Management Science

Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment

Sinan Aral, Dylan Walker

Please scroll down for article—it is on subsequent pages

Tie Strength, Embeddedness, and Social Influence: A Large-Scale Networked Experiment

Sinan Aral

MIT Sloan School of Management, Cambridge, Massachusetts 02142, sinan@mit.edu

Dylan Walker

School of Management, Boston University, Boston, Massachusetts 02215, dtwalker@bu.edu

We leverage the newly emerging business analytical capability to rapidly deploy and iterate large-scale, microlevel, in vivo randomized experiments to understand how social influence in networks impacts consumer demand. Understanding peer influence is critical to estimating product demand and diffusion, creating effective viral marketing, and designing "network interventions" to promote positive social change. But several statistical challenges make it difficult to econometrically identify peer influence in networks. Though some recent studies use experiments to identify influence, they have not investigated the social or structural conditions under which influence is strongest. By randomly manipulating messages sent by adopters of a Facebook application to their 1.3 million peers, we identify the moderating effect of tie strength and structural embeddedness on the strength of peer influence. We find that both embeddedness and tie strength increase influence. However, the amount of physical interaction between friends, measured by coappearance in photos, does not have an effect. This work presents some of the first large-scale in vivo experimental evidence investigating the social and structural moderators of peer influence in networks. The methods and results could enable more effective marketing strategies and social policy built around a new understanding of how social structure and peer influence spread behaviors in society.

1. Introduction

The newly emerging capability to rapidly deploy and iterate microlevel, in vivo randomized experiments in complex social and economic settings at population scale is, in our view, one of the most significant innovations in modern business analytics and what separates today's analytics from the analytical processes of the last several decades. As more and more social interactions and commercial transactions are digitized and mediated by online platforms, our ability to 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. This new tool kit portends a sea change in our scientific understanding of human behavior and dramatic improvements in business and social policy as a result. We leverage this new tool kit to better understand one of the most significant recent developments in consumer behavior: how social influence in networks impacts consumer demand.

Social influence in networks is recognized as a key factor in the propagation of ideas, behaviors, and economic outcomes in society. Understanding the role social influence plays in spreading economic behaviors is therefore critical to constructing sound policy in both the public and private sectors. Emerging online systems, which increasingly connect people and mediate their interactions, now provide opportunities to acquire microlevel data at population scale (e.g., Eagle et al. 2010, Golder and Macy 2011) and to conduct experiments that address endogeneity (e.g., Aral and Walker 2011a, Bakshy et al. 2012a). These two advantages can be leveraged to create new experimental analytical methods that yield real-time, context-specific inferences about the role of social influence in consumer demand, marketing, and public policy. Moreover, large-scale data from randomized experiments permit the detection of nuanced or subtle effects that are economically important but difficult to observe with observational analytics. It is precisely these nuanced effects that are in danger of being eclipsed by bias in endogenous processes, making randomized experimentation in large-scale social systems a vital new tool in the arsenal of modern business analytics.

A primary question in understanding the role of social influence in the diffusion of new products, ideas, behaviors, and outcomes is how heterogeneity in the relationships between individuals impacts the influence they exert on one another. Despite decades of observational research, results in this domain remain inconsistent and elusive. This is perhaps unsurprising since statistical challenges like simultaneity (Godes and Mayzlin 2004), homophily (Aral et al. 2009), unobserved heterogeneity (Van den Bulte and Lilien 2001), time-varying factors (Van den Bulte and Lilien 2001), and other contextual and correlated effects (Manski 1993) make it difficult to distinguish causal peer influence from other confounds that create behavioral clustering in network space and time.

Research has employed instrumental variables methods (e.g., Tucker 2008) and high-dimensional propensity score matching (HDPSM) (Aral et al. 2009) to identify peer influence and distinguish it from homophily and other confounding factors in observational data. HDPSM has recently been shown to reduce bias in causal estimates of peer influence by up to 75% (Eckles and Bakshy 2014), which is useful since most available data in this domain are currently observational. Nonetheless, controlling for unobservable factors like latent homophily remains difficult (Shalizi and Thomas 2011).

As an alternative to observational analysis, experimental network studies using random assignment can provide a more robust means of identifying causal peer effects and distinguishing influence from confounding factors. Some recent experiments have demonstrated a role for peer influence in product adoption (Aral and Walker 2011a, 2012; Bakshy et al. 2012b; Bapna and Umyarov 2012), health behaviors (Centola 2010), and altruism (Leider et al. 2009). Though these studies use experiments to address confounds and identify peer influence in different network contexts, they have not investigated in vivo the social or structural conditions under which influence is strongest, an area identified as a critical new frontier in the science of social influence (Aral 2011).[1] Two of the most widely studied social factors theorized to affect the strength of social influence are *embeddedness*, the extent to which individuals share common peers, and *tie strength*, the significance or intensity of relationships. We therefore investigate how embeddedness and tie strength moderate social influence in product adoption, while simultaneously controlling for confounding factors that can bias inference in networked settings.

We conducted a randomized experiment measuring social influence in the adoption of a commercial application among 1.3 million users of the online social network Facebook.com. Using novel techniques of randomized experimentation in networked environments and statistical analysis, we identify and distinguish influence-driven outcomes from spontaneous outcomes and examine the roles social embeddedness and tie strength play in the level of influence exerted between peers. Theoretically, we extend the definition of *tie strength* beyond the frequency of communication to examine several precisely defined measures of the strength of ties (SoT) that describe the nature of the relationship between individuals and their peers in a concrete manner, including (a) the social context of the relationship (how individuals met, know one another, or interact with each other, e.g., whether peers attended the same college, come from the same hometown, or share common institutional affiliations), (b) the recency of the relationship (e.g., whether peers currently live in the same town), (c) the overlap of common interests (e.g., being fans of the same Facebook pages or joining the same Facebook groups), and (d) the frequency of physical interaction (e.g., copresence in online photos).

Our study complements and extends prior work on the role that individual attributes play in social influence processes. Prior research has demonstrated that, in spreading processes, not all *individuals* are created equal (by analyzing how individuals' attributes moderate diffusion; Aral and Walker 2012). Our work extends this line of inquiry and demonstrates that not all *relationships* are created equal (by theorizing and analyzing the role of relationship characteristics in moderating diffusion). Our approach can be generalized and extended to a multitude of systems of interest to researchers in economics, sociology, marketing, management, information systems, and other quantitative social sciences in which large-scale randomized field experiments are rapidly gaining traction as a powerful new analytical tool. The methods and results are also relevant to managers and policy makers applying experimental analytics to a variety of real business and social policy arenas in which social influence is a primary driver of behavior change.

2. A New Experimental Paradigm for Large-Scale, Social Scientific, Business Analytics

The phrase "Big Data" suggests that the main power in modern analytical processes is in the size and scale of the observational data we now collect (Mayer-Schonberger and Cukier 2013). Our view, however, is that the real power lies in the granularity of the data (not just its scale) combined with a new ability to engineer and randomize social settings to (a) robustly estimate the causal effects of different policy alternatives,

---

[1] Two notable exceptions to this are the studies by Bond et al. (2012) and Bakshy et al. (2012a), who examined the moderating role of tie strength (defined by interaction volume) on influence in voting behaviors and advertising, respectively. In this work, we go further by simultaneously considering multiple different categories of tie strength.

(b) explore the heterogeneity in these causal effects across subpopulations of consumers, and (c) unpack the nuanced behavioral mechanisms that underlie and explain the causal outcomes of policy experiments. These new tools portend a sea change in our understanding of human behavior, for both business analytics and social science. By understanding the causal behavioral mechanisms underlying the outcomes of specific policies, how and why those outcomes vary across different consumers, and how they change over time, we can develop more contextual and personalized, and therefore more effective, business and social policies.

Organizations have employed multivariate "A/B" testing of online products and services for several years now. But recently the precision and complexity of the experimental tool kit has increased dramatically. As online platforms have scaled to support hundreds of millions of simultaneous users and platform design has become more open and precise, the ability to test complex dynamic hypotheses about social and behavioral phenomena has expanded. Application programming interfaces and explicit developer controls now expose functional platform elements that researchers can use to execute new experimental designs. 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 microtasks (Mason and Watts 2012, Rand and Nowak 2011, Suri and Watts 2011); and formal collaboration with platform developers and website administrators enables researchers to achieve even more comprehensive and precise experimental control in large-scale in vivo environments (Bakshy et al. 2012a, b; Muchnik et al. 2013).

These digital tools are enabling a new era of experimental social science and analytics that has begun to reveal robust evidence of the nuanced causal determinants of consumer behavior and the implied effectiveness of different social policies and business strategies. For example, recent large-scale digital experiments have revealed specific insights about product adoption and engagement (Aral and Walker 2011a, 2012; Bapna and Umyarov 2012; Taylor et al. 2013), social commerce and advertising (Aral and Taylor 2014, Bakshy et al. 2012a, Tucker 2011), information sharing and diffusion (Bakshy et al. 2012b), herd behaviors in cultural markets (Muchnik et al. 2013, Salganik et al. 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), 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. 2012). As a reference, we summarize the focus, context, experimental procedures, and scale of recent large-scale digital experiments in Table 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 (e.g., Suri and Watts 2011), and individual (e.g., Aral and Walker 2012) levels to reveal important insights about human behavior at the 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 nuanced experimental tests of microlevel 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. 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 (2012) 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 business policies tailored for particular users. Furthermore, creating a deep understanding of the data-generating processes that explain the average behavioral effects of policy interventions prepare organizations to respond to changes in the data-generating processes and to know how their interventions will create dynamic changes in behavior over time.

