

Seeding Strategies for Viral Marketing: An Empirical Comparison

Seeding strategies have strong influences on the success of viral marketing campaigns, but previous studies using computer simulations and analytical models have produced conflicting recommendations about the optimal seeding strategy. This study compares four seeding strategies in two complementary small-scale field experiments, as well as in one real-life viral marketing campaign involving more than 200,000 customers of a mobile phone service provider. The empirical results show that the best seeding strategies can be up to eight times more successful than other seeding strategies. Seeding to well-connected people is the most successful approach because these attractive seeding points are more likely to participate in viral marketing campaigns. This finding contradicts a common assumption in other studies. Well-connected people also actively use their greater reach but do not have more influence on their peers than do less well-connected people.

Keywords: viral marketing, seeding strategy, word of mouth, social contagion, targeting

The future of traditional massmedia advertising is uncertain in the modern environment of increasingly prevalent digital video recorders and spam filters. Marketers must realize that 65% of consumers consider themselves overwhelmed by too many advertising messages, and nearly 60% believe advertising is not relevant to them (Porter and Golan 2006). Such information overload can cause consumers to defer their purchase altogether (Iyengar and Lepper 2000), and strong evidence indicates consumers actively avoid traditional marketing instruments (Hann et al. 2008).

Other empirical evidence also reveals that consumers increasingly rely on advice from others in personal or professional networks when making purchase decisions (Hill, Provost, and Volinsky 2006; Iyengar, Van den Bulte, and Valente 2011; Schmitt, Skiera, and Van den Bulte 2011). In particular, online communication appears increasingly important as more websites offer user-generated content,

such as blogs, video and photo sharing opportunities, and online social networking platforms (e.g., Facebook, LinkedIn). Companies have adapted to these trends by shifting their budgets from above-the-line (mass media) to below-the-line (e.g., promotions, direct mail, viral) marketing activities.

Not surprisingly then, viral marketing has become a hot topic. The term “viral marketing” describes the phenomenon by which consumers mutually share and spread marketing-relevant information, initially sent out deliberately by marketers to stimulate and capitalize on word-of-mouth (WOM) behaviors (Van der Lans et al. 2010). Such stimuli, often in the form of e-mails, are usually unsolicited (De Bruyn and Lilien 2008) but easily forwarded to multiple recipients. These characteristics parallel the traits of infectious diseases, such that the name and many conceptual ideas underlying viral marketing build on findings from epidemiology (Watts and Peretti 2007).

Because viral marketing campaigns leave the dispersion of marketing messages up to consumers, they tend to be more cost efficient than traditional massmedia advertising. For example, one of the first successful viral campaigns, conducted by Hotmail, generated 12 million subscribers in just 18 months with a marketing budget of only \$50,000. Google’s Gmail captured a significant share of the e-mail provider market, even though the only way to sign up for the service was through a referral. A recent viral advertisement by Tipp-Ex (“A hunter shoots a bear!”) triggered nearly 10 million clicks in just four weeks.

However, to enjoy such results, firms must consider four critical viral marketing success factors: (1) content, in that the attractiveness of a message makes it memorable (Berger and Milkman 2011; Berger and Schwartz 2011; Gladwell 2002; Porter and Golan 2006); (2) the structure of the social

Oliver Hinz is Chaired Professor of Information Systems esp. Electronic Markets, Technische Universität Darmstadt (e-mail: hinz@wi.tu-darmstadt.de). Bernd Skiera is Chaired Professor of Electronic Commerce, Department of Marketing, Goethe-University of Frankfurt (e-mail: skiera@wiwi.uni-frankfurt.de). Christian Barrot is Assistant Professor of Marketing and Innovation (e-mail: christian.barrot@the-klu.org), and Jan U. Becker is Assistant Professor of Marketing and Service Management (e-mail: jan.becker@the-klu.org), Kühne Logistics University. The authors thank Carsten Takac, Philipp Schmitt, Martin Spann, Lucas Bremer, Christian Messerschmidt, Daniele Mahoutchian, Nadine Schmidt, and Katharina Schnell for helpful input on previous versions of this article. In addition, three anonymous *JM* referees provided many helpful suggestions. This research is supported by the E-Finance Lab Frankfurt.

network (Bampo et al. 2008); (3) the behavioral characteristics of the recipients and their incentives for sharing the message (Arndt 1967); and (4) the seeding strategy, which determines the initial set of targeted consumers chosen by the initiator of the viral marketing campaign (Bampo et al. 2008; Kalish, Mahajan, and Muller 1995; Libai, Muller, and Peres 2005). This last factor is of particular importance because it falls entirely under the control of the initiator and can exploit social characteristics (Toubia, Stephen, and Freud 2010) or observable network metrics. Unfortunately, there is a “need for more sophisticated and targeted seeding experimentation” to gain “a better understanding of the role of hubs in seeding strategies” (Bampo et al. 2008, p. 289).

The conventional wisdom adopts the influentials hypothesis, which states that targeting opinion leaders and strongly connected members of social networks (i.e., hubs) ensures rapid diffusion (for a summary of arguments, see Iyengar, Van den Bulte, and Valente 2011). However, recent findings raise doubts. Van den Bulte and Lilien (2001) show that social contagion, which occurs when adoption is a function of exposure to other people’s knowledge, attitudes, or behaviors (Van den Bulte and Wuyts 2007), does not necessarily influence diffusion, and yet it remains a basic premise of viral marketing. Such contagion frequently arises when people who are close in the social structure use one another to manage uncertainty in prospective decisions (Granovetter 1985). However, in a computer simulation, Watts and Dodds (2007) show that well-connected people are less important as initiators of large cascades of referrals or early adopters. Their finding, which Thompson (2008) provocatively summarizes by implying “the tipping point is toast,” has stimulated a heated debate about optimal seeding strategies, though no research offers an extensive empirical comparison of seeding strategies. Van den Bulte (2010) thus calls for empirical comparisons of seeding strategies that use sociometric measures (i.e., metrics that capture the social position of people).

In response, we undertake an empirical comparison of the success of different seeding strategies for viral marketing campaigns and identify reasons for variations in these levels of success. In doing so, we determine whether companies should care about the seeding of their viral marketing campaigns and why. In particular, we study whether well-connected people really are more difficult to activate, participate more actively in viral campaigns, and have more influence on their peers than less well-connected people. In contrast with previous studies that rely on analytical models or computer simulations, we derive our results from field experiments and from a real-life viral marketing campaign.

We begin this article by presenting literature relevant to viral marketing and social contagion theory. We introduce our theoretical framework, which disentangles the determinants of social contagion, and present four seeding strategies. Next, we empirically compare the success of these seeding strategies in two complementary field experiments (Studies 1 and 2) that aim at spreading information and inducing attitudinal changes. Then, we analyze a real-life viral marketing campaign designed to increase sales, which provides an economic measure of success. After we identify the determinants of success, we conclude with a discussion of our research contributions, managerial implications, and limitations.

Theoretical Framework

When information about an underlying social network is available, seeding on the basis of this information, as typically captured by sociometric data, seems promising (Van den Bulte 2010). Such a strategy can distinguish three types of people: “hubs,” who are well-connected people with a high number of connections to others; “fringes,” who are poorly connected; and “bridges,” who connect two otherwise unconnected parts of the network. The sociometric measure of degree centrality captures connectedness within the local environment (for details, see the Web Appendix at <http://www.marketingpower.com/jmnov11>), such that high-degree-centrality values characterize hubs, whereas low-degree-centrality values mark fringes. In contrast, the sociometric betweenness centrality measure describes the extent to which a person acts as a network intermediary, according to the share of shortest communication paths that pass through that person (see the Web Appendix). Thus, bridges earn high values on betweenness centrality measures.

Determinants of Social Contagion

Following Van der Lans et al. (2010), we propose a four-determinant model of social contagion to determine the success of viral marketing campaigns. First, individual i receives a viral message from sender s , who can be either a friend or the campaign initiator that makes i aware of and informed by the message with information probability I_i . Individual i then may become active and participate in the campaign with participation probability P_i . Given participation, individual i passes the message to a set of recipients J_i , where n_i is the number of recipients ($|J_i| = n_i$), such that it provides a measure of used reach. The number of expected referrals R_i by individual i , then, is the product of the information probability (I_i), the probability of participating (P_i), and the used reach (n_i): $R_i = I_i \times P_i \times n_i$.

The conversion rate $w_{i,j}$ linearly influences the number of expected successful referrals SR_i of individual i on recipients j ($j \in J_i$), given by

$$(1) \quad SR_i = I_i \times P_i \times n_i \times \sum_{j=1}^{n_i} \frac{w_{i,j}}{n_i}.$$

If a sender i has the same conversion rate for all recipients, such that $w_{i,j} = w_i \forall j \in J_i$, the number of expected successful referrals can be rewritten as

$$(2) \quad SR_i = I_i \times P_i \times n_i \times w_i.$$

All these determinants are a function of i ’s social position, though they also may be influenced by the characteristics of the sender s and the conversion rate w_s . Despite the lack of empirical comparisons of seeding strategies for viral marketing campaigns, various studies in marketing, sociology, and epidemiology have analyzed the influence of the social position (captured by sociometric measures) on different determinants, such as whether hubs are more likely to persuade their peers. In Table 1, we summarize these findings according to the determinants of information probability I_i , participation probability P_i , used reach n_i , and conversion rate w_i .

