Identification of peer effects through social networks." *Journal of Econometrics* 150(1), 41–55.

Brock, W. A. and Durlauf, S. N. (2001). "Discrete choice with social interactions." *The Review of Economic Studies* 68(2), 235–260.

Burke, M. A. and Sass, T. R. (2013). "Classroom peer effects and student achievement." *Journal of Labor Economics* 31(1), 51–82.

Centola, D. (2010). "The spread of behavior in an online social network experiment." *Science* 329(5996), 1194–1197.

Centola, D. (2011). "An experimental study of homophily in the adoption of health behavior." *Science* 334(6060), 1269–1272.

Centola, D. and Macy, M. (2007). "Complex contagions and the weakness of long ties." *American Journal of Sociology* 113(3), 702–734.

Coppock, A. and Sircar, N. (2013). "Design of field experiments under unknown interference structures." Working paper.

Cox, D. R. (1958). *Planning of Experiments*. New York: Wiley-Interscience.

Dieye, R., Djebbari, H., and Barrera-Osorio, F. (2014). "Accounting for peer effects in treatment response." IZA Discussion Paper No. 8340, Institute for the Study of Labor, Bonn, Germany.

Domingos, P. and Geoff, H. (2000). "Mining high-speed data streams." *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Eckles, D. and Bakshy, E. (2014). "Bias and high-dimensional adjustment in observational studies of peer effects." Working paper, Data Science Team, Facebook, Menlo Park, CA.

Eckles, D., Karrer, B., and Ugander, J. (2014). "Design and analysis of experiments in networks: Reducing bias from interference." Working paper, Data Science Team, Facebook, Menlo Park, CA.

Erdős, P. and Rényi, A. (1959). "On random graphs. I." *Publicationes Mathematicae* 6, 290–297.

Erdős, P. and Rényi, A. (1960). "On the Evolution of Random Graphs." *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, 17–61.

Erdős, P. and Rényi, A. (1961). "On the strength and connectedness of random graphs." *Acta Mathematica Academiae Scientiarum Hungarica* 12(1–2), 261–267.

Fowler, J. H. and Christakis, N. A. (2010). "Cooperative behavior cascades in human social networks," *Proceedings of the National Academy of Sciences USA* 107(12), 5334–5338.

Graham, B. S. and Hahn, J. (2005). "Identification and estimation of the linear-in-means model of social interactions." *Economics Letters* 88(1), 1–6.

Granovetter, M. (1978). "Threshold models of collective behavior." *American Journal of Sociology* 83(6), 1420–1443.

Hasan, S. and Bagde, S. Forthcoming. "Peers and network growth: Evidence from a natural experiment." *Management Science.*

Hajek, J. (1971). "Comment on 'An Essay on the Logical Foundations of Survey Sampling, Part One.'" in *The Foundations of Survey Sampling*, V. P. Godambe and D. A. Sprott, eds., 236. Holt, Rinehart, and Winston.

Halloran, M. E. and Struchiner, C. J. (1991). "Study designs for dependent happenings." *Epidemiology* 2, 331–338.

Halloran, M. E. and Struchiner, C. J. (1995). "Causal inference in infectious diseases." *Epidemiology*, 6, 142–151.

Heckathorn, D. D. (1997). "Respondent-driven sampling: A new approach to the study of hidden populations." *Social Problems* 44(2), 174–199.

Hinz, O., Skiera, B., Barrot, C., and Becker, J. U. (2011). "Seeding strategies for viral marketing: An empirical comparison." *Journal of Marketing* 75(6), 55–71.

Horton, J., Rand, D. G., and Zeckhauser, R. J. (2010). "The online laboratory: Conducting experiments in a real labor market." *Experimental Economics* 14(3), 399–425.

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.

Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley: New York.

Hudgens, M. G. and Halloran, E. (2008). "Toward causal inference with interference." *Journal of the American Statistical Association* 103(482), 832–842.

Kearns, M., Suri, S., and Montfort, N. (2006). "An experimental study of the coloring problem on human subject networks." *Science* 313(5788), 824–827.

Kempton, R. A. (1997). "Interference between plots." In *Statistical Methods for Plant Variety Evaluation*, R. A. Kempton and P. N. Fox, eds., 101–116. London: Chapman & Hall.

Leavitt, H. J. (1951). "Some effects of certain communication patterns on group performance." *The Journal of Abnormal and Social Psychology* 46(1), 38.

Lee, L. (2007). "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects." *Journal of Econometrics* 140(2), 333–374.

Lee, S., Kim, P., and Jeong, H. (2006). "Statistical properties of sampled networks," *Physical Review E* 73, 016102.

Leider, S., Möbius, M. M., Rosenblat, T., and Do, Q.-A. (2009). "Directed altruism and enforced reciprocity in social networks." *The Quarterly Journal of Economics* 124(4), 1815–1851.

Leskovec, J., Kleinberg, J., and Faloutsos, C. (2005). "Graphs over time: Densification laws, shrinking diamaters and possible explanations." In *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Leskovec, J. and Faloutsos, C. (2006). "Sampling from large graphs," In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 631–636.