Our work builds on a specific line of research into social influence in information diffusion and product adoption in networks (Aral and Walker 2011a, 2012; Bakshy et al. 2012a, b; Bapna and Umyarov 2012; Taylor et al. 2013; Tucker 2011). In particular, we extend an experimental method for measuring heterogeneity of treatment effects in the impact of influence-mediating messages on peer behavior from the individual level (studied by Aral and Walker 2012) to the relationship level. As we describe below, message randomization can be combined with statistical analysis to measure influence and susceptibility in networks and the heterogeneity of treatment effects across observable characteristics of senders and recipients. We extend this approach to examine structural and dyadic relationship characteristics that moderate the influence someone exerts on his or her peers. In particular, we focus on embeddedness and tie strength because these are two of the most widely studied relationship specific factors theorized to moderate social influence.

Table 1   Examples of Large-Scale, Digital, Experimental Social Science Research

Study	Focus	Context	Experimental procedure	Scale
Product diffusion and engagement				
Present study	Impact of tie strength and embeddedness on social influence	Facebook app adoption	Randomize influence-mediating messages at the dyad level	1.3 M 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	1.5 M users
Aral and Walker (2012)	Social influence and susceptibility in product adoption	Facebook app adoption	Randomize influence-mediating messages at the individual level	1.3 M experimental users, 12 M observational users
Bapna and Umyarov (2012)	Social influence in paid product adoption	Last.fm paid subscription adoption	Randomize gifted paid subscriptions at the individual level	40 K experimental users, 1.2 M observational users
Taylor et al. (2013)	Downstream selection effects in online sharing behavior	Facebook offers	Randomize sharing mechanisms at platform level	1.2 M 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	23 M users, 148 K ads, 101 M user-ad pairs
Tucker (2011)	Social advertising	Unnamed nonprofit	Randomize social ad text and targeting	630 ads, 13 K impressions
Aral and Taylor (2014)	Incentivized peer referral marketing	Online floral delivery site	Randomize incentive structure: Selfish, generous, and fair	637 users
Information sharing and diffusion				
Bakshy et al. (2012b)	Social influence in information diffusion	Facebook news feed	Randomize influence-mediating messages at the individual level	253 M users, 76 M URLs, 1.2 B 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	14 K users
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	116 K users, 101 K comments, 10 M user-comment impression pairs
Tucker and Zhang (2011)	Popularity information and choice	Wedding service vendor website	Randomize availability of popularity information	3 vendor categories, 90 K click-throughs
Health behaviors				
Centola (2010)	Social influence in the spread of health behaviors	Artificial health network	Randomize alters/network structure	1,500 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	61 M 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 and 2,360 students
Bapna et al. (2011)	Reciprocity games	Facebook game	Randomize anonymity	190 users, 77 games
Innovation contest performance				
Boudreau and Lakhani (2011)	Innovation contests	NASA TopCoder contest	Randomize incentives and sorting on competitive regime	
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	10 K experimental users, 100 K observational users

3. Theory

3.1. Social Influence
Understanding how word-of-mouth (WOM) "buzz" about a product, service, opinion, or behavior can impact its adoption has long been considered crucial to how firms can promote diffusion (Arndt 1967, Brown and Reingen 1987, Engel et al. 1969, Godes and Mayzlin 2004, Katz and Lazarsfeld 1955, Manchanda et al. 2008). Traditionally, WOM has been specified as information (often about opinions, preferences, or choices) deliberately exchanged through face-to-face interactions, though more recently the term has been applied to online or technology-enabled information exchange between individuals or from one individual to a group of others (as in the case of consumer product reviews). Researchers in economics and marketing have employed the phrase "observational learning" to refer to circumstances where a consumer observes (usually) aggregated decisions of a population prior to making her selection from a set of alternatives (Banerjee 1992, Bikhchandani et al. 1998, Zhang 2010).

Yet the distinction between WOM and observational learning is not always clear. For example, many online social networks automatically disseminate information about an individual's actions to his immediate peers (e.g., "Brian is reading *A Tale of Two Cities*") in a way that does not indicate aggregate decisions in larger populations or preferences among a set of clear alternatives. Social exchanges such as these straddle the boundary between WOM and observational learning. Notions of customer intent, opinion, preference, and content valence that are typically studied in WOM research may be ambiguous or subjective in this type of information exchange, and observational learning models may not be well suited to situations in which the set of alternatives is very large and information about peer decisions and outcomes are sparsely distributed.

As Godes et al. (2005) highlighted, the conventional definition of WOM is not suitable for a variety of social interactions that mediate information and influence. They instead proposed the more encompassing term *social influence* to describe "an action or actions that is taken by an individual not actively engaged in selling the product or service and that impacts others' expected utility for that product or service" (Godes et al. 2005, pp. 416–417). In the remainder of this section, we describe and expand on the specific notions of the *channel*, *content*, and *impact* of social influence proposed by Godes et al. (2005). We then discuss how these theoretical dimensions of social influence informed and guided the design of our randomized field experiment.

The *channel of social influence* refers to the medium through which influence is communicated or transmitted. Several dimensions specify the channel, such as the number of senders and recipients involved, which

may be one-to-one (as in the case of personal email), one-to-many (as in the cases of email lists, online recommendations, group invitations, and automated peer referrals), many-to-many (as in the case of polls in online community forums), or many-to-one (as in the case of voting on online forum comments). Other salient dimensions include how the recipients are selected, the credibility of the channel, and whether the channel is mediated by a third party. The *content of social influence* refers to the information that is transmitted over the channel. For example, information can include individual decisions or outcomes relating to product features or product adoption, factual information about product features, or subjective opinions about the product as in the case of peer recommendations or customer reviews. Salient dimensions of the content are the subjectivity (fact versus opinion) and whether the content is personalized to the intended recipients. Finally, the *impact of social influence* refers to the overall effect social influence may have on the actions of others. Salient dimensions of impact primarily relate to how it is measured, e.g., whether the impact is inferred or measured directly and what it means to "impact" an outcome. From our perspective, a key dimension of impact is the causal effect of an individual on their peers' behavior. As Aral (2011, p. 217) argued, defining social influence as creating behavior change or "[h]ow the behaviors of one's peers change the likelihood that (or extent to which) one engages in a behavior" is essential to making effective marketing and public policy decisions because effective policy requires an understanding of how behavior is likely to change as a result of an intervention.

The specification of social influence we outline is useful in relating existing research on differing forms of social influence to one another. Moreover, it informs the design of our experiment. We use firm mediation to control the delivery of automated notifications with impersonal content to randomly selected peer targets and assess the impact of social influence by directly measuring peer adoption response. We discuss the ramifications of these design choices in more detail in the experimental design section.

3.2. Influence and Susceptibility
Prior research on social influence has focused on the dual notions of influence and susceptibility (or influenceability). The idea that some individuals are more influential than others and therefore play a catalyzing role in spreading opinions, innovations, and products (Coleman et al. 1957, Gladwell 2002, Rogers 2003, Valente 1995, Van den Bulte and Joshi 2007) is sometimes referred to as "the influentials hypothesis." Other research, focusing on the complementary idea that individual *susceptibility to influence* is the dominant driving mechanism behind diffusion in social networks,

is represented in a variety of theoretical threshold-based contagion models in which behavior adoption occurs when some number or proportion of one's peers have adopted beyond one's intrinsic adoption threshold (e.g., Granovetter 1978, Valente 1996, Watts and Dodds 2007).[2] Though studies estimating the importance of influentials and susceptibles in the diffusion of products or behaviors in real-world networks significantly lag theoretical and simulation models of influence-based contagion, a recent observational study examined the combined notions of influence and susceptibility in social influence processes (Iyengar et al. 2011). More recent work has empirically identified individual influence and susceptibility, demonstrated that both mechanisms together determine the propagation of behaviors in social networks, and also explored *dyadic influence*, in which the influence exerted by an individual on their peer depends on dyadic or pairwise characteristics of both parties (Aral and Walker 2012). We extend this work by examining how characteristics of the relationship between two people moderate the influence they exert on one another. Specifically, we estimate dyadic influence arising from heterogeneity in the embeddedness and tie strength of a relationship while controlling for heterogeneity in individual influence and susceptibility as well as for tendencies toward noninfluenced spontaneous adoption.

3.3. Impact of Social Embeddedness and Tie Strength on Social Influence

3.3.1. Embeddedness. Network embeddedness, or the number of friends that two individuals in a relationship share in common (Easley and Kleinberg 2010, p. 55), has long been theorized to affect the level of trust, altruism, cooperation, and communication in relationships. Embedded relationships are likely to conduct greater social influence because the presence of third-party ties increases the level of trust between embedded peers (Uzzi 1997). As the relationship is "on display" in a social sense, recommendations from embedded peers are likely to be truthful revelations of product experiences or the perceived benefits of the recommended product for the party receiving the recommendation (Granovetter 1985, Uzzi 1996). Embeddedness also engenders greater cooperation because news of noncooperative behavior spreads quickly in the network, making it harder for the uncooperative actor to maintain friendly ties with third parties. In this way, embeddedness enables the development of cooperative norms that facilitate mutual helping relationships (Coleman 1988, Granovetter 1985). Embedded relationships also typically create opportunities

for greater knowledge transfer between individuals (Reagans and McEvily 2003) and more fine-grained information flows (Uzzi 1997) that are multifaceted in that they provide information across multiple topics or dimensions of topics (Aral and Van Alstyne 2011). For example, when discussing a product, two consumers in an embedded relationship may share more information about the product, more knowledge about different dimensions or features of the product, and more fine-grained knowledge of the product, its uses, and its strengths and weaknesses compared to other similar products. Greater trust, cooperation, and fine-grained information exchange is likely to increase the influence conducted in a relationship, and thus we expect embedded relationships to convey greater influence. We adopt the conventional network structural measure of embeddedness, defined in this context as the *number of common friends* shared by individuals and their peers.