TABLE 1
Previous Research

Studies	Context	Reason for Contagion	Participation Probability P_i	Used Reach n_i	Social Position Has Positive Influence on ...			Empirically Tested Seeding Strategy
					Expected Number of Referrals R_i	Conversion Rate w_i	Expected number of Successful SR_i	Recommendation for Optimal Seeding Strategy
Coleman, Katz, and Menzel (1966)	Product (low risk)	A, BU, NP	Hub		Hub			Hub
Becker (1970)	Product (low risk)	A, BU, NP	Hub		Hub			Hub
	Product (high risk)		Fringe		Fringe			Fringe
Simmel (1950); Porter and Donthu (2008)	Messages	A	Fringe					
	Messages	A						
Watts and Dodds (2007)	—	—	Fringe		Fringe		Fringe	Fringe
Leskovec, Adamic, and Huberman (2007)	Product (low risk)	A, BU	Hub		Hub	Fringe		
Anderson and May (1991); Kemper (1980)	Epidemiology	A	Hub		Hub		Hub	Hub
	Epidemiology	A						
Granovetter (1973); Rayport (1996)	Messages	A	Bridge		Bridge			Bridge
	Messages	A						
Iyengar, Van den Bulte, and Valente (2011)	Product (high risk)	A, BU	Hub		Hub	Hub	Hub	Hub
Study 1	Messages	A	Controlled				✓	Hub, fringe, bridge, random
Study 2	Messages	A					✓	Hub, fringe, bridge, random
Study 3	Product (low risk)	A, BU	✓	✓	✓	✓	✓	Hub, fringe, random

Notes: A = awareness, BU = belief updating, NP = normative pressure, and i = focal individual. Expected number of referrals: $R_i = P_i \times n_i$; Successful number of referrals: $SR_i = w_i \times R_i$.

Effect of social position on information and participation probability. A viral marketing campaign aims to inform consumers about the viral marketing message and to encourage them to participate in the campaign by sending the message to others. In investigating the impact of social position on information probability, Goldenberg et al. (2009) indicate that hubs tend to be better informed than others because they are exposed to innovations earlier through their multiple social links. In his reanalysis of Coleman, Katz, and Menzel's (1966) medical innovation study, Burt (1987) also recognizes that some people experience discomfort when peers whose approval they value adopt an innovation they have not yet adopted; in this case, social contagion (reflected in a greater probability to participate) results from normative pressure and status considerations. This mechanism could explain why Coleman, Katz, and Menzel (1966) find that highly integrated people (e.g., hubs) are more likely to adopt an innovation early than are more isolated people.

However, in some cases, hubs do not adopt innovations first (Becker 1970), such as when the innovation does not suit the hub's opinion, which may mean that adoption occurs first at the fringes of the network (Iyengar, Van den Bulte, and Valente 2011). Another potential explanation of hubs' lower participation probability stems from information overload effects. Because hubs are exposed to so many contacts, they possess a wealth of information and thus might be more difficult to activate (Porter and Donthu 2008; Simmel 1950) or less likely to participate in viral marketing campaigns. Overall, information and participation probabilities remain difficult to disentangle; for the purposes of this study, we assume that all receivers of viral marketing messages are aware of them. This assumption is likely to hold for our three empirical studies, and our main findings remain unchanged even when it does not. The only difference is that the participation probability would also capture the probability that a person is informed (and thus aware) of the viral marketing campaign.

Effect of social position on used reach. Epidemiology studies indicate that hub constellations foster the spread of diseases (Anderson and May 1991; Kemper 1980), which suggests in parallel that hubs should be more attractive for seeding viral marketing campaigns. However, it is unclear whether hubs actively and purposefully make use of their potential reach. The deliberate use of reach is a common assumption, and yet only Leskovec, Adamic, and Huberman (2007) actually confirm that hubs send more messages. Furthermore, their definition of hubs relies on messaging behavior, such that it cannot offer generalizable evidence of the assumption that hubs actively use their greater reach potential.

In contrast, we anticipate that individual i 's used reach (first generation), added to the used reach of successive generations that originate from i 's initial direct reach (second and further generations), which we call i 's "influence domain" (Lin 1976), depends on the number of others who already have received the message. In this setting, bridges are advantageous because they can forward the message to different parts of the network (Granovetter 1973) that have not yet been infected with the viral campaign.

Effect of social position on conversion rate. A person's social position might also indicate the degree of persuasiveness, as measured by the conversion rate—namely, the share of referrals that lead to successful referrals. Iyengar, Van den Bulte, and Valente (2011) find that hubs are more likely to be heavy users and that, therefore, their influence is more effective, because they act in accordance with their own recommendations (e.g., by making heavy use of the innovation). Leskovec, Adamic, and Huberman (2007) find that the success rate per recommendation decreases with the number of recommendations made, which implies that people have influence over a limited number of friends but not over everybody they know. This result indicates that the conversion rate decreases when hubs use their full reach potential, though it does not preclude the notion that hubs instead might select relevant subsets of recipients from among their peers and thus achieve high conversion rates. Thus, the effect of social position on the conversion rate remains unclear. Goldenberg et al. (2009) still make what they call the conservative assumption that hubs are not more persuasive than others, though without empirical support.

Seeding Strategies

As our review in Table 1 shows, little consensus exists regarding recommendations for optimal seeding strategies. Four studies recommend seeding hubs, three recommend fringes, and one recommends bridges. We analyze these discrepant recommendations in turn.

If at least one of the determinants I_i , P_i , n_i , or w_i increases with the connectivity of the sender i , and the remaining determinants are not correlated with higher connectivity, hubs should be the targets of initial seeding efforts, because they spread viral information best—as indicated in Hanaki et al. (2007), Van den Bulte and Joshi (2007), and Kiss and Bichler (2008). Using hubs as initial seeding points implies a "high-degree seeding" strategy.

In contrast, Watts and Dodds's (2007) computer simulations of interpersonal influence indicate that targeting well-connected hubs to maximize the spread of information works only under certain conditions and may be the exception rather than the rule. They propose instead that a critical mass of influenceable people, rather than particularly influential people, drives cascades of influence. The impact on triggering critical mass is not even proportional to the number of people who hubs directly influence; rather, according to Dodds and Watts (2004), the people most easily influenced have the highest impact on the diffusion. Moreover, if hubs suffer from information overload because of their central position in the social network (Porter and Donthu 2008; Simmel 1950), they must filter or validate the vast amount of information they receive, such that they may be less susceptible to information received from anyone outside their trusted network. In their analytical model, Galeotti and Goyal (2009) propose targeting low-degree members instead, on the fringe of the network, if the probability of adopting a product increases with the absolute number of adopting neighbors. Sundararajan (2006) similarly suggests seeding the fringes rather than the hubs, which we refer to as a "low-degree seeding" strategy.

When analyses focus on the influence domain, encompassing referrals beyond the first generation, it also becomes necessary to consider centrality beyond the local

environment. Bridges who connect otherwise separated subnetworks have vast influence domains, such that seeding them might enable information to diffuse throughout parts of the network and prevent a viral message from simply circulating in an already infected, highly clustered subnetwork. Accordingly, Rayport (1996) recommends exploiting the strength of weak ties (i.e., bridges; Granovetter 1973) to spread a marketing virus. From an opposite perspective, Watts (2004) similarly recommends eliminating bridges to prevent epidemics. We thus refer to the idea of seeding bridges as a “high-betweenness seeding” strategy.

Finally, if there is no correlation between social position and the determinants I_i , P_i , n_i , and w_i , or if the opposing influences of the determinants nullify one another, there should be no differences across the proposed strategies or a random targeting. We also test this “random seeding” strategy, which further serves as a benchmark situation in which no information about the social network is available.

Method

We use three studies to empirically compare the success of seeding strategies and identify which of the determinants are influenced by the people’s social position. Our three studies encompass two types of settings that are particularly relevant for viral marketing. First, viral marketing campaigns primarily aim to spread information, create awareness, and improve brand perceptions, which are noneconomic goals. Second, other campaigns attempt to increase sales through mutual information exchanges between adopters and prospective adopters, to trigger belief updating, such that we can use an economic measure of success.

These goals map well onto the classification that Van den Bulte and Wuyts (2007) provide to describe five reasons for social contagion, the first two of which are especially relevant for viral marketing campaigns. First, people may become aware of the existence of an innovation through WOM provided by previous adopters in a simple information transfer. Second, people may update their beliefs about the benefits and costs of a product or service. Third, social contagion may occur through normative pressures, such that people experience discomfort when they do not comply with the expectations of their peer group. Fourth, social contagion can be based on status considerations and competitive concerns (i.e., the level of competitiveness between two people). Fifth, complementary network effects might cause social contagion, in which the benefit of using a product or service increases with the number of users.

To examine both types of viral marketing campaigns, we conduct two experimental studies (Studies 1 and 2) that simulate viral marketing campaigns in which social contagion mainly involves simple information transfers and results in greater awareness as a noneconomic measure of success. The aim is to compare the success of different seeding strategies. In Study 3, we examine a viral marketing campaign in which social contagion relies on belief updating and results in sales (i.e., economic measure). In Table 2, we summarize the complementary setup of the three studies, which helps us overcome some individual limitations of each study.