Mani, A., Rahwan, I., and Pentland, A. (2013). "Inducing peer pressure to promote cooperation." *Scientific Reports* 3.

Manski, C. F. (1993). "Identification problems in the social sciences." *Sociological Methodology* 23.

Manski, C. F. 2013. "Identification of treatment response with social interactions." *The Econometrics Journal* 16(1): S1–S23.

Mason, W. and Suri, S. 2012. "Conducting behavioral research on Amazon's Mechanical Turk." *Behavioral Research* 44, 1–23.

Mason, W. and Watts, D. J. 2012. "Collective learning in networks." *Proceedings of the National Academy of Sciences* 109(3), 764–769.

Middleton, J. A. and Aronow, P. M. 2011. "Unbiased estimation of the average treatment effect in cluster-randomized experiments." Working paper, New York University, New York.

Muchnik, L., Aral, S., and Taylor, S. 2013. "Social influence bias: A randomized experiment." *Science* 341(6146), 647–651.

Newman, M. E. J. 2002. "Random graphs as models of networks." Working paper, Santa Fe Institute.

Neyman, J. 1923. "On the application of probability theory to agricultural experiments: Essay on principles, Section 9." Roczniki Nauk Roiniczych, tom X, 1–51 [in Polish]; reprinted in English with discussion in *Statistical Science* 5, 463–480.

Press, W. H. 2009. "Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research." *Proceedings of the National Academy of Sciences* 106(52), 22387–22392.

Rand, D. G. and Nowak, M. A. 2011. "The evolution of antisocial punishment in optional public goods games." *Nature Communications. Article* 434.

Rosenbaum, P. R. 2007. "Interference between units in randomized experiments." *Journal of the American Statistical Association* 102, 477.

Rosenthal, R. and Jacobson, L. 1968. *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*. New York: Holt, Rinehart and Winston.

Rubin, D. B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studie." *Journal of Educational Psychology* 66, 688–701.

Rubin, D. B. 1986. "Which ifs have causal answers." *Journal of the American Statistical Association* 81(396), 961–962.

Rubin, D. B. 1990. "Comment: Neyman (1923) and causal inference in experiments and observational studies." *Statistical Science* 5(4), 472–480.

Salganik, M. J., Dodds, P. S., and Watts, D. J. 2006. "Experimental study of inequality and unpredictability in an artificial cultural market." *Science* 311(5762), 854–856.

Salganik, M. J. and Heckathorn, D. D. 2004. "Sampling and estimation in hidden populations using respondent-driven sampling." *Sociological Methodology* 34(1), 193–240.

Shalizi, C. R. and Thomas, A. C. 2011. "Homophily and contagion are generically confounded in observational social network studies." *Sociological Methods and Research* 40(2), 211–239.

Shore, J., Ethan, B., and Lazer, D. 2014. "Facts and figuring: An experimental investigation of network structure and performance in information and solution spaces." Harvard Business School Working Paper, No. 14-075.

Sobel, M. 2006. "What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference." *Journal of the American Statistical Association* 101, 1398–1407.

Stumpf, M., Wiuf, C., and May, R. 2005. "Subnets of scale-free networks are not scale-free: Sampling properties of networks." *Proceedings of the National Academy of Sciences* 102(12), 4221–4224.

Suri, S. and Watts, D. J. 2011. "Cooperation and contagion in web-based, networked public goods experiments." *PLoS One* 6(3), e16836.

Taylor, S. J. 2014. "Social influence from online social signals: Three experiments." PhD thesis, NYU Stern School of Business.

Taylor, S., Bakshy, E., and Aral, S. 2013. "Selection effects in online sharing: Consequences for peer adoption." *14th ACM Conference on Electronic Commerce*, Philadelphia, PA.

Taylor, S., Muchnik, L., and Aral, S. 2014. "Identity and opinion: A randomized experiment." Working paper, MIT Sloan School of Management.

Taylor, S. J., Bakshy, E., Eckles, D., and Aral, S. 2014. "Asymmetric effects of personalized social cues: Evidence from field experiments." Working paper.

Tchetgen Tchetgen, E. J. and VanderWeele, T. J. 2012. "On causal inference in the presence of interference." *Statistical Methods in Medical Research* 21(1), 55–75.

Thomas, A. C. and Finegold, M. 2013. "Protocols for randomized experiments to identify network contagion." Working paper, CMU.

Tucker, C. 2011. "Social advertising." Working paper, MIT Sloan School of Management, Cambridge, MA.

Tucker, C. and Zhang, J. 2010. "Growing two-sided networks by advertising the user base: A field experiment." *Marketing Science* 29(5), 805–814.

Tucker, C. and Zhang, J. 2011. "How does popularity information affect choices? A field experiment." *Management Science* 57(5), 828–842.

Ugander, J., Karrer, B., Backstrom, L., and Kleinberg, J. 2013. "Graph cluster randomization: Network exposure to multiple universes." *19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 329–337. ACM, New York.

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

Wang, H., Fan, W., Yu, P. S., and Han, J. 2003. "Mining concept-drifting data streams using ensemble classifiers" *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

Watts, D. J. 2002. "A simple model of global cascades on random networks." *Proceedings of the National Academy of Sciences* 99(9), 5766–5771.