3.3.2. Tie Strength. Granovetter (1973, p. 1361) defines tie strength as "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie." He notes that strong ties also typically display multiplexity and the exchange of multiple types of content through the relationship (e.g., information, advice, support; Sundararajan et al. 2013). Strong tie relationships are more likely to conduct influence because they convey greater trust, more fine-grained information exchange, and cooperation (Coleman 1988). Although tie strength is a multidimensional theoretical construct, most studies use measures of the frequency of interaction to proxy for the strength of ties. For example, recent work relating social influence to political mobilization concluded that strong ties were associated with greater social influence, but defined tie strength purely in terms of the frequency of online interactions between peers (Bond et al. 2012). Prior theory has defined tie strength in terms of a number of aggregate measures that include frequency of interaction, surveyed relationship category (such as "friend," "neighbor," "relative," or "acquaintance"), and perceived importance or intimacy (Brown and Reingen 1987, Frenzen and Davis 1990, Granovetter 1983). These extensions of the definition of tie strength beyond interaction frequency are clearly meaningful. At the same time, survey instruments that codify the nature of relationships between individuals and their peers are subject to perception bias and typically do not scale well to large systems. Aggregating the nature of relationship categories into a single measure is also, in some sense, undesirable because it obscures meaningful differences in the type, quality, and context of relationships and reduces our ability to detect the impact of the different dimensions of tie strength on influence.

We expand this conceptualization of tie strength to capture several different dimensions of relationships

---

[2] In this context, a susceptible individual is one with a low intrinsic threshold.

that may be relevant to the strength of social influence. We treat tie strength as a collection of well-defined measures about the historical and current context of the relationships between individuals and their peers and simultaneously examine the distinct impact of these different aspects of tie strength. We define the strength of ties as including (a) the social context of the relationship (how individuals met, know one another, or interact with each other, e.g., whether two Facebook friends attended the *same college*, come from the *same hometown*, or the *number of common institutional affiliations* they share), (b) the recency of the relationship (e.g., whether two Facebook friends currently live in the *same town*), (c) the overlap of common interests (e.g., the *number of common Facebook pages they are "fans" of* or the *number of common Facebook groups* they have joined), and (d) frequency of the interaction (e.g., friends' *copresence in photos* online).[3] We expect greater social affiliation and interaction is predictive of greater influence conducted between friends, whereas similarity in preferences and interests is predictive of correlations in noninfluenced spontaneous adoption, though the theoretical distinctions between different types of tie strength are not yet well developed enough to provide clear theoretical predictions about how they would moderate the degree of influence conducted in a given relationship.

3.3.3. The Endogeneity of Social Structure and Influence. The review of relevant theories of tie strength and embeddedness in the previous two sections immediately calls attention to the endogenous processes that govern them. It is therefore perhaps unsurprising that previous empirical research on embeddedness, tie strength, and social influence has been hampered by endogeneity and spurious correlation. In real-world networks, social embeddedness and tie strength are often correlated with each other and with homophily, making their measurement difficult to untangle in practice (Rogers 2003). Individuals tend to form closer relationships with similar peers, close friends are more likely to share more friends in common, and friendships become stronger with shared common experiences and friends. Nonetheless, tie strength, embeddedness, and homophily can be clearly distinguished theoretically. Although it may be less common, an individual and her peer may share many mutual friends, despite being dissimilar on demographic or personality dimensions. Similarly, some close friends may have nonintersecting peer groups. Endogeneity among these tie characteristics is exacerbated in observational and survey-based studies that do not properly control for selection bias.

Natural social influence processes often involve endogenous communication patterns.[4] Individuals select into sending, receiving, or soliciting influence-mediating communications to or from their peers. As a consequence, studies of social influence, embeddedness, and tie strength that do not explicitly control for selection biases in communication patterns confound our understanding of how tie characteristics moderate influence. For example, some studies on social influence and tie characteristics report that dissimilarity between peers is correlated with increased influence (e.g., Gilly et al. 1998). In contrast, other studies contend that more homophilous ties, though more likely to be activated, are not associated with any more (or less) influence (Brown and Reingen 1987). Studies have also examined the role of embeddedness in social influence. For example, recent work on influence in the decision to join Facebook in response to an email invitation indicates that less embedded ties are associated with greater influence, leading to a higher probability of positive response (Ugander et al. 2012). None of these studies account or control for selective communication. Burt (2005) discusses another potential confounding factor: weak ties are more likely to transmit novel information, by virtue of being less socially embedded.[5] These considerations highlight that inferring the impact of tie strength and social embeddedness on influence is difficult because influence-mediating communications are inherently endogenous. One notable exception is provided by the work of De Bruyn and Lilien (2008), who disentangled selection bias in communication tendencies by explicitly modeling multiple stages of interaction in influence processes in an experimental study of word-of-mouth viral marketing. However, this approach relies on the ability of researchers to correctly model the stages of interaction in social influence processes and does not generalize well to contexts in which we lack intuition about which social processes are at work.

---

[3] One could interpret these variables as indicators of homophily rather than tie strength. However, we reserve the discussion of whether homophily explains variation in behaviors to the interpretation of the parameter estimates on spontaneous adoption. Although selection into common experiences could be driven by preference similarity, such common experiences also typically strengthen ties. Our intention is to distinguish variance in correlated peer behaviors explained by influence from variance explained by homophily and confounding factors. From such a perspective, common experience that creates influence is an example of the impact of tie strength on influence, whereas common experience that does not create influence but explains correlated peer behaviors is an example of preference similarities (peers with common experiences making similar choices in the absence of influence). We leverage randomization of influence-mediating messages to make these empirical distinctions.

[4] Communication patterns may include information seeking behavior (such as solicitation of peer advice), passive subscriptions to information sources (such as blogs, twitter feeds, or Facebook news feeds), or active forwarding of information to peers (such as personalized referrals or invitations).

[5] In the Godes et al. (2005) framework, this corresponds to selection bias in both the channel and content of social influence.

Our study design disentangles the impact of both the frequency of interaction and the novelty of information exchanged from other aspects of tie strength that characterize the nature of the relationship between individuals and their peers by (a) controlling the channel of influence (through randomized recipient selection) and (b) holding message content constant. In our design, influence-mediating communication is controlled by a third party, allowing us to randomize message target selection and to homogenize content. Messages sent from individuals to their peers contain approximately the same information, allowing us to study the impact of embeddedness and tie strength holding information diversity or novelty constant. Message target randomization also ensures that the number (frequency) of influence-mediating messages sent from individuals to their peers is independent of embeddedness and tie strength.

Communication patterns between individuals and their peers certainly play a role in influence and in part comprise what makes individuals influential on their peers. However, understanding the impact of social embeddedness and tie strength on influence, holding communication patterns constant is important for two reasons. First, it contributes to our understanding of how recipients of an influence-mediating message would respond differently to more or less embedded peers and peers with whom they are strongly or weakly connected. Such insights are critical to viral marketing initiatives designed to target advertisements at those most likely to maximize the diffusion products and services in a population through their natural or intrinsic influence (Aral et al. 2009, 2013). Second, it can inform policies that operate outside of the scope of natural influence (such as individual and network-based interventions and peer-oriented incentive schemes), which are deliberately designed to impact and alter natural communication and information flow patterns (Aral and Taylor 2014).

4. Empirical Methods

We partnered with a firm that develops commercial applications hosted on the popular social networking website Facebook.com. A commercial Facebook application was designed and publically released in concert with the launch of the experiment. This application provides users the opportunity to share information and opinions about movies, actors, directors, and the film industry in general. As adopters used the application, automated notifications were delivered to randomly selected peers in their local social networks. Data on individual attributes of adopters and their local peers, time-stamped delivery of automated notifications, and subsequent time-stamped adoption responses were collected throughout the course of the experiment.

The experimental design employs *message target randomization* to deliver automated notifications[6] to randomly selected peers of existing application users. In this scheme, packets of notifications are generated when application users take one of several actions within the Facebook application (e.g., when the user rates a movie or friends a celebrity). These notifications are then delivered to randomly chosen subsets of the application user's Facebook friends. The random selection of a set of recipient peers is performed on a per-packet basis (i.e., a different set of recipient peers is randomly chosen each time a packet is sent from an application user). This design is illustrated in Figure 1, which displays a diagram of the delivery of two packets over sequential time periods.

At time $t1$, the application user performs a packet-generating action within the application and a packet of notifications is generated. At time $t2$, randomly chosen peers of the application user are designated as recipients and receive the notifications. At time $t3$, the application user performs a second packet-generating action within the application, and a second packet of notifications is generated and delivered at time $t4$ to a (different) set of randomly chosen peer recipients. At any given time throughout the course of the experiment, peers of an application user received 0, 1, 2, or more influence-mediating messages from their application user friend. The exposure of a peer to influence-mediating messages (notifications) over time is exogenously determined as a function of the randomization procedure. Over time, peers are assigned to risk groups (corresponding to the number of influence-mediating messages received), where risk is monotonically and randomly increasing over the course of the experiment. We discuss the implication of message target randomization on our modeling strategy and censoring procedures in the section that follows.

Automated notifications (passive viral messages) have several advantages over alternate types of influence-mediating communication for the purpose of our experimental design (Aral and Walker 2011a, b, 2012). First, the automated notifications channel is mediated by a third party, allowing the desired level of experimental control and randomization of peer targets. Since the recipients of notifications and the decisions of whether and when to send them are all automated, selection effects on the part of the sender that might otherwise introduce bias can be avoided.

Second, the content of automated notifications employed in the experiment included only impersonal information about the sender's use of the application

---

[6] Notifications are a form of messaging on Facebook that appear in a designated Notifications inbox on the recipient's Facebook menu and are not integrated into other areas, such as Facebook news feeds, and as such are not subject to ranking or filtering mechanisms.