Experimental Comparison of Seeding Strategies

In Studies 1 and 2, we compare the success of our four seeding strategies in different conditions and confirm the robustness of the results across different settings. Trusov, Bodapati, and Bucklin (2010) point out the necessity to conduct such experiments: In analyzing data from a major social networking site, they find that only approximately one-fifth of a user’s friends actually influence that user’s activity on the site. However, they cannot discern how responsive the “top influencers” are or whether marketers should use information about underlying social networks to seed their viral marketing campaigns. Therefore, they call for further research that uses straightforward field experiments. Because such experiments can help identify best-practice strategies, we compare the four seeding strategies in two small-scale field experiments.

Study 1: Comparison of seeding strategies in a controlled setting. We begin with a controlled setup to ensure internal validity and control for willingness to actively participate P_i (see Table 1). We recruited 120 students from a German university. The recruitment and commitment processes ensured relatively similar participants in terms of communication activity across treatments because all of them expressed a willingness to contribute actively. Therefore, we expect minimal variation in activity levels, compared with a study in which respondents are unaware of their participation or do not come into direct contact with the experimenter. A prerequisite for participation was maintaining an account on a specified online social networking platform (similar to Facebook). Using proprietary software, we automatically gathered each participant’s friends list from the platform and then applied an event-based approach to specify boundaries, such that we discarded all links to friends who did not participate in the experiment. The software Pajek calculated the sociometric measures (degree centrality and betweenness centrality; see the Web Appendix at <http://www.marketingpower.com/jmnov11>) for each participant.

The social network thus generated consisted of 120 nodes (i.e., participants) with 270 edges (i.e., friendship relations). Degree centrality ranged from 1 to 17, with a mean of 4.463 and a standard deviation of 3.362. In other words, the participants had slightly more than four friends each, on average, in the respective, bounded social network. The correlation (.592, $p < .01$) between the degree centrality in this small, bounded network created by the artificial boundary specification strategy (using the criteria “participation in experiment”) and the degree centrality of the entire network hosted by this social networking platform (6.2 million unique users, November 2009) is striking. It also supports Costenbader and Valente’s (2003) claim that some centrality metrics are relatively robust across different network boundaries. That is, the boundary we applied does not appear to bias degree centrality, even for a sub sample that comprises as little as .002% of the entire social network. The betweenness centrality ranged from 0 to .320, with a standard deviation of .053.

We used these sociometric measures to implement our four seeding strategies. The seeding relied on the message function of the social networking platform, such that we sent unique tokens of information to a varying subset of participants (the total population remained unchanged throughout this experiment) and traced the contagion pro-

TABLE 2
Summary of Studies

	Study 1	Study 2	Study 3
Seeding strategies	Four seeding strategies: •High degree (HD) •Low degree (LD) •High betweenness (HB) •Random (control)	Four seeding strategies: •High degree (HD) •Low degree (LD) •High betweenness (HB) •Random (control)	Three seeding strategies: •High degree (HD) •Low degree (LD) •Random
Social contagion through	Awareness (advertisement)	Awareness (advertisement)	Belief updating (service referral)
Motivation	Extrinsic motivation for sharing (experimental remuneration)	Intrinsic motivation (funny video about university)	Extrinsic motivation for sharing (additional airtime for referral)
Seeding size	10% of network size 20% of network size	7% of network size	Entire network
Seeding timing	Sequential	Parallel	—
Social network	120 nodes (small network), 270 edges	1380 nodes (medium-sized network), 4052 edges	208,829 nodes (very large network), 7,786,019 edges
Number of treatments	$16 = 4 \times 2 \times 2$	4	—
Number of replications	2 (4 treatments missing)	1	—
Number of experimental settings	28	4	—
Boundary of network	Artificial	Natural	Natural
Design strengths	Test of causality, strong control due to experimental setup, identification of individual behavior due to specific IDs	Test of causality, realistic scenario	Large real-world network based on firm data, identification of determinants
Design weaknesses	Repeated measures due to sequential timing, artificial scenario	Potential interaction between treatments, activity level of individuals not controlled for, individuals cannot be identified	Missing edges between noncustomers (HB could not be tested), causality cannot be tested
Specific finding	HD and HB are comparable and outperform random by +39%–52% and LD by factor 7–8	HD and HB are comparable and outperform random by +60% and LD by factor 3	HD outperforms random by factor 2 and LD by factor 8–9

cess. These tokens were to be shared by initial recipients with friends, who in turn were to spread them further. All receivers were asked to enter the tokens on a website that we created for this purpose, along with details about from whom they received these tokens (called the “referrer”). Because each participant was provided with unique login information for this website, we could observe the number of tokens entered on the website (and thus the number of successful referrals SR_i) by each individual i for each of the seeding strategies. Furthermore, we could distinguish whether the recipient received the tokens directly from the experimenter (“Seeded by Experimenter”) or through viral spreading from friends. We prohibited and did not observe the use of forums or mailing lists to spread the tokens.

The experiment used a $4 \times 2 \times 2$ full-factorial design. Following the strategies we defined previously, we seeded the tokens every few days to hubs (high-degree seeding), fringes (low-degree seeding), bridges (high-betweenness seeding), or a random set of participants. We varied the number of initial seeds, such that the tokens were sent to either 12 (10%) or 24 (20%) of the 120 participants.

We also varied the payment levels for successful referrals to account for the potential effects of extrinsic motivation (incentive for sharing yes/no). When they received no incentive for sharing, participants earned remuneration only if they correctly entered the secret token (~.40 EUR per token; for detailed instructions, see the Web Appendix at <http://www.marketingpower.com/jmnov11>). In contrast, under the incentives-for-sharing condition, they received an additional monetary reward when they were named as a referrer (.25 EUR per correctly entered token and .20 EUR per referral; for detailed instructions, see the Web Appendix).

Therefore, we systematically varied the $4 \times 2 \times 2 = 16$ different treatments, with two replications per treatment. The limitations of the social networking platform’s messaging system prevented us from replicating four specific treatments, so we obtained a total of 28 experimental settings (the potential maximum was $4 \times 2 \times 2 \times 2 = 32$ experimental settings). Although we systematically varied the treatments, we placed the low-incentive-for-sharing before the high-incentive-for-sharing settings to avoid confusing participants with different incentive instructions. We always

seeded one token and then captured all responses two weeks after the seeding.

Overall, 55% of the participants actively spread or entered unique tokens, resulting in 1155 responses. The average number of tokens spread per experimental setting was 41.25, with a standard deviation of 19.21. To compare the success of the strategies, we use a random-effects logistic regression analysis that accounts for individual behavioral differences according to each participant's responsiveness in each experimental setting. We use the number of correctly entered tokens as a dependent variable, which can be 1 if the token was correctly reported and 0 if otherwise. With 120 participants and 28 experimental settings, we obtained 3360 observations. As the independent variables, we included dummy-coded treatment variables that reflect our full-factorial design, as we detail in Table 3.

The model achieves a pseudo R-square of 15.5%. The proportion of unexplained variance accounted for by subject-specific differences due to unobserved influences, labeled ρ , is greater than 90%. Compared with random seeding, the high-degree seeding strategy yields a much higher likelihood of response (odds ratio = 1.53) that is similar to the high-betweenness seeding strategy (odds ratio = 1.39). In contrast, the low-degree seeding strategy dramatically decreases the likelihood of response (odds ratio = .19).

Our treatment variable, high seeding (dummy coded as 0 = 12 seeds and 1 = 24 seeds), positively influences response likelihood. Furthermore, the type of incentive offered drives the high odds ratio estimate, which might explain why extrinsic motivation in the form of monetary incentives is popular for viral marketing (e.g., recruit-a-friend campaigns offering rewards such as price discounts or coupons for successful referrers; Biyalogorsky, Gerstner, and Libai 2001). Finally, the participants who received the token from the experimenter ("seeded by experimenter") exhibited a higher response likelihood, which is not surprising because the information probability in this case equals 1.

TABLE 3

Individual Probability to Respond (i.e., Entering the Correct Token at Experimental Website [Random Effects Logit Model, Study 1])

Variable	Odds Ratio	SE
Seeding Strategy		
Low degree	.19***	.04
High betweenness	1.39*	.28
High degree	1.53**	.31
High seeding	1.89***	.28
High incentives	38.11***	26.07
Seeded by experimenter	14.36***	3.74
Random coefficient: User ID		
$\ln(\delta_u^2)$	3.36	.17
δ_u	5.36	.46
P	.90	.02
R ² (pseudo)	.16	
N	3360	

* $p < .1$.

** $p < .05$.

*** $p < .01$.

Notes: Two-tailed significance levels. n.s. = not significant. Reference category: "random" seeding strategy, "low seeding," "no incentives," and "was not seeded by experimenter."

To compare the various seeding strategies directly, we also varied the contrast specifications but left the rest of the model unchanged, which produced the conditional odds ratio matrix in Table 4. As Table 4 indicates, both high-degree and high-betweenness seeding increase response likelihood, in contrast with the random seeding strategy, by 39%–53%. Compared with the low-degree seeding strategy (second column of Table 4), all other strategies are five to eight times more successful. However, a comparison of the two most successful seeding strategies, high betweenness and high degree, does not yield significant differences. This result has key implications for marketing practice, in that degree centrality as a local measure is much easier to compute than betweenness centrality, which requires information about the structure of the entire network.

In summary, we find that the low-degree seeding strategy is inferior to the other three seeding strategies and that both high-betweenness and high-degree seeding outperform the random seeding strategy but yield comparable results. However, we also acknowledge that this experiment might suffer from sequential effects, which would limit the validity of our separate analysis of each experimental setting. The behavior of a respondent in one experimental setting might be influenced by his or her experience in prior experimental settings. This problem is driven by the limited number of participants in our experiment. Therefore, in Study 2 we include more participants and avoid sequential effects by implementing the four seeding strategies simultaneously.

Study 2: Comparison of seeding strategies in a field setting. In a second field experiment, we focused on the entire online social network of all students enrolled in the MBA program at the same university as in Study 1. Thus, the network boundary is defined by participation in the program. We collected contact information for 1380 students (1380 nodes, 4052 edges) by crawling the same social networking platform to collect information on friendships, and we then calculated the sociometric measures as in Study 1.

TABLE 4

Conditional Odds Ratios of Seeding Strategies (Study 1)

	Low Degree	Random	High Betweenness	High Degree
Low degree	—	.19***	.13***	.12***
Random	5.37***	—	.72*	.65**
High betweenness	7.47***	1.39*	—	.91 ^{n.s.}
High degree	8.19***	1.53**	1.10 ^{n.s.}	—

* $p < .1$.

** $p < .05$.

*** $p < .01$.

Notes: Two-tailed significance levels. n.s. = not significant. Read the second column as follows: The odds that a person reacts to the strategy of random seeding is 5.37 times as large as that for low-degree seeding, 7.47 times as large in the strategy of high-betweenness seeding as for low-degree seeding, and 8.19 times as large in the strategy of high-degree seeding as for low-degree seeding. The conditional odds ratio of the two seeding strategies relate inversely. For example, the odds ratios of random and low degree relate as follows: .19 = 1/5.37.

The mean degree centrality (standard deviation) is 5.872 (7.318). Study 2 also reveals a high and significant correlation between degree centrality in the bounded network (1380 MBA students at the university) and degree centrality in the entire network of the social networking platform (6.2 million unique users in November 2009). The Pearson correlation of .824 ($p < .001$) thus suggests that the number of friends reported is also a good indicator of degree centrality in abounded network.

As a proxy for the level of activity, we also used the time since the last profile update in Study 2. We acquired information about 849 update time stamps (we could not access 531 due to the privacy restrictions set by users). On average, users updated their profile 25.7 weeks ago ($Mdn = 15.0$), and we observed a weak but significant correlation between degree centrality and time (in weeks) since the last profile update ($r = -.192, p < .01$). We also observed a correlation between betweenness centrality and time since the last profile update ($r = -.154, p < .01$). These negative correlation simply that participants who updated their profiles more recently (and probably update them more frequently) are also more central in the social network. In other words, activity correlates with centrality and may be an additional determinant of the viral spread of information in this setting. In terms of gender (805 male, 569 female, and 6 missing observations), male participants were more central, such that the average female participant had .92 fewer connections than the average male ($p < .05$). However, this gender difference becomes insignificant if we control for activity.

The experimental setup for Study 2 was somewhat different. First, the four treatment groups (hubs, bridges, fringes, or random sample) were all seeded on the same day. Second, we eliminated the incentive variation, such that we did not use extrinsic monetary incentives to stimulate participation. Third, we did not vary the seeding size and sent a reminder out to the initial seeds seven days after the initial seeding. The seeding included 95 participants in each of the four treatments (70 on Day 1, 25 on Day 2), or 7% of the total network (which is in line with Jain, Mahajan, and Muller 1995). The seed message contained a

unique URL for a website with a funny video that we produced about the participants' university (the landing page and video were identical for all treatments). By producing a new video specifically for this second field experiment, we ensured that the viral marketing stimulus (i.e., content) was unknown to all participants. Furthermore, we predicted that the link to the video would be distributed preferentially to fellow students (from which we obtained mutual online social network relationships), rather than to others outside the university's social network. In other words, the social network for Study 2 should represent a coherent, self-contained social community. One MBA student served as the initiator who seeded the message to others, according to the chosen seeding strategy. In addition to the link to the particular entry page, the message indicated that the addressees could find a funny video about the university that had just been created by the initiator.

We tracked website visits for the entry pages and video download pages of the four sites (one for each strategy) for 19 days. Figure 1 compares the success of the seeding strategies. The rank order with respect to their success, across both dependent variables, is consistent with the results from our first experiment. That is, high-degree and high-betweenness seeding clearly outperform both low-degree and random seeding. For example, in terms of videos watched, the high-degree seeding strategy yielded more than twice the number of responses than did random seeding. Information about social position thus made it possible to more than double the number of responses.

We also estimated two random-effects linear models (one for the entry page, one for the video page) in which we treated each of the 19 days as a unit of observation, for which we have four observations. The dependent variable is thus the number of unique visits for the entry page and the number of unique video requests from the video page. We included the seeding and reseeded (reminder) days as dummy variables and added another dummy variable to account for weekends. The seeding strategies also are coded as dummy variables, and the experimental day is a unit specific random coefficient. Table 5 illustrates the results.

FIGURE 1
Development of the Number of Unique Visits Over Time (Study 2)

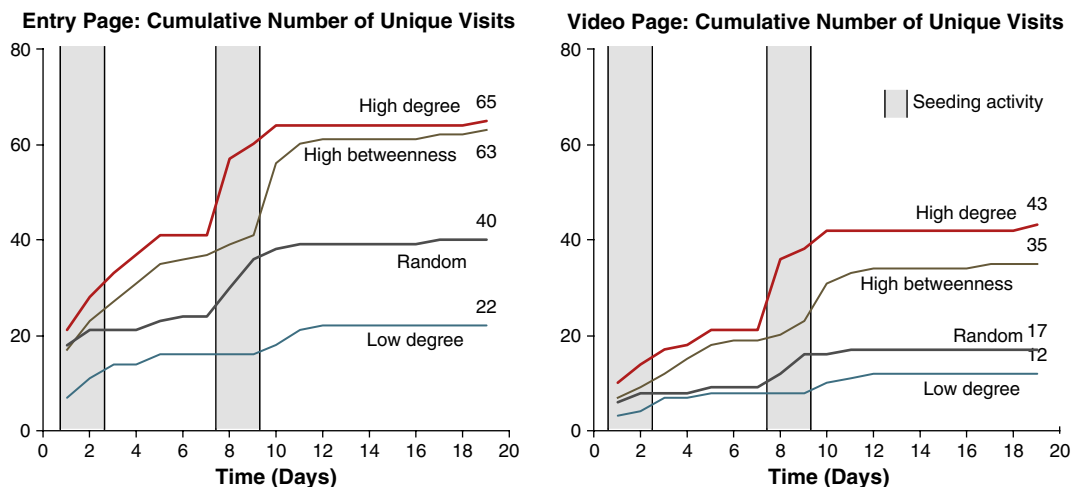


TABLE 5
Number of Visits per Day (Random Effects Model, Study 2)

Variable	Entry Page Unique Visits		Video Page Unique Visits	
	Coefficient	SE	Coefficient	SE
High-degree seeding	2.263***	.766	1.623***	.547
High-betweenness seeding	2.158***	.766	1.211**	.547
Random seeding	.947 ^{n.s.}	.766	.263 ^{n.s.}	.547
Seeding day or reseeding	7.128***	1.276	4.005***	.727
Weekend	-1.026 ^{n.s.}	1.569	-.636 ^{n.s.}	.794
Intercept	-.127 ^{n.s.}	.911	-.078 ^{n.s.}	.524
Random Coefficient: Experimental Day				
δ_e	2.375		1.642	
P	.519		.290	
R ² (overall)	.475		.436	

* $p < .1$.

** $p < .05$.

*** $p < .01$.

Notes: Two-tailed significance levels. n.s. = not significant. Reference categories: "low-degree seeding" and "weekdays" and "no seeding day."

The models for both the entry and video pages are highly significant, with explained overall variances (adjusted R^2) of 47.5% and 43.6%. The results in Figure 1 confirm our previous observations: High-degree and high-betweenness seeding yield comparable results and are three times more successful than low-degree seeding and 60% more successful than random seeding. Days with seeding or reseeding activities yield more unique visits. Responsiveness declined on weekends (albeit insignificantly), perhaps due to the overall higher level of online activity by these students on weekdays.

In summary, Study 2 supports our findings from Study 1 that seeding to hubs and bridges is preferable to seeding to fringes. However, we also note the potential for interactions among the activities associated with the four seeding strategies in Study 2. For example, a participant might have watched the video after receiving a message from seeding strategy A and then receive a nearly identical message from seeding strategy B, in which case this participant is unlikely to click the link again to watch the same video. Thus, seeding strategies that foster faster diffusion may have an advantage that could bias the result and lead to over estimations of the success of high-degree and high-betweenness seeding in contrast with random and low-degree seeding. However, in Study 1, such crossings were not possible as a result of the sequential timing, and the results remained the same. Neither Study 1 nor Study 2 identifies the reasons for the superiority of specific seeding strategies. In addition, we cannot distinguish between first- and second-generation referrals. We address these shortcomings in Study 3.