Figure 1   Randomized Targeting of Influence-Mediating Messages
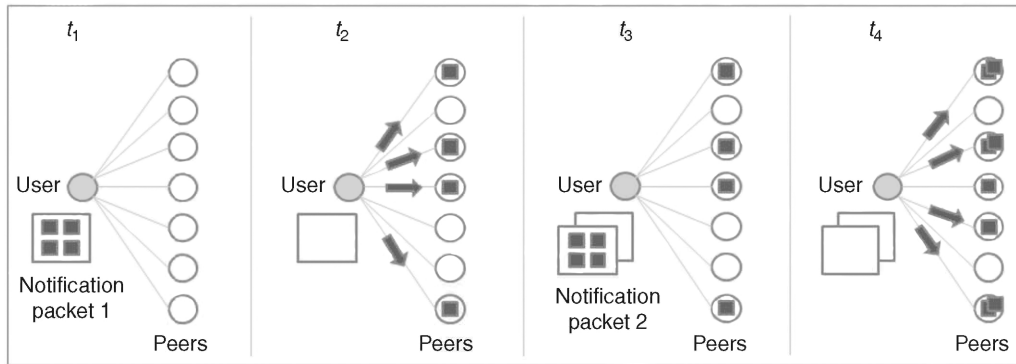
Notes. A diagram depicting the message target randomization employed in the experiment is shown. Notification packets are generated when an application user takes a packet-generating action within the Facebook application. For each packet that is generated, the notifications in the packet are distributed to a randomly selected subset of the application user's peers. The figure displays two sequential packet distributions. Different recipient targets are randomly chosen at the time of distribution for each packet.

(as is typical of information exchange in observational learning scenarios). The inclusion of only impersonal information in influence-mediating messages allows for the measurement of the impact of social influence while holding constant the potentially large degree of heterogeneity present in personalized, sender-created content. Heterogeneity in the content of messages created by individuals is known to have a significant effect on social influence. In particular, the effects of content heterogeneity on influence have been studied in the context of the positive valence of the message content (Berger and Milkman 2009) and the effectiveness of viral features that allow message tailoring or personalization (Aral and Walker 2011a). Although content heterogeneity can and in all likelihood does play a major role in what makes individuals influential, simultaneous variation of content and relationship attributes can confound measurements of the effect of relationship attributes on social influence.

Third, the delivery of notifications to only a random subset of an individual's peers permits direct comparison of the responses of treated peers to those of peers of the same application user that were not treated. When the targets of potentially influential communications are randomized among peers of the same application user, any homophilous structure between an application user and his treated and untreated peers and the propensity to select a particular peer to notify are held constant and are identical for recipient and nonrecipient peer groups. Other unobserved factors that could potentially drive influenced adoption, such as off line or alternative online communications, can also be cleanly distinguished with this design, because recipient and nonrecipient peers in expectation share similar propensities to receive and be affected by such communications on average. Moreover, homophily in unobserved attributes (latent homophily) that may be indicated by the very existence of a relationship of

peers with a common friend (see Shalizi and Thomas 2011) will be equally represented in recipient and nonrecipient peer groups. Differences in adoption outcomes between recipient and nonrecipient peers can then be attributed solely to the influence-mediating messages they received.

5.   Analysis and Results

5.1.   Data and Descriptive Statistics

Throughout the 44-day experimental period, beginning on November 28, 2009, we collected individual-level profile data from 7,730 application users and their 1.3 million distinct peers as well as time-stamped click stream data on notification delivery and subsequent peer responses. During this time, 41,686 automated notifications were delivered to randomly chosen peer targets of application users, resulting in 967 peer adoptions, a 13% increase in product adoption. The user data we collected included the social network of adopters, all mutual ties between peers, and individual-level profile data including *age, gender, relationship status, hometown, current town, college attendance, affiliations, Facebook pages, Facebook group membership,* and *tagged appearance in photos.*[7] Descriptive statistics are provided in Tables A1 and A2 of the online technical appendix (available at http://web.mit.edu/sinana/www/MSTA14.pdf).

5.2.   Model Specification

Following prior work on influence identification and social contagion, we adopt a hazard modeling approach

---

[7] Facebook affiliations typically indicate some past experience such as working for a company or belonging to the same institution or society. Facebook pages typically indicate interest in activities, brands, products, bands, and media personalities. Facebook groups are an online version of social groups that allow users to interact with one another in centralized discussion forums and view news and events pertaining to that group.

(Aral and Walker 2011a, 2012; Iyengar et al. 2011; Nam et al. 2010; Van den Bulte and Lilien 2001). We employ a Cox proportional hazard model to estimate the hazard of a peer (of an existing application user) adopting. The model simultaneously estimates the impact of social embeddedness and multiple measures of tie strength on influenced and spontaneous adoption while controlling for the moderating effect of individual attributes on influenced and spontaneous adoption:

$$\lambda_j(i,j,t)$$
$$= \lambda_0(t)\exp\{\beta_N N_j(t) + \beta_{Spont}^i X_i + \beta_{Spont}^j X_j + \beta_{Pref}^{SoT} SoT_{ij}$$
$$+ \beta_{Pref}^{Embed} Embed_{ij} + \beta_{Infl}^i X_i N_j(t) + \beta_{Susc}^j X_j N_j(t)$$
$$+ \beta_{Infl}^{SoT} N_j(t) SoT_{ij} + \beta_{Infl}^{Embed} N_j(t) Embed_{ij}\},$$

where λ_j is the hazard for a peer j of a user i adopting, $N_j(t)$ is the number of notifications received by a peer j, $Embed_{ij}$ is the embeddedness of the relationship between user i and peer j (the number of friends shared by individual i and peer j), and SoT_{ij} is a vector of the tie strength attributes characterizing the relationship between individual i and peer j. These models include a rich set of covariate controls for demographic and individual-level characteristics of individuals (X_i) and their peers (X_j) (including *age, gender,* and *relationship status*). The coefficient β_N captures the raw impact of influence, holding constant heterogeneity in individual or tie attributes—it represents the marginal impact of a peer receiving influence-mediating messages irrespective of the individual attributes of individual i and peer j or their dyadic tie attributes. The coefficient β_{Spont}^i captures the tendency of peers of application users with attributes X_i to spontaneously adopt in the absence of influence ($N_j = 0$). The coefficient β_{Spont}^j captures the tendency of peers with own attributes X_j to spontaneously adopt in the absence of influence ($N_j = 0$). The coefficient β_{Pref}^{SoT} captures the extent to which the measure of tie strength indicates a similarity in preference for a peer adopting the product spontaneously in the absence of influence ($N_j = 0$) given that her application user friend has adopted. This is the variation in correlated peer behaviors explained by homophily along tie strength features. The coefficient Similarly, β_{Pref}^{Embed} captures the extent to which peers with embedded ties to existing application users will have a preference to adopt the product spontaneously in the absence of influence ($N_j = 0$). This is the variation in correlated peer behaviors explained by homophily along the embeddedness features. The coefficient β_{Infl}^i represents individual influence—it captures the impact of individual attributes of an application user (X_i) on the hazard of her peer adopting due to influence per influence-mediating message received. The coefficient β_{Susc}^j represents individual susceptibility to influence—it captures the impact of peer attributes (X_j) on peers' haz-

ard to adopt due to influence per influence-mediating message received. The coefficient β_{Infl}^{SoT} captures the impact of the tie strength measure (for the tie between application user i and peer j) on influence-driven adoption per influence-mediating message received. The coefficient β_{Infl}^{Embed} captures the impact of the embeddedness of the tie between application user i and peer j on influence-driven adoption by j per influence-mediating message received. The model estimation provides good concordance with observed data, and Wald, log rank, and likelihood ratio test statistics indicate strong likelihood, significance, and goodness of fit. (See Table A3 in the online technical appendix.)

Because peers of application users randomly receive (multiple) notifications from their application user friends throughout the course of the experiment, we employed interval censoring to transition users from one risk group (e.g., the hazard associated with receiving one influence-mediating notification) to the next (e.g., the hazard associated with receiving two influence-mediating notifications). Peer adoption outcomes may be correlated because peers of a given application user share a common application user friend. To account for this, the hazard model employs robust errors clustered on the identity of the application user friend.

5.3. Results
Our model specification enables us to estimate two quantities of theoretical interest: influence-based adoption and spontaneous adoption. Influence-based adoption measures the degree to which a particular relationship characteristic moderates the influence an individual has over their peers or the degree to which that characteristic is associated with changing someone's behavior from not adopting to adopting the application. Spontaneous adoption, on the other hand, measures the degree to which a particular relationship characteristic predicts a correlated latent preference to adopt the product. For example, if having attended the same college is associated with spontaneous adoption but not influence-based adoption, then attending the same college is capturing preference similarities that predict adoption by a peer of a current adopter. If, on the other hand, having attended the same college is associated with influence-based adoption but not spontaneous adoption, then individuals influence those friends who attended the same college as they did more than those friends who attended different colleges than they did.

These distinctions are critical to marketing policy and networked interventions in general. Predictors of spontaneous adoption can inform targeting: They generate good sets of advertising targets whose preference similarities to current adopters make them likely to respond positively to advertisements about the product under consideration. On the other hand, moderators of influence can inform peer referral strategies: they highlight

good sets of relationship pairs in which incentives to propagate social influence may work well to create additional product adoptions (Aral and Taylor 2014). For example, incentives to "invite friends" to products or discounts for friend and family adoption may work well in relationship pairs that conduct influence. Model estimations of the impact of relationship characteristics on influence-based and spontaneous adoption are tabulated in Table 2. (For full model estimations, see Table A3 in the online technical appendix.)