Comparison of the Effect of Seeding Strategies on the Determinants of Social Contagion in a Real-Life Viral Marketing Campaign (Study 3)

For Study 3, a mobile phone service provider stimulated referrals (through text messages) to attract new customers. The provider tracked all referrals, so we can compare the economic success of different seeding strategies and analyze the influence of the corresponding sociometric measures on all determinants of social contagion (Table 1) in a real-life setting. This helps us identify the reasons for

any differences. Thus, Study 3 enables us to decompose the effect of the different determinants that drive the social contagion process, including participation probability P_i , the used reach n_i , the mean conversion rate w_i of all referrals made by ion the expected number of referrals R_i , and the expected number of successful referrals SR_i . The viral marketing campaign of the mobile phone service provider featured text messages sent to the entire customer base ($n = 208,829$ customers), promising a 50% higher reward than the regular bonus of € 10 worth of airtime for each new customer referred in the next month. In total, 4549 customers participated in the campaign, initiating 6392 first-generation referrals, which was a 50% increase over the average number of referrals. We anticipate that social contagion works through belief updating, as prospective customers talk to adopters about the product. Furthermore, in Becker's (1970) terms, we classify this product as a low-risk offering (cf. trials of untested drugs).

Our analysis of the social contagion process is based on a rich data set; each referral activity was logged in the online referral system of the company, because customers had to initiate the referral messages to friends online. Successful referrals were confirmed during the registration process of the new customers, who had to identify their referrer to trigger the payment of the referral premium. Thus, we gathered information about whether customers acted on the stimulus of the referral, captured by the variable program participation P_i , as well as the number of referrals R_i and the number of successful referrals SR_i . The mean conversion rate per referrer w_i can be inferred from a comparison of R_i and SR_i . We used individual-level communication data and the number of text messages to others to calculate the (external) degree centrality. (In total, we evaluated more than 100 million connections.¹) We assumed that any telephone call or text message between people (independent of the direction) reflected social ties. Thus, degree centrality

¹All individual-level data were made anonymous with a multi-stage encryption process, undertaken by the firm before the analysis. At no point was any sensitive customer information, such as names or telephone numbers, disclosed.

equals a count of the total number of unique communication relationships. However, the service can only be referred to current non customers, so the degree centrality metric accounts only for ties that customers had to people outside the service network at the beginning of the viral marketing campaign, which makes it a form of external degree centrality. We lacked information about the relationships of people who were not customers, so we could not measure betweenness centrality and test the high-betweenness seeding strategy in Study 3.

We used the following customer characteristics as covariates: demographic information including age (in years) and gender (1 = female, and 0 = male); service-specific characteristics, such as customer tenure (i.e., length of the relationship with the company in months); and the tariff plan. We operationalized the tariff plan with a dichotomous variable to indicate whether the customer chose a community tariff (=1, including a reduced per-minute price for calls within the network) or a one-price tariff (=0). Furthermore, we used two measures of customers' trust in the service: payment type (dichotomous variable: automatic = 1, and manual = 0) and refill policy (dichotomous variable: automatic = 1, and manual = 0). In the case of automatic payment and refill, customers provided credit card details to the service provider. Finally, we included information about the acquisition channel for each customer (1 = offline/retail, and 0 = online). As additional controls, we include information on the individual service usage of the customer, namely, average monthly airtime (in minutes) and monthly short message service (SMS)—that is, the average monthly number of SMS sent by a customer.

Our model reflects the two-stage process for each participant, who first decides whether to participate (P_i) and then chooses to what extent to participate (n_i). A specific characteristic of the first stage is the relatively large share of zeros (i.e., nonparticipants), whereas observed values for the second stage are count measures and highly skewed. This data structure requires specific two-stage regression models: either inflation models, such as the zero-inflated Poisson (ZIP) regression (Lambert 1992), or hurdle models, such as the Poisson-logit hurdle regression (PLHR) model (Mullahy 1986). We use a PLHR model, which combines a logit model to account for the participation decision and a zero-truncated Poisson regression to analyze the actual outcomes of participation (e.g., number of successful referrals).² In our PLHR specification, the binary variable P_i indicates whether individual i participates in the referral program (hurdle or logit model). In addition, Used Reach n_i indicates how many referrals the individual i initiates, conditional on the decision to participate ($P_i = 1$). As an extension, Converted Reach $CR = (n_i \times w_i)$ indicates how many successful referrals individual i initiates, again conditional on the decision to participate. Note that Used Reach n_i and Converted Reach CR_i are equivalent to Referrals R_i and Successful Referrals SR_i , respectively, conditional on a program participation probability $P_i = 1$. These variables

provide the dependent variables in the Poisson regression of our PLHR specification.

Let P_i^* be the latent variable related to P_i , n_i^* be the censored variable related to n_i , and $CR_i^* = (n_i^* \times w_i)^*$ be the censored variable related to $CR = (n_i \times w_i)$. Together with the explanatory variable of (external) degree centrality and the covariates (age, gender, payment type, refill policy, acquisition channel, and customer tenure), the PLHR can be specified as follows:

$$(3) \quad P_i = \begin{cases} 1 & \text{if } P_i^* > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{where } P_i^* = \beta_{P0i} + \beta_{Pij} \times X_{Pij} + \varepsilon_{Pi},$$

$$(4) \quad n_i = \begin{cases} n_i^* & \text{if } P_i^* > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{where } n_i^* = \beta_{UR0i} + \beta_{URij} \times X_{URij} + \varepsilon_{URi},$$

and

$$(5) \quad CR = (n_i \times w_i) = \begin{cases} (n_i \times w_i)^* & \text{if } CR_i^* > 0 \\ 0 & \text{otherwise,} \end{cases} \quad \text{where } (n_i \times w_i)^* = \beta_{CR0i} + \beta_{CRij} \times X_{CRij} + \varepsilon_{CRi}.$$

Thus, X_{ij} contains the explanatory variables j (i.e., degree centrality and the covariates) and the error terms ε_{Pi} , ε_{URi} , and ε_{CRi} , which represent unobserved influences on participation probability, used reach, and the number of successful referrals.

Seeding strategies in first-generation models. In a first step, we restricted our analysis to first-generation models; we only considered referrals directly initiated by customers who received the seeding stimulus during the viral marketing campaign. Table 6 contains the parameter estimates of the PLHR model. The results can be interpreted in two stages: first, what drives the participation of seeded customers in the viral marketing campaign (logit component = LC) and, second, among these participants, what influences the number of referrals and successful referrals (Poisson component = PC). With regard to the covariates' impact on program participation, we find significant effects of the demographic variables gender ($\beta_{2n}^{LC} = -.2171, p < .01$) and age ($\beta_{3n}^{LC} = -.0209, p < .01$), indicating that male and older customers are more likely to participate, as are customers with short customer tenures who have just recently adopted the service ($\beta_{8n}^{LC} = -.0016, p < .01$). The latter finding aligns with cognitive dissonance theory, in that these customers might communicate shortly after their purchase decision to reduce dissonance (Festinger 1957). Furthermore, we found that customers acquired online are more strongly engaged in the (online-based) referral program than customers acquired through the retail channel ($\beta_{6n}^{LC} = -.9843, p < .01$). A one-price tariff seems easier to communicate; seeded customers with that tariff option are more likely to participate in the referral program ($\beta_{7n}^{LC} = -.1433, p < .01$). With regard to the usage covariates, we find positive and significant values for monthly airtime ($\beta_{9n}^{LC} = .0007, p < .01$) and monthly SMS ($\beta_{10n}^{LC} = .0010, p < .01$). The influences of most covariates are comparable between the logit and Poisson regression stages, except the acquisition channel, in that retail customers are less likely to participate in the program, but if they do, they exhibit significantly greater

²We choose PLHR over ZIP because the logit stage of the former is designed to determine what leads to participation (i.e., identifying referrers, in which we are interested), whereas the inflation stage of ZIP tries to detect "sure zeros" (i.e., nonparticipants).

activity than online customers ($\beta_{6n}^{PC} = .7388, p < .01$). The same applies to the usage covariates: Here, monthly airtime ($\beta_{9n}^{PC} = -.0007, p < .1$) and monthly SMS ($\beta_{10n}^{PC} = -.0018, p < .01$) show negative effects on participation.

However, the influence of (external) degree centrality varies between the stages of the model. In the logit regression stage, degree centrality has a positive and significant influence on the likelihood to participate P_i in the referral program ($\beta_{1n}^{LC} = .0032, p < .01$). Confirming the results of Studies 1 and 2, this finding shows that customers with high degree centrality are more likely to participate than those with low degree centrality (average degree centrality of participants = 45.3 vs. nonparticipants = 36.5). However, in the Poisson regression stage that analyzes only the group of active referrers, the effect of degree centrality is mixed. We find a significant, positive effect on used reach ($\beta_{1n}^{PC} = .0012, p < .01$), such that customers with high-degree centrality are not only more likely to participate but also more active when participating in the viral marketing campaign. However, we find no significant effect of degree

centrality on the referral success of active referrers ($\beta_{1CR}^{PC} = -.0002, n.s.$).