Table 2   Impact of Embeddedness and SoT on Influence

	Influence	Spontaneous
	Hazard ratio (SE)	Hazard ratio (SE)
	β_{Infl}^{Embed}	β_{Pref}^{Embed}
Num. common friends	1.0063***	1.0077***
	(0.0016)	(0.0008)
	β_{Infl}^{SoT}	β_{Pref}^{SoT}
Hometown (same)	1.0171	1.4724
	(0.2272)	(0.2568)
Hometown (different)	1.3735*	0.7187
	(0.1684)	(0.2361)
Current town (same)	2.2899***	0.4686
	(0.3094)	(0.7221)
Current town (different)	0.3171*	1.9300*
	(0.6699)	(0.3418)
College (same)	8.5540**	0.3646
	(0.8389)	(1.1272)
College (different)	0.5878	0.7664
	(0.4956)	(0.3529)
Num. common affiliations	2.2548**	0.8184
	(0.3740)	(0.3288)
Num. common pages	1.0031	1.0067***
	(0.0023)	(0.0010)
Num. common groups	1.0074	1.0335***
	(0.0057)	(0.0044)
Num. photos together	0.9977	1.0142***
	(0.0031)	(0.0018)

Notes. This table reports parameter estimates and standard errors from the single failure proportional hazards model. Variables reported include categorical dummy variables indicating the following: *Hometown* (*same, different*), whether the individual and peer come from the same or different hometowns (unreported hometown corresponds to the holdout); *Current town* (*same, different*), whether the individual and peer live in the same or different current towns (unreported current town corresponds to the holdout); *College* (*same, different*), whether the individual and peer attended the same or different colleges (unreported college corresponds to the holdout); *Num. common pages*, the number of Facebook pages shared in common between the individual and their peer; *Num. common groups*, the number of Facebook groups shared in common between the individual and their peer; *Num. photos together*, the number of photos in which both the individual and peer appear. Hazard ratios in the influence column correspond to variables crossed with N_j (the number of notifications received by the peer) and indicate the effect of attribute-driven adoption. Hazard ratios in the spontaneous column correspond to uncrossed variables and represent spontaneous or preference-related adoption.

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

5.3.1.   Effects of Social Embeddedness and Tie Strength on Influence. Results illustrating the impact of social embeddedness and tie strength on influence are displayed in Figure 2. The forest plot displays the hazard ratios, standard errors (boxes), and 95% confidence intervals (whiskers) of influence-driven adoption associated with embeddedness (number of common friends) and tie strength attributes. The hazard ratios displayed are relative to the baseline case (a hypothetical blank social network profile) and all categorical dummy variables such as *Same college* and *Diff. college* catalogue all categories for which the measure is defined. The holdout sets are cases for which either the application user or peer has not reported the measure in their social network profile. For example, the three exhaustive categories for college affiliation relations and their associated encoding are (1) the peers went to the same college (*Same college* = 1; *Diff. college* = 0), (2) the peers went to different colleges (*Same college* = 0; *Diff. college* = 1), and (3) one or both peers do not report their college attendance (*Same college* = 0; *Diff. college* = 0).

We observe several interesting patterns in the results detailing the impact of embeddedness and different measures of tie strength on influence. First, tie measures that capture peers' joint participation in common social or institutional contexts between individuals and their Facebook friends are associated with greater influence. Individuals exert 125% more influence on friends for each institutional affiliation they share in common ($p < 0.05$). Attending the same college as one's friend is associated with a 1,355% increase in influence ($p < 0.01$) compared to attending different colleges. This represents the largest impact on influence of the categorical measures of tie strength we considered. In contrast, coming from the same hometown is not significantly associated with influence, perhaps suggesting that this measure accurately captures more casual or even incidental social contexts (e.g., weak ties resulting from Facebook users' desires to keep in contact with casual acquaintances).

Second, tie strength measures associated with current or recent social contexts exhibit differing impacts on influence. Individuals exhibit 622% more influence on friends that live in the same current town ($p < 0.01$). This is interesting because ties between friends currently living in the same town may indicate joint involvement in more recent social contexts (e.g., friendships that are more recent or recently relevant). Interestingly, appearing in photos with peers, an indicator of offline interaction at in-person events, is not significantly associated with influence.

Third, tie strength measures associated with common interests or preferences do not moderate influence. Individuals are no more or less influential on peers

Figure 2   Influence Associated with Embeddedness and Tie Strength
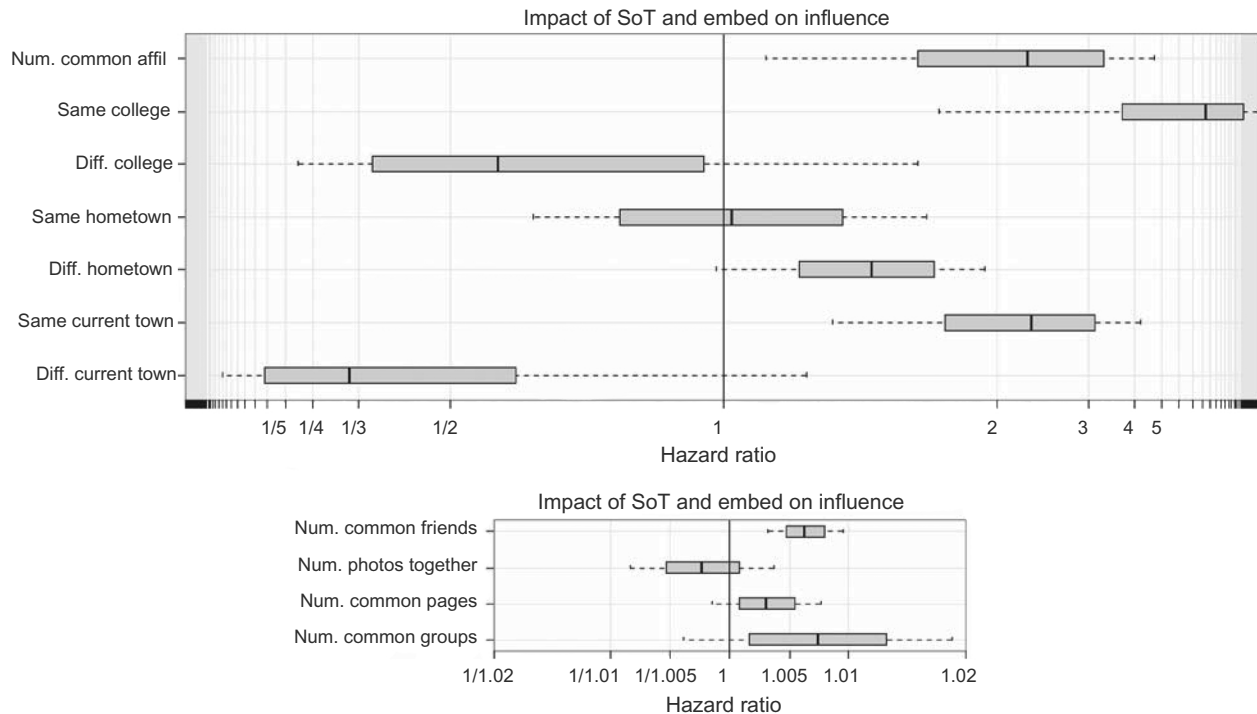
Notes. The effects of tie strength and embeddedness and tie strength measures on influenced adoption as shown. The figure displays hazard ratios (HRs) representing the percentage increase (HR > 1) or decrease (HR < 1) in adoption hazards associated with each tie attribute. Boxes are standard errors. Whiskers are 95% confidence intervals.

with whom they share common Facebook pages or are comembers of online groups.

Finally, individuals are more influential on peers with whom they are more embedded, exhibiting a (multiplicative) 0.6% increase in influence for each friend they share in common ($p < 0.001$). The impact of embeddedness on influence, though comparatively smaller than the tie strength measures considered here, remains economically significant, as the number of common friends can be quite large. For example, for a friendship that shares 10 friends in common, the result implies over a 6.5% increase in influence.[8]

These results indicate the importance of social embeddedness in influence processes and the subtlety in the relationship between influence and measures of tie strength. They also highlight the importance of using disaggregated measures of tie strength. It should be noted that although our experimental design and analysis allow for identification of the *causal* impact of peer influence, we estimate how relationship-level covariates are *correlated* with the extent or impact of influence. This is an important distinction from a policy standpoint, because our results do not indicate that

exogenous changes to relationship-level covariates will yield a change in influence or influenceability corresponding to our estimates. For example, our findings *do not* indicate that exogenously inducing an individual to move to the same current town as a peer will lead to a sixfold increase in influence over that peer. Our findings are robust to the inclusion of multiple different factors that control for heterogeneity in network degree, Facebook profile completeness, Facebook activity levels, nonlinearity in the response to influence-mediating messages, and stratification over the number of influence-mediating messages sent by application users. (See the online technical appendix for more details on the tests of model robustness.)

5.3.2.   Social Embeddedness and Tie Strength as Predictors of Spontaneous Adoption. Results on the correlation between spontaneous (preference) adoption and tie characteristics are displayed in Figure 3.

Tie characteristics (tie strength and embeddedness) associated with spontaneous adoption of the product by peers of existing adopters indicate preference similarity between peers—the extent to which the measure captures similarity in the (latent) preference to adopt the product when a friend has already adopted. Some tie strength measures that seem to relate to common social contexts are good predictors of preference similarity, whereas others are not. Sharing common affiliations or

---

[8] Using the exact estimate of the influence associated with common friends of 1.0063 (taken from Table 2), the example of 10 friends in common leads to an increase in influence of $(1.0063)^{10} = 1.065$, or a 6.5% increase in influence.