This result is further confirmed when we analyze the mean conversion rate of referrals per referrer $w_i = CR_i/n_i^*$ (see Table 7). Again, we do not find a significant effect of degree centrality for active referrers ($\beta_{1w} = .0001, n.s.$). Thus, our results offer no support for the assumption that participating central customers are more persuasive referrers or better selectors of potential referral targets.

Next, considering that viral marketing campaigns can be costly, we attempt to identify customers who are most likely to participate and generate (successful) referrals. We use the estimated participation probability calculated from the results of the selection model (see Table 6, “Logit Component”) to group the full customer base into cohorts and then compare these cohorts according to their observed participation, referral, and conversion rates and degree centrality (see Table 8). The top 5000 cohort corresponds to a high-degree and the bottom 5000 cohort to a low-degree seeding strategy; the results in the “Average” column correspond to random seeding.

TABLE 6
Determinants of Number of Referrals, Number of Successful Referrals, and Influence Domain
(Poisson-Logit Hurdle Regression Models, Study 3)

		Used Reach n^*		Converted Reach $CR = (n^*w)^*$		Conditional Influence Domain ID_i^T	
Variable		Coefficient	SE	Coefficient	SE	Coefficient	SE
Logit Component							
Degree centrality	β_1	.0022***	.0003	.0021***	.0004	.0021***	.0004
<i>Covariates</i>							
Gender	β_2	-.2287***	.0318	-.2527***	.0337	-.2527***	.0337
Age	β_3	-.0202***	.0013	-.0190***	.0014	-.0189***	.0014
Payment type	β_4	.0913**	.0389	.0668 ^{n.s.}	.0412	.0668 ^{n.s.}	.0412
Refill policy	β_5	-.0122 ^{n.s.}	.0376	.0174 ^{n.s.}	.0396	.0174 ^{n.s.}	.0396
Acquisition channel	β_6	-.9848***	.0506	-1.0535***	.0545	-1.0530***	.0545
Tariff plan	β_7	-.1371***	.0395	-.1253***	.0419	-.1253***	.0420
Customer tenure	β_8	-.0016***	.0001	-.0016***	.0001	-.0016***	.0001
Monthly airtime	β_9	.0007***	.0002	.0007***	.0002	.0007***	.0002
Monthly SMS	β_{10}	.0010***	.0002	.0009***	.0002	.0010***	.0002
Intercept		-2.3825***	.0727	-2.534***	.0770	-2.5341***	.0770
Poisson Component							
Degree centrality	β_1	.0026***	.0006	.0001 ^{n.s.}	.0033	-.0025***	.0007
<i>Covariates</i>							
Gender	β_2	-.1539***	.0519	-.1177 ^{n.s.}	.2206	-.3323***	.0496
Age	β_3	-.0098***	.0020	-.0104 ^{n.s.}	.0087	-.0112***	.0019
Payment type	β_4	-.3908***	.0581	-.1807 ^{n.s.}	.2492	-.1379**	.0544
Refill policy	β_5	-.3417***	.0761	-.1369 ^{n.s.}	.2819	.0033 ^{n.s.}	.0608
Acquisition channel	β_6	.7408***	.0561	.5902**	.2592	.6273***	.0548
Tariff plan	β_7	-1.0740***	.0473	-1.1426***	.2067	-.5322***	.0466
Customer tenure	β_8	-.0007***	.0002	-.0007 ^{n.s.}	.0007	-.0013***	.0001
Monthly airtime	β_9	-.0007*	.0004	-.0000 ^{n.s.}	.0000	.0000 ^{n.s.}	.0003
Monthly SMS	β_{10}	-.0018***	.0005	-.0001 ^{n.s.}	.0019	.0005 ^{n.s.}	.0004
Intercept		.9811***	.0941	-1.4633***	.4181	1.1008***	.0905
Log-likelihood value		-25,163		-19,850		-23,723	
BIC		50,596		39,969		47,714	
N				208,829			

* $p < .1$.

** $p < .05$.

*** $p < .01$.

Notes: Two-tailed significance levels. n.s. = not significant.

TABLE 7
Determinants of Conversion Rates, Active Referrers (Poisson Regression, Study 3)

Variable		Poisson Regression Model	
		Conversion $w = CR/n^*$	
		Coefficient	SE
Degree centrality	β_1	-.0001 ^{n.s.}	.0004
Covariates			
Gender	β_2	.0788**	.0350
Age	β_3	.0047***	.0015
Payment type	β_4	.1104***	.0431
Refill policy	β_5	.1079***	.0409
Acquisition channel	β_6	-.4951***	.0616
Tariff plan	β_7	.4638***	.0472
Customer tenure	β_8	.0002*	.0001
Intercept		-1.2014***	.0848
Log-likelihood value		-5,074	
N		4,549	

* $p < .1$.

** $p < .05$.

*** $p < .01$.

Note: Two-tailed significance levels. n.s. = not significant. For the Poisson regression model, we used $\ln(n)$ as offset variable.

The results in Table 8 clearly confirm the positive correlation between degree centrality and the success of viral marketing: As the estimated participation probability increases, observed participation, referral, and conversion rates (i.e., total number of participants, referrals, or successful referrals divided by number of seeded customers in the cohort) and degree centrality increase as well. Thus, the participation rate of the top 5000 cohort is a multiple of that of the bottom 5000 cohort (4.4% vs. .5%), with a much higher average degree centrality (70.8 vs. 18.0). A high-degree seeding strategy would be nearly nine times as successful as a low-degree strategy. Compared with the average value of a random strategy, the top 5000 cohort participation and degree centrality are twice as high; therefore, targeting hubs doubles the performance of random seeding for a sample of the same size. Thus, Study 3 clearly shows the significant, positive effect of degree centrality on viral marketing participation and activity, in strong support of a

high-degree seeding strategy. However, the results of the Poisson regression model do not indicate higher referral success of hubs within the group of active referrers.

Seeding strategies in multiple-generation models. In the second step of our empirical analysis, we extended the measure of success to account for a fuller range of the effects of seeding efforts by including more than one generation of referrals. First-generation referrals initiate a viral process that should continue in further generations. The extent of this viral branching may differ across seeding strategies, due to their ability to reach different parts of the social network. Thus, the optimal seeding strategy might change if we consider multiple generations.

To capture this form of success, we measured all subsequent referrals that originated from a first-generation referral during the campaign. We limited the observation period to 12 months because the company repeated the referral campaign 13 months after the initial seeding. During our observation period, the company did not engage in other promotions that directly focused on referrals, nor did we find any anomalies (e.g., drastic increases or decreases) in company-owned or competitive marketing spending.

In the first year after the campaign, 20.8% of all first-generation referrals became active referrers themselves, and 5.8% did so multiple times. We observed viral referral chains with a maximum length of 29 generations; on average, every first-generation referral during the campaign led to .48 additional referrals.

The dependent variable for this analysis is the influence domain of all successful referrals of a specific first-generation customer, which equals the number of successful first-generation referrals, plus the number of successful referrals in successive generations during the subsequent 12 months.³ For example, Figure 2 depicts the influence domain of a referral customer X that spans 22 additional successful referrals over seven generations.

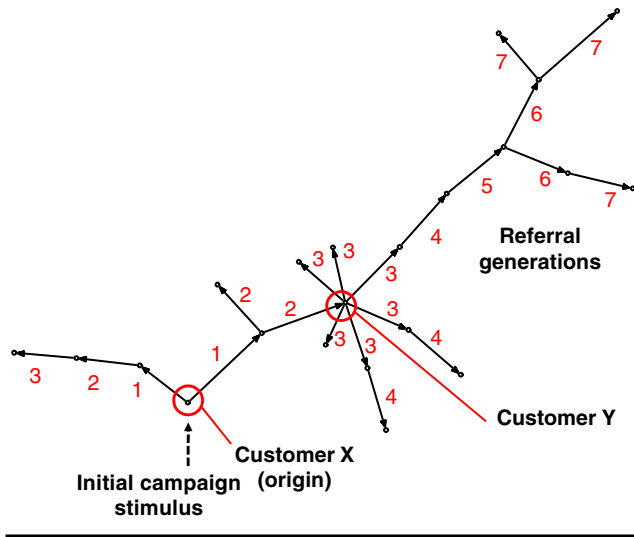
³Note that, by definition, there is no overlap of influence domains between two origins; every referred customer has an in-degree of 1, and only one specific referrer is rewarded for every new customer.

TABLE 8
Relationship of Conversion Rates and Degree Centrality, Full Sample (Study 3)

	Customer Cohort (According to Estimated Participation Probabilities)					Average
	Top 5000	Top 10,000	Top 20,000	Top 50,000	Bottom 5000	
Participants						
Total participation	220	378	671	1385	24	—
Participation rate	4.4%	3.8%	3.4%	2.8%	.5%	2.2%
Referrals						
Total referrals	292	489	856	1783	26	—
Referral rate	5.8%	4.9%	4.3%	3.6%	.5%	3.0%
Successful referrals						
Total conversions	191	330	598	1233	19	—
Conversion rate	3.8%	3.3%	3.0%	2.5%	.4%	2.0%
Average degree centrality	70.83	60.21	52.13	45.42	18.01	36.48

Notes: Top (Bottom) 5,000/10,000/... refer to the cohort of customers with the highest (lowest) estimated participation probabilities, according to the coefficient estimates of the logit component reported in Table 6. We calculated rates by dividing the total number of participants, referrals, or successful referrals by the total number of customers in the cohort.

FIGURE 2
Influence Domain of a Referral Campaign
Participant (Study 3)



The parameter estimates of the PLHR model for this multiple-generation model appear in the right-hand column of Table 6. The dependent variable ID_i^T is conditional on program participation ($P_i = 1$). When we compare the regression model parameters across the different dependent variables, we find similar results for Influence Domain ID_i^T and Used Reach n_i (e.g., no significant effect of the usage covariates) but with one important difference: Our focal variable, degree centrality, is negative ($\beta_{ID}^{PC} = -.00205$, $p < .01$) in the Poisson regression model. That is, among the participants, more central customers have a smaller influence domain. The observed network structure of the referral processes offers a potential explanation of this surprising

result. For hubs, we mostly observe short referral chains (if at all), whereas fringe customers who participate in the viral marketing campaign demonstrate significantly longer referral chains. For example, in Figure 2, the fringe customer X reacts to the campaign and refers the service. Within two generations, this referral reaches actor Y, who initiates a total of 15 additional referrals, which increases the influence domain of X to 22.

Because we find a positive effect of high degree centrality in the selection model but a negative effect in the regression model, the overall effect of degree centrality remains unclear. For a simple test, we performed both an ordinary least squares (OLS) and a simple Poisson regression, with the unconditional Influence Domain ID_i^R as the dependent variable for the complete sample of 208,829 customers who received the viral marketing campaign stimulus. According to the results in Table 9, the standardized beta for degree centrality is positive and significant in both regressions ($\beta_{ID}^{OLS} = .010$, $\beta_{ID}^{PR} = .002$; $p < .001$), such that the overall effect of high degree centrality as a selection criterion for seeding a viral marketing campaign is positive. Therefore, high-degree seeding remains the more successful strategy, even if we account for a multiple-generation viral process.

Robustness checks. To check the robustness of our findings, we also analyzed our data with a set of alternative approaches, including a Poisson-based (ZIP regression) and probit-OLS combinations (e.g., Tobit Type II models). Our core results pertaining to the influence of degree centrality, including its positive influence on the selection stage to determine the likelihood of participation and its negative effect on the influence domain in the regression stage, hold across all tested models. The results also remain unchanged when we incorporate alternative individual-level covariates, such as monthly mobile charges, that represent the attractiveness of a customer to the provider.

Unlike Studies 1 and 2, Study 3 does not allow us to assess the causal effect of the seeding strategy on referral

TABLE 9
Determinants of Unconditional Influence Domain (OLS Model, Study 3)

Variable		OLS Model Unconditional Influence Domain (ID_i^R)		Poisson Regression Model Unconditional Influence Domain (ID_i^R)	
		Standard Coefficient	SE	Coefficient	SE
Degree centrality	β_1	.010***	.000	.002***	.000
Covariates					
Gender	β_2	-.015***	.002	-.358***	.028
Age	β_3	-.024***	.000	-.024***	.001
Payment type	β_4	.001 ^{n.s.}	.002	.013 ^{n.s.}	.033
Refill policy	β_5	.000 ^{n.s.}	.002	.003 ^{n.s.}	.033
Acquisition channel	β_6	-.025***	.002	-.680***	.039
Tariff plan	β_7	-.016***	.002	-.433***	.030
Customer tenure	β_8	-.031***	.004	-.002***	.000
Intercept		.091***	.004	-1.509***	.058
R^2 (pseudo)		.05		.03	
N		208,829		208,829	

* $p < .1$.

** $p < .05$.

*** $p < .01$.

Notes: Two-tailed significance levels. n.s. = not significant.

success unambiguously. However, it reflects a real-life marketing application and is based on detailed firm data, such that it strikingly illustrates the power of network information in real-life situations.

General Discussion

Research Contribution

Inspired by conflicting recommendations in previous studies regarding optimal seeding strategies for viral marketing campaigns, this article empirically compares the performance of various proposed strategies, examines the magnitude of differences, and identifies determinants that are responsible for the superiority of a particular seeding strategy. To the best of our knowledge, such an experimental comparison of seeding strategies is unprecedented; previous literature is based solely on mathematical models and computer simulations. Our real-life application provides some answers to controversies about whether hubs are harder to convince, whether they make use of their reach, and whether they are more persuasive.

Marketers can achieve the highest number of referrals, across various settings, if they seed the message to hubs (high-degree seeding) or bridges (high-betweenness seeding). These two strategies yield comparable results and both outperform the random strategy (+52%) and are up to eight times more successful than seeding to fringes (low-degree seeding). The superiority of the high-degree seeding strategy does not rest on a higher conversion rate due to a higher persuasiveness of hubs but rather on the increased activity of hubs, which is in line with previous findings (e.g., Iyengar, Van den Bulte, and Valente 2011; Scott 2000). This finding is persistent even when we control for revenue of customers and thus demonstrates the importance of social structure beyond customer revenues and customer loyalty.

Research that suggests a low-degree seeding strategy is usually based on the central assumption that highly connected people are more difficult to influence than less connected people because highly connected people are subject to the influence of too many others (e.g., Watts and Dodds 2007). Our results (Studies 1 and 3) reject this assumption and underline Becker's (1970) suggestion: Hubs are more likely to engage because viral marketing works mostly through awareness caused by information transfer from previous adopters and through belief updating, especially for low-risk products. The low perceived risk means hubs do not hesitate before participating. Furthermore, when social contagion occurs mostly at the awareness stage, the possible disproportionate persuasiveness of hubs is irrelevant. As long as the social contagion occurs at the awareness stage through simple information transfer, hubs are not more persuasive than other nodes (Godes and Mayzlin 2009; Iyengar, Van den Bulte, and Valente 2011).

Our analysis of a mobile service provider's viral marketing campaign reveals that hubs make slightly more use of their reach potential. Furthermore, for the group of participating customers, we find a negative influence of greater connectivity on the resulting influence domains. Although in epidemiology studies infectious diseases spread through hubs, we find that well-connected people do not use their

greater reach potential fully in a marketing setting. Spreading information is costly in terms of both time invested and the effort needed to capture peers' attention. Furthermore, hubs may be less likely to reach other previously unaffected central actors, such that they are limited in their overall influence domain. The compelling findings from sociology and epidemiology thus appear to have been incorrectly transferred to targeting strategies in viral marketing settings.

Nevertheless, the social network remains a crucial determinant of optimal seeding strategies in practice because a social structure is much easier to observe and measure than communication intensity, quality, or frequency. Furthermore, we find robust results even when we control for the level of communication activity. Therefore, companies should use social network information about mutual relationships to determine their viral marketing strategy.

Managerial Implications

Viral marketing is not necessarily an art rather than a science; marketers can improve their campaigns by using sociometric data to seed their viral marketing campaigns. Our multiple studies show that information about the social structure is valuable, in that seeding the "right" consumers yields up to eight times more referrals than seeding the "wrong" ones. In contrast with random seeding, seeding hubs and bridges can easily increase the number of successful referrals by more than half. Thus, it is essential for marketers to adopt an appropriate seeding strategy and use sociometric data to increase their profits. We conclude that adding metrics related to social positions to customer relationship management databases is likely to improve targeting models substantially.

Many companies already have implicit information about social ties that they could use to calculate explicit sociometric measures. Telecommunication providers can exploit connection data (as we did in Study 3), banks possess data about money transfers, e-mail providers might analyze e-mail exchanges, and companies can evaluate behaviors in company-owned forums. Many companies also have indirect access to information on social networks, such as Microsoft through Skype or Google through its Google mail services, and they could begin to use information obtained this way (e.g., Hill, Provost, and Volinsky 2006; Aral, Muchnik, and Sundararajan 2009). Such network information is further available in the form of friendship data obtained from online social communities such as Facebook or LinkedIn (e.g., Hinz and Spann 2008).

Remarkably, we reveal that to target a particular subnetwork (e.g., students of a particular university, Study 2) with a viral marketing message, the use of the respective subnetwork's sociometric measures is not absolutely required to implement the desired seeding strategies. Instead, because the sociometric measures of subnetworks and their total network are highly correlated, marketers can use the sociometric measures of the total network, without undertaking the complex task of determining exact network boundaries. Conversely, this appealing result also allows marketers to feel confident in inferring the connectivity of a person in an overall network from information about his or her connectivity in a natural subnetwork. Moreover, because betweenness centrality requires knowledge about the structure of

the entire network, as well as a complex, time-consuming computation, degree centrality seems to be the best sociometric measure for marketing practice (see also Kiss and Bichler 2008).

According to these insights, marketers should pick highly connected persons as initial seeds if they hope to generate awareness or encourage transactions through their viral marketing campaigns because these hubs promise a wider spread of the viral message. As long as the social contagion operates at the awareness stage through information transfer, we do not observe that hubs are more persuasive. Thus, the sociometric measure of degree centrality cannot be used to identify persuasive seeding points. The use of demographics and product-related characteristics (see Table 7) seems more promising for this purpose. Study 1 reveals that monetary incentives for referring strongly increase the spread of viral marketing messages, which supports the use of such incentives. However, they would also make viral marketing more costly than is commonly assumed.