Figure 3   Spontaneous (Preference) Adoption Associated with Embeddedness and Tie Strength
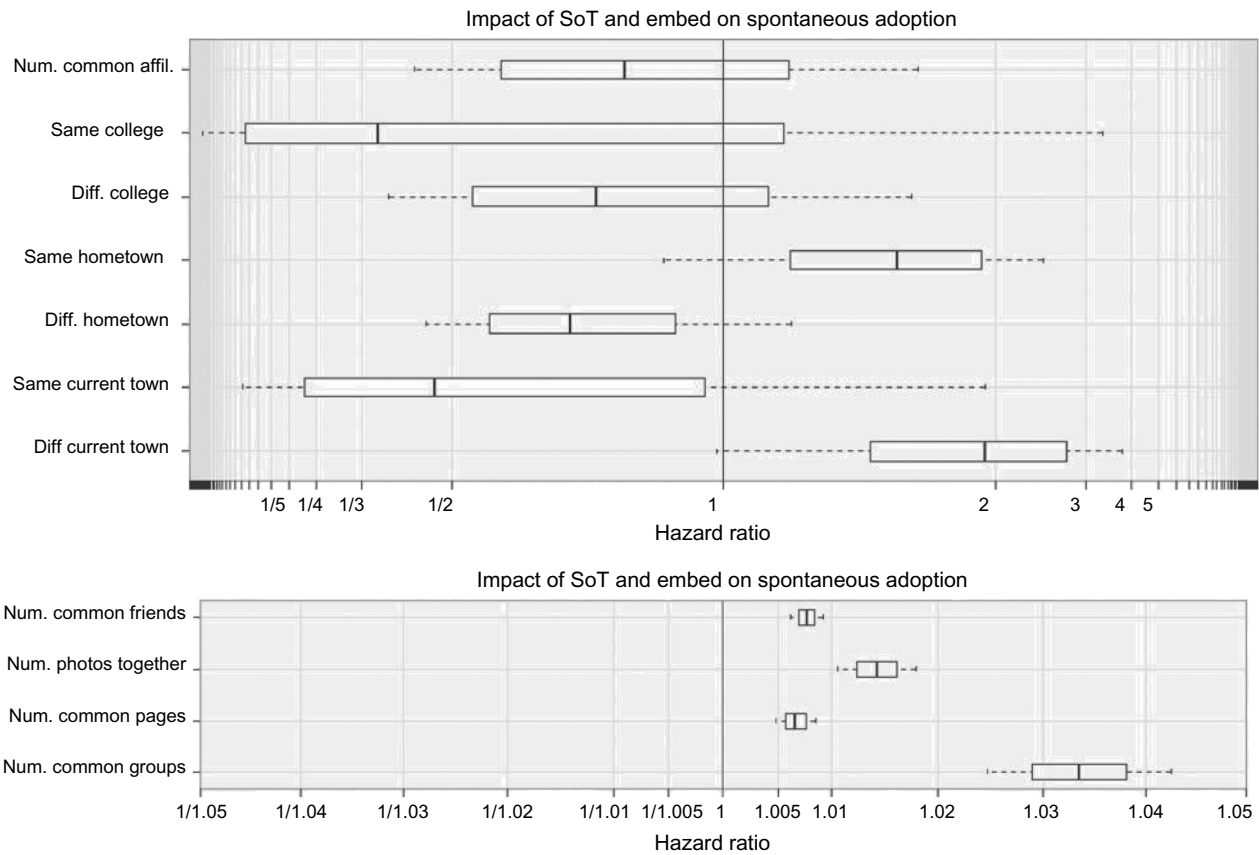
Notes. The effects of tie strength and embeddedness and tie strength measures on spontaneous (preference) adoption are shown. The figure displays hazard ratios (HRs) representing the percentage increase (HR > 1) or decrease (HR < 1) in spontaneous adoption hazards associated with each tie attribute. Boxes are standard errors. Whiskers are 95% confidence intervals.

attending the same college as an adopter of the product is not significantly associated with a tendency to spontaneously adopt. However, each photo that a peer shares in common with an adopter is associated with a 1.4% increase in the hazard to adopt spontaneously ($p < 0.001$). Coming from the same hometown as a peer who has adopted the product is associated with a 105% increase in the hazard to adopt spontaneously ($p < 0.05$). This indicates that hometown may be a good (latent) proxy for individual preferences for the product. One explanation for this pattern in the results could be that current friends influence us more, but that our preference-driven behaviors are more correlated with past, nonrecent social contexts. In other words, we are more influenced by friends in the same current town, but our preferences are more correlated with friends from the hometown in which we grew up and with friends that currently do not live in the same town.

Each additional fan page that a peer shares in common with an adopter of the product is associated with a 0.7% increase in the hazard to adopt spontaneously ($p < 0.001$). This indicates that declared preferences and interests (not directly related to the product) also capture preferences for the product. Each online group

that a peer participates in with an adopter of the product is associated with a 3.4% increase in the hazard to spontaneously adopt ($P < 0.001$). This indicates that online social activities capture latent dimensions of preference for the product. We note that the above interpretation of spontaneous adoption estimates are subject to the caveat that influence through offline channels unrelated to the explicit channel of automated notifications that we focus on here may exist, though prior research suggests such effects are low because active offline channels are used much less often because of the increased effort required compared to automated (passive) messaging (Aral and Walker 2011a). In addition, researchers should use caution in interpreting how tie attributes moderate influence, because some unobserved individual attributes may be correlated with the tendency to form particular types of ties. For example, the average number of common photos that an individual appears in with their peers may be correlated with extroversion.

6.   Robustness
Though our estimates are based on a large-sample in vivo randomized experiment, which provides a certain

degree of assurance against omitted variables and questions about causal interpretations of our results, we conducted multiple tests of the robustness of our parameter estimates to ensure the reliability of our findings.

To assess possible multicollinearity among model covariates, variance inflation factors were calculated for all time-independent covariates in the model. Results are presented in Table A4 in the online technical appendix. All variance inflation factors are well below the conventionally accepted threshold (VIF < 5), indicating the multicollinearity is not an issue in our model and does not significantly impact our findings.

To assess to the extent to which our estimates of the moderating impact of tie strength on influence depend upon embeddedness, the model was estimated with our operational definition of embeddedness (the number of common friends) omitted from the model. The results, presented in Table A5 in the online technical appendix, indicate that our findings on the moderating impact of tie strength on influence generally hold when embeddedness (number of common friends) is excluded from the model.

To assess whether our main findings were affected by variation in the network degree of the adopting user i, we estimated two models, one with an additional control for network degree, shown in Table A6a in the online technical appendix, and one with both the standalone degree variable (network degree of i) and the interaction term (number of notifications $*$ network degree of i), shown in Table A6b in the online technical appendix. We find that the network degree of user i is not significant and does not significantly alter the remaining estimates, indicating that our findings are robust to variation in network degree of the adopting user.

We find the standalone control for network degree of user i is only marginally significant, whereas the interaction control for network degree of user i and the number of notifications is significant but similar in magnitude and opposite in sign to that of the standalone control term, indicating that the impact of the two are both small and tend to cancel one another in cases when influence is present. Importantly, the inclusion of both standalone and interaction terms with network degree does not significantly alter the remaining estimates, indicating that our findings are robust to variation in the network degree of the adopting user.

Variation in profile completeness within Facebook can be extensive in terms of the number of photos, groups, pages, and affiliations indicated by individuals and their peers. To establish that our results on the moderating impact of tie strength represented by common photos, groups, pages, and affiliations are robust to controls for profile completeness, we estimated two models, one with standalone profile completeness controls, shown in Table A7a in the online

technical appendix, and one with both standalone (profile completeness term) and interaction (number of notifications $*$ profile completeness) controls, shown in Table A7b in the online technical appendix. We find that our main findings are generally robust to inclusion of controls for both standalone and interacted profile completeness for individuals and their peers.

The message target randomization scheme we employed delivers automated notifications from an individual to randomly selected subsets of their peers. Since the generation of notifications is contingent on user activity within the application (according to limitations placed on Facebook application messaging to limit spam) and because this activity may be correlated with other (potentially latent) attributes, we assess the robustness of our findings by estimating a model with a control for the number of messages sent in Table A8 in the online technical appendix.

Our findings are generally robust to the inclusion of an additional control for the number of notifications sent. The reduction in the primary influence effect (number of notifications received) is expected and is a consequence of the parceling out of the main influence effect across the actual number of notifications received by a peer and the average probability to receive a notification for a peer j of a user i, as embodied by the number of notifications sent by user i (which is related to the average probability to receive notifications through a constant of proportionality as the inverse of the degree of user i).

However, because the control for number of notifications sent is significant in the above estimation, we performed two additional robustness checks by estimating stratified models with stratification over terms related to the number of notifications sent. These specifications test whether our findings are a result of correlation between the number of notifications sent and some latent variable (X) that itself is both correlated with either interest in the application or a tendency to spontaneously adopt in the absence of influence and is homophilously distributed among application users and their peers.

A model stratified by whether the average notifications sent by a user are greater than average (nns_greater_than_average) with a control for the number of notifications sent is displayed in Table A9a in the online technical appendix. The model shows that influence and the modulation of influence by tie attributes remain strong and highly significant, and our estimations do not change substantially relative to our original model, indicating little benefit to stratification. Moreover, the number of notifications sent is no longer significant after stratification, indicating that the residual impact of the number of notifications sent cannot be attributed to a latent variable correlated with

an interest or tendency to adopt spontaneously in the absence of influence.

The conclusion is similarly supported by estimation of a second model stratified by the number of notification sent, displayed in Table A9b in the online technical appendix. This second model shows that influence and the modulation of influence by tie attributes remain strong and highly significant, and our estimations do not change substantially relative to our original model, indicating little benefit to stratification. Both stratified models have a substantially worse fit to the empirical data (in terms of measures of likelihood and concordance), are less parsimonious, allow for a large amount of unspecified heterogeneity (arising from the nonspecification of the large number of baseline hazards for each strata), and hinder interpretation of covariate estimates, relative to the main model presented in this paper. We therefore retain these models as secondary robustness checks.

The moderating impact of tie strength and embeddedness (the primary focus of this work) is best assessed through the main model presented in the results section of this paper. In this model, the impact of receiving one or more notifications is modeled with a response term (in the proportional hazard model) that is linear in the number of notifications received. However, in theory we expect the marginal response of a peer receiving an additional notification to decay nonlinearly as the number of notifications increases beyond a threshold.

Martingale residuals, which assess the extent to which the number of notifications received departs from linearity, are displayed in Figure A3 in the online technical appendix. The martingale residuals plot indicates relatively linear response (approximately zero slope), with a minor departure from linearity that begins for peers receiving four or more notifications. Peers receiving four or more notifications are relatively infrequent in our data (only 40 out of 1.3 million peers received five or more notifications; less than 300 peers received four or more notifications; less than 1,000 peers received more than two notifications). Nonetheless, we assess the departure from linearity of the main influence effect by estimating our original model with a term that is quadratic in the number of notifications received in Table A10 in the online technical appendix.

We do find significance for the quadratic term in number of notifications received, consistent with the reasoning outlined above. It should be noted that inclusion of a term that is quadratic in number of notifications received makes interpretation of the estimates and confidence intervals of moderating tie strength and embeddedness factors less clear because quadratic interaction terms need also be included and their estimates combined to correctly assess moderating factor effect sizes, significance, and confidence intervals, making

interpretation and comparison of the impact of embeddedness and different tie strength measures difficult. In this sense our original model is more suitable for assessing the moderating impact of tie strength and embeddedness on influence.

Finally, we minimize the effects of interference in our experiment by recruiting subjects from a large sparse graph structure, implementing a recruitment campaign that minimizes the connections between recruited subjects, analyzing only local peer effects from a recruited user to their peer, and right-censoring observations with multiple treated peers to parameterize our ignorance of the effects of multiple treated peers. There are multiple approaches to interference in networked experiments. We discuss the strengths and weaknesses of these approaches in the next section. Our approach, which represents a design approach to minimizing interference, maintains a parsimonious view of peer effects and reduces the number of assumptions we have to make to perform inference on the estimands we care about. One consequence of such an approach is that although our estimates robustly generalize to local peers effects (estimates of the effect of influence on direct peers), they may not generalize to describe the full complexity of the dynamic spread of social influence in the network (e.g., including direct and first-, second-, and third-order—and so on—indirect influence effects cascading throughout the network).

Although our experimental design strategy allows for causal identification of the moderating impact of tie strength and embeddedness on peer influence, minimizes the effects of interference, and circumvents endogeneity in online communication patterns, it is not without its limitations. First, our results may have limited generalizability to other contexts, for example, to cases where there is a significant financial cost to adopting a product or service or for populations of inactive (or rarely active) users.[9] Second, the use of firm-mediated automated (passive) peer-to-peer messages, although advantageous in their ability to break endogenous online communication patterns, have been found to be less effective in inducing adoption on a per-message basis than active personalized messaging features and, when overused, may be regarded as spammy or may even be blocked (contingent on platform-specific options) on a per-user basis. For an in-depth comparison of passive and active messaging, see Aral and Walker (2011a). Third, some forms of endogeneity may still be present and uncontrolled for, such as offline word-of-mouth communications

---

[9] The spread of WOM by nonadopters or nonusers of a product may arise, for example, from preconceived notions concerning a products utility or by mimicking WOM heard from others. Speculative, mimicked, and spurious WOM is an interesting potential topic for future inquiry.

between peers. These other endogenous effects may impact the interpretation of our spontaneous adoption results. Though prior research suggests that offline word of mouth (that necessarily requires conscious effort) is significantly less likely to impact adoption in our specific context (Aral and Walker 2011a), future research into offline influence mechanisms and the comparison of online to offline influence is encouraged. Future work should also address additional contexts, where the correlation between social structural measures and influence may differ. In addition, focus on alternative social measures, such as structural equivalence and brokerage, may yield meaningful insights into the dynamics of social influence.

7. Discussion: The Future of Large-Scale Randomized Experiments in Networks

Although recent studies have powerfully demonstrated the scientific potential and empirical effectiveness of a new paradigm of business analytics based on large-scale digital experimentation, many challenges remain in perfecting the science of these dynamic, often networked experiments. First, choosing the sample frame and designing the sampling and/or recruitment strategy for networked experiments is essential to avoiding selection bias. Recent work has demonstrated that some samples based on network propagation methods, such as respondent driven sampling, can create biased inference because they systematically miss certain types of edges and nodes (Airoldi et al. 2011). These studies also conclude, however, that comprehensive propagation sampling, which samples all edges and peers in each step of the snowball propagation (e.g., Aral et al. 2009), are not subject to these biases because they collect complete local network information. Furthermore, selection effects can create bias when recruited subjects do not represent the population of network nodes they intend to represent. Recruitment campaigns must therefore be specifically designed to avoid recruitment selection bias (Aral and Walker 2011a). These sampling and recruitment choices serve as informative examples of the types of biases that can be created by well-known and widely used sampling methods. Researchers must therefore carefully attend to sampling choices in the design of networked experiments and the inference techniques they use to evaluate experimental evidence.

Second, the potential for interference, leakage, or contamination in networked experiments is well documented (Aral and Walker 2011a, Aronow and Samii 2011). Since treated nodes are connected to nontreated nodes through complex network structure, even randomly assigned treatments can violate the stable unit treatment value assumption (SUTVA) and interfere with one another to create systematic biases. As a

starting point, the likelihood of interference and its effects on bias and variance can be empirically estimated (e.g., Muchnik et al. 2013). Several solutions to interference in networked experiments are also currently being developed along two primary dimensions: design strategies and inference strategies.

Design strategies attempt to set up network experiments so as to minimize the likelihood of interference. For example, as we do in this paper, prior work has recruited subjects from large sparse graph structures and analyzed only local peer effects from a recruited user to their peer to minimize the potential for interference across treatments in local network neighborhoods (e.g., Aral and Walker 2011a, 2012). More recently, clustered graph randomization strategies have been developed to reduce the likelihood of interference. Ugander et al. (2013), for example, propose graph clustering to analyze average treatment effects under social interference and develop an efficient exact algorithm to compute the probabilities for each node's network exposure. Using these probabilities as inverse weights, Ugander et al. (2013) demonstrate that a Horvitz–Thompson estimator can provide an effect estimate that is unbiased, provided that the exposure model has been properly specified. They show that proper cluster randomization can lead to exponentially lower estimator variance when experimentally measuring average treatment effects under interference.

Inference strategies, on the other hand, attempt to correct interference bias at the point of inference (Aral and Walker 2011a, Aronow and Samii 2011, Middleton and Aronow 2011). One approach has been to right-censor contaminated nodes at the point in time after which, during a networked experiment, they are likely to be exposed to interference (e.g., Aral and Walker 2011a). This approach parameterizes the ignorance of the researcher with regard to the true exposure model by limiting inference to conservative effects that are likely to be protected from interference. Other inference strategies, however, begin by hypothesizing an exposure model, and then base inference techniques on estimands that are unbiased if the exposure model is accurately specified (e.g., Aronow and Samii 2011, Middleton and Aronow 2011). These techniques assume the form of interference is known and provide conservative estimators of the randomization variance of the average treatment effects given social interference, relying in part on the accuracy of the hypothesized exposure model.

Third, power calculations for the design of networked experiments are critical and can often spell the difference between a waste of costly resources and the successful detection of meaningful signals in noisy environments and populations with a large degree of heterogeneity. Networked environments pose

unique challenges to a priori estimation of statistical power across treatment groups, particularly in the case where treated population sizes may grow (through diffusion and as a consequence of treatment) and endogenous response to treatment may require censoring of subjects to reduce leakage effects or other kinds of interference that effectively limit statistical power. The lack of preexisting estimates of adoption rates and responses to treatment in the precise context being examined are compounded by networked environments that may amplify small discrepancies between rate or response parameter estimates and their true values in a nonlinear fashion. When possible, pilot studies provide the cleanest means of assessing statistical power associated with experimental designs. However, in many circumstances pilot studies are not feasible due to constraints on timing, resource costs, or limitations set in place by third-party affiliates. In such cases, observational data from the system being studied may be used to make reasonable assessments of context-specific parameter estimates for adoption, spreading, or treatment responses in general (for a good example, see Bapna and Umyarov 2012). In addition, future studies may wish to employ methods of adaptive or conditional treatment randomization that assign subjects to treatment groups conditional on the current state of treatment group sizes or average responses to treatment.

Finally, it is important to precisely consider the mechanism of randomization and its effects on the biases we have described, in particular its implications for SUTVA, interference, sampling, and power. The mechanism of treatment randomization in networked experiments can take many forms. For example, randomized treatments can represent changes to the available channels of influence-mediating messages (e.g., Aral and Walker 2011a, 2012; Bakshy et al. 2012b; Taylor et al. 2013), changes to network structure (e.g., Centola 2010, Kearns et al. 2006, Mason and Watts 2012, Rand and Nowak 2011, Suri and Watts 2011), encouragement of behaviors in particular nodes (e.g., Bapna and Umyarov 2012), manipulations of available social cues (e.g., Bakshy et al. 2012a, Bond et al. 2012, Tucker 2011), manipulations of incentives for networked propagation (e.g., Aral and Taylor 2014), and randomization of population-level social signals (e.g., Muchnik et al. 2013, Salganik et al. 2006). Each of these mechanisms of randomization has different implications for inference, the characteristics of resultant samples, the likelihood of interference, the generalizability of parameter estimates, and, more broadly, the conclusions that can be draw from experimental results. Rather than delineate the effects of each mechanism, here we simply stress the need for networked experimental research to systematically consider the mechanism of randomization in the design and analysis of networked experiments and leave for future work the detailed assessment of the implications of each strategy.

The methods used in this study can be generalized to a wide variety of contexts to further our understanding of social contagions and better inform data-driven decisions in several policy domains including marketing, public health, and politics. In general, future work must be diligent in addressing the unique challenges created by large-scale networked experimentation. The complexity of the experimental environments in which social science is now possible creates great opportunities for social and economic research. But, these opportunities can only be truly realized if the challenges we have outlined are appropriately addressed.