Finally, expertise in the domain of social networks is valuable for seeding purposes. Thus, online communities such as Facebook might begin to offer information on members' social positions to third-party marketers or provide the option to seed to a specific target group according to sociometric measures. Specialized service providers might adopt a similar idea to tailor their offerings with respect to optimal seeding. Business models might reflect social network information collected from communication relationships or domain expertise in certain subject domains, such that they target the most highly connected or intermediary persons with specific marketing information and thus maximize the success of viral marketing campaigns. For example, the Procter & Gamble subsidiaries Vocal point and Tremor already use social network information to introduce new products, enjoying doubled sales in some test locations.

Limitations and Directions for Further Research

Although we designed our experiments carefully, some of their shortcomings might limit the validity of the results. In Study 1, order effects may exist because the sets of participants in the different experimental settings are not disjunctive. We designed Study 2 to avoid such an overlap, but the parallel timing in that experiment could lead to interrelationships across the different seeding strategies. However, as we noted previously, seeding strategies that lead to faster diffusion might have performed better, which also reflects reality well: In a world in which people become

increasingly exposed to multiple viral campaigns competing for their attention and participation, seeding strategies that lead to faster diffusion may be advantageous for the initiator of the particular viral marketing campaign.

The student sample and the artificial information content of the experiments constitute additional limitations in our experimental studies, though these shortcomings do not systematically favor one strategy over another. It would be useful to conduct similar experiments with different samples and less artificial information, such as real-life viral marketing campaigns.

In our real-life application, we could compare only high- and low-degree and random seeding strategies. We did not differentiate referrals with regard to the profit of the referred customers (for an analysis of the value of customers acquired through referral programs, see Schmitt, Skiera, and Van den Bulte 2011), but our robustness checks show no significant regression-stage effects of additional covariates on converted reach or influence domain. Our results regarding the effects of degree centrality also remain unchanged across the robustness checks.

Most current research focuses on individual choices and treats the choices of partners (within the social network) as exogenous; as we do, these studies assume that the network remains fixed for the duration of the study and unaffected by it. However, this strong assumption ignores the likely effects of dynamics inherent to real-life social networks (e.g., Steglich, Snijders, and Pearson 2010). Marketing response models that incorporate these effects would be a fruitful avenue for further research. Another extension might incorporate information on the dyadic level, such as tie strength, which could distinguish the success of referrals from a specific customer. In addition, from a marketing perspective, it would be useful to intensify research regarding the effects of incentives, as we find their impact to be quite significant in Study 1 (e.g., Schmitt, Skiera, and van den Bulte 2011; Aral, Muchnik, and Sundararajan 2011).

Our combination of two experimental studies and an *ex post* analysis of a real-world viral marketing campaign provides a strong argument that hubs and bridges are key to the diffusion of viral marketing campaigns. We cannot confirm recent findings that question the exposed role of hubs for the success of viral marketing campaigns, but our analysis of the different determinants in Study 3 yields additional explanations for why hubs make more attractive seeding points. These findings should serve as inputs to create more realistic computer simulations and analytical models.

REFERENCES

- Anderson, Roy M. and Robert M. May (1991), *Infectious Diseases of Humans: Dynamics and Control*. New York: Oxford University Press.
- Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009), "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences*, 106 (51), 21544–49.
- , ———, and ——— (2011), "Engineering Social Contagions: Optimal Network Seeding and Incentive Strategies," (accessed July 21, 2011), [available at <http://ssrn.com/abstract=1770982>].
- Arndt, Johan (1967), "Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research*, 4 (August), 291–95.
- Bampo, Mauro, Michael T. Ewing, Dineli R. Mather, David Stewart, and Mark Wallace (2008), "The Effects of the Social Structure of Digital Networks on Viral Marketing Performance," *Information Systems Research*, 19 (3), 273–90.

- Becker, Marshall H. (1970), "Sociometric Location and Innovativeness: Reformulation and Extension of the Diffusion Model," *American Sociological Review*, 35 (2), 267–82.
- Berger, Jonah and Katy Milkman (2011), "Social Transmission, Emotion, and the Virality of Online Content," (accessed July 21, 2011), [available at <http://marketing.wharton.upenn.edu/documents/research/Virality.pdf>].
- and Eric Schwartz (2011), "What Do People Talk About? Drivers of Immediate and Ongoing Word-of-Mouth," *Journal of Marketing Research*, forthcoming.
- Biyalogorsky, Eyal, Eitan Gerstner, and Barak Libai (2001), "Customer Referral Management: Optimal Reward Programs," *Marketing Science*, 20 (1), 82–95.
- Burt, Ronald S. (1987), "Social Contagion and Innovation: Cohesion Versus Structural Equivalence," *American Journal of Sociology*, 92 (6), 1287–1335.
- Coleman, James S., Elihu Katz, and Herbert Menzel (1966), *Medical Innovation: A Diffusion Study*. Indianapolis: Bobbs-Merrill.
- Costenbader, Elizabeth and Thomas W. Valente (2003), "The Stability of Centrality Measures When Networks Are Sampled," *Social Networks*, 25 (4), 283–307.
- De Bruyn, Arnaud and Gary L. Lilien (2008), "A Multi-Stage Model of Word-of-Mouth Influence Through Viral Marketing," *International Journal of Research in Marketing*, 25 (3), 151–63.
- Dodds, Peter S. and Duncan J. Watts (2004), "Universal Behavior in a Generalized Model of Contagion," *Physical Review Letters*, 92 (21), 1–5.
- Festinger, Leon (1957), *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Galeotti, Andrea and Sanjeev Goyal (2009), "Influencing the Influencers: A Theory of Strategic Diffusion," *RAND Journal of Economics*, 40 (3), 509–532.
- Gladwell, Malcolm (2002), *The Tipping Point: How Little Things Can Make a Big Difference*. Boston: Back Bay Books.
- Godes, David and Dana Mayzlin (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Study," *Marketing Science*, 28 (4), 721–39.
- Goldenberg, Jacob, Sangman Han, Donald R. Lehmann, and Jae Weon Hong (2009), "The Role of Hubs in the Adoption Process," *Journal of Marketing*, 73 (2), 1–13.
- Granovetter, Mark S. (1973), "The Strength of Weak Ties," *American Journal of Sociology*, 78 (6), 1360–80.
- (1985), "Economic Action and Social Structure: The Problem of Embeddedness," *American Journal of Sociology*, 91 (3), 481–510.
- Hanaki, Nobuyuki, Alexander Peterhansl, Peter S. Dodds, and Duncan J. Watts (2007), "Cooperation in Evolving Social Networks," *Management Science*, 53 (7), 1036–50.
- Hann, Il-Horn, Kai-Lung Hui, Sang-Yong T. Lee, and Ivan P.L. Png (2008), "Consumer Privacy and Marketing Avoidance: A Static Model," *Management Science*, 54 (6), 1094–1103.
- Hill, Shawndra, Foster Provost, and Chris Volinsky (2006), "Network-Based Marketing: Identifying Likely Adopters via Consumer Networks," *Statistical Science*, 21 (2), 256–76.
- Hinz, Oliver and Martin Spann (2008), "The Impact of Information Diffusion on Bidding Behavior in Secret Reserve Price Auctions," *Information Systems Research*, 19 (3), 351–68.
- Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30 (2), 195–212.
- Iyengar, Sheena S. and Mark Lepper (2000), "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, 76 (6), 995–1006.
- Jain, Dipak, Vijay Mahajan, and Eitan Muller (1995), "An Approach for Determining Optimal Product Sampling for the Diffusion of a New Product," *Journal of Product Innovation Management*, 12 (2), 124–35.
- Kalish, Shlomo, Vijay Mahajan, and Eitan Muller (1995), "Waterfall and Sprinkler New-Product Strategies in Competitive Global Markets," *International Journal of Research in Marketing*, 12 (2), 105–119.
- Kemper, John T. (1980), "On the Identification of Superspreaders for Infectious Disease," *Mathematical Biosciences*, 48 (1/2), 111–27.
- Kiss, Christine and Martin Bichler (2008), "Identification of Influencers—Measuring Influence in Customer Networks," *Decision Support Systems*, 46 (1), 233–53.
- Lambert, Diane (1992), "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing," *Technometrics*, 34 (1), 1–14.
- Leskovec, Jure, Lada A. Adamic, and Bernardo A. Huberman (2007), "The Dynamics of Viral Marketing," *ACM Transactions on the Web*, 1 (1), 228–37.
- Libai, Barak, Eitan Muller, and Renana Peres (2005), "The Role of Seeding in Multi-Market Entry," *International Journal of Research in Marketing*, 22 (4), 375–93.
- Lin, Nan (1976), *Foundations of Social Research*. New York: McGraw-Hill.
- Mullahy, John (1986), "Specification and Testing of Some Modified Count Data Models," *Journal of Econometrics*, 33 (3), 341–65.
- Porter, Constance E. and Naveen Donthu (2008), "Cultivating Trust and Harvesting Value in Virtual Communities," *Management Science*, 54 (1), 113–28.
- Porter, Lance and Guy J. Golan (2006), "From Subservient Chickens to Brawny Men: A Comparison of Viral Advertising to Television Advertising," *Journal of Interactive Advertising*, 6 (2), 30–38.
- Rayport, Jeffrey (1996), "The Virus of Marketing," *Fast Company* (accessed March 24, 2011), [available at <http://www.fastcompany.com/magazine/06/virus.html>].
- Schmitt, Philipp