8. Conclusion

The availability of microlevel data at population scale has been recognized as a crucial opportunity in the advancement of modern business analytics. At the same time, microlevel experimentation in large networked environments is a new frontier for analytics that has the potential to circumvent problematic issues of causal identification and endogeneity that have hindered our understanding of the detailed social influence processes involved in the propagation of behaviors and economic outcomes in society. Advancement of the science of social influence is vital to both marketing strategy and public policy where firms or governments seek to leverage social influence to encourage the spread of products or promote positive behaviors while curtailing negative ones. Research on social influence has predominantly focused on whether influence plays a role in the diffusion of a product or behavior and the relative size of the effect. However, recently the focus has shifted to examining when and under what individual, social, and structural conditions influence is stronger or weaker. This latter focus, which we adopt in this study, is important for policy because it can reveal which interventions (e.g., viral incentives, social interventions, targeting or network-based marketing) work best for which relationships.

We conducted a large-scale randomized experiment to identify the impact of tie strength characteristics and social embeddedness on influence. This work presents some of the first large-scale experimental evidence investigating the social and structural moderators of peer influence in networks. Results from our study shed light on the role that relationship and social structural characteristics play on influence-based propagation. We found that relationship characteristics, which capture joint participation in institutional or common social contexts, were associated with the greatest influence. For example, individuals exerted an over 13-fold increase in influence over peers with whom they attended the same college. Some measures

of the recency of social context in the relationships between individuals and their peers were associated with increased influence, whereas others were not. For example, individuals exerted an over sixfold increase in influence over peers currently living in the same town, but did not exert more influence over peers with whom they coappeared in online photos. Interestingly, measures of tie strength based on common interests were not associated with influence, though they were good predictors of preference similarity in the adoption of the product we studied. Finally, individuals exhibited greater influence on peers with whom they shared embedded relationships. This latter effect was both subtle and economically significant, highlighting the importance of large-scale randomized experiments in detecting nuanced effects in the face of endogeneity, bias, and confounding factors.

Acknowledgments

References

Airoldi E, Bai X, Carley K (2011) Network sampling and classification: An investigation of network model representations. *Decision Support Systems* 51(3):506–518.

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

Aral S, Taylor S (2014) Viral incentive systems: A randomized field experiment. Working paper, Stern School of Business, New York University, New York.

Aral S, Van Alstyne MW (2011) The diversity-bandwidth tradeoff. *Amer. J. Sociology* 117(1):90–171.

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

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

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

Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. USA* 106(51):21544–21549.

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

Arndt J (1967) Role of product-related conversations in the diffusion of a new product. *J. Marketing Res.* 4(3):291–295.

Aronow PM, Samii C (2011) Estimating average causal effects under general interference. Working paper, Yale University, New Haven, CT.

Bakshy E, Eckles D, Yan R, Rosenn I (2012a) Social influence in social advertising: Evidence from field experiments. *Proc. 13th ACM Conf. Electronic Commerce* (ACM, New York), 146–161.

Bakshy E, Rosenn I, Marlow C, Adamic L (2012b) The role of social networks in information diffusion. *Proc. 21st Internat. Conf. World Wide Web* (ACM, New York), 519–528.

Banerjee AV (1992) A simple model of herd behavior. *Quart. J. Econom.* 107(3):797–817.

Bapna R, Umyarov A (2012) Are paid subscriptions on music social networks contagious? A randomized field experiment. Working paper, University of Minnesota, Minneapolis.

Bapna R, Gupta A, Rice S, 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, Umyarov A (2012) One-way mirrors and weak-signaling in online dating: A randomized field experiment. Working paper, University of Minnesota, Minneapolis.

Berger JA, Milkman KL (2009) Social transmission, emotion, and the virality of online content. Working paper, University of Pennsylvania, Philadelphia.

Bikhchandani S, Hirshleifer D, Welch I (1998) Learning from the behavior of others: Conformity, fads, and informational cascades. *J. Econom. Perspectives* 12(3):151–170.

Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, Fowler JH (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298.

Boudreau KJ, Lakhani KR (2011) High incentives, sorting on skills—Or just a "taste" for competition? Field experimental evidence from an algorithm design contest. Working paper, London Business School, London.

Brown JJ, Reingen PH (1987) Social ties and word-of-mouth referral behavior. *J. Consumer Res.* 14(3):350–362.

Burt RS (2005) *Brokerage and Closure: An Introduction to Social Capital* (Oxford University Press, New York).

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.

Coleman J (1988) Free riders and zealots: The role of social networks. *Sociol. Theory* 6(1):52–57.

Coleman J, Katz E, Menzel H (1957) The diffusion of an innovation among physicians. *Sociometry* 20(4):253–270.

De Bruyn A, Lilien G (2008) A multi-stage model of word-of-mouth influence through viral marketing. *Internat. J. Res. Marketing* 25(3):151–163.

Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. *Science* 328(5981):1029–1031.

Easley D, Kleinberg J (2010) *Networks, Crowds, and Markets* (Cambridge University Press, Cambridge, UK).

Eckles D, Bakshy E (2014) Bias in observational studies of social contagion. Working paper, Data Science Team, Facebook, Menlo Park, CA.

Engel JF, Kegerreis RJ, Blackwell RD (1969) Word-of-mouth communication by the innovator. *J. Marketing* 33(3):15–19.

Fowler JH, Christakis NA (2010) Cooperative behavior cascades in human social networks. *Proc. Natl. Acad. Sci. USA* 107(12): 5334–5338.

Frenzen JK, Davis HL (1990) Purchasing behavior in embedded markets. *J. Consumer Res.* 17(1):1–12.

Gilly MC, Graham JL, Wolfinbarger MF, Yale LJ (1998) A dyadic study of interpersonal information search. *J. Acad. Marketing Sci.* 26(2):83–100.

Gladwell M (2002) *The Tipping Point: How Little Things Can Make a Big Difference* (Back Bay Books, New York).

Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Marketing Sci.* 23(4):545–560.

Godes D, Mayzlin D, Chen Y, Das S, Dellarocas C, Pfeiffer B, Libai B, Sen S, Shi M, Verlegh P (2005) The firm's management of social interactions. *Marketing Lett.* 16(3/4):415–428.

Golder S, Macy M (2011) Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333(6051):1878–1881.

Granovetter M (1973) The strength of weak ties. *Amer. J. Sociology* 78(6):1360–1380.

Granovetter M (1978) Threshold models of collective behavior. *Amer. J. Sociology* 83(6):1420–1443.

Granovetter M (1983) The strength of weak ties: A network theory revisited. *Sociol. Theory* 1:201–233.

Granovetter M (1985) Economic action and social structure: The problem of embeddedness. *Amer. J. Sociology* 91(3):481–510.

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

Iyengar R, Van den Bulte C, Valente TW (2011) Opinion leadership and social contagion in new product diffusion. *Marketing Sci.* 30(2):195–112.

Katz E, Lazarsfeld P (1955) *Personal Influence: The Part Played by People in the Flow of Mass Communications* (Transaction Publishers, Piscataway, NJ).

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

Leider S, Möbius MM, Rosenblat T, Do Q-A (2009) Directed altruism and enforced reciprocity in social networks. *Quart. J. Econom.* 124(4):1815–1851.

Manchanda P, Xie Y, Youn N (2008) The role of targeted communication and contagion in product adoption. *Marketing Sci.* 27(6):961–976.

Manski CF (1993) Identification problems in the social sciences. *Sociol. Methodology* 23:1–56.

Mason W, Watts DJ (2012) Collaborative learning in networks. *Proc. Natl. Acad. Sci. USA* 109(3):764–769.

Mayer-Schonberger V, Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, 1st ed. (Eamon Dolan/Houghton Mifflin Harcourt, Boston).

Middleton JA, Aronow PM (2011) Unbiased estimation of the average treatment effect in cluster-randomized experiments. Working paper, New York University, New York.

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

Nam S, Manchanda P, Chintagunta PK (2010) The effect of signal quality and contiguous word of mouth on customer acquisition for a video-on-demand service. *Marketing Sci.* 29(4):690–700.

Rand DG, Nowak MA (2011) The evolution of antisocial punishment in optional public goods games. *Nature Comm.* 2:TArticle 434.

Reagans R, McEvily B (2003) Network structure and knowledge transfer: The effects of cohesion and range. *Admin. Sci. Quart.* 48(2):240–267.

Rogers EM (2003) *Diffusion of Innovations*, 5th ed. (Free Press, New York).

Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.

Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol. Methods Res.* 40(2):211–239.

Sundararajan A, Provost F, Oestreicher-Singer G, Aral S (2013) Information in digital, economic, and social networks. *Inform. Systems Res.* 24(4):883–905.

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

Taylor S, Bakshy E, Aral S (2013) Selection effects in online sharing: Consequences for peer adoption. *14th ACM Conf. Electronic Commerce, Philadelphia.*

Tucker C (2008) Identifying formal and informal influence in technology adoption with network externalities. *Management Sci.* 54(12):2024–2038.

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

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

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

Ugander J, Backstrom L, Marlow C, Kleinberg J (2012) Structural diversity in social contagion. *Proc. Natl. Acad. Sci. USA* 109(16):5962–5966.

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

Uzzi B (1996) The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *Amer. Sociol. Rev.* 61(4):674–698.

Uzzi B (1997) Social structure and competition in interfirm networks: The paradox of embeddedness. *Admin. Sci. Quart.* 42(1):35–67.

Valente T (1996) Social network thresholds in the diffusion of innovations. *Soc. Networks* 18(1):69–89.

Valente TW (1995) *Network Models of the Diffusion of Innovations (Quantitative Methods in Communication Subseries)* (Hampton Press, NJ).

Van den Bulte C, Joshi YV (2007) New product diffusion with influentials and imitators. *Marketing Sci.* 26(3):400–421.

Van den Bulte C, Lilien G (2001) Medical innovation revisited: Social contagion versus marketing effort. *Amer. J. Sociology* 106(5):1409–1435.

Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *J. Consumer Res.* 34(4):441–458.

Zhang J (2010) The sound of silence: Observational learning in the U.S. kidney market. *Marketing Sci.* 29(2):315–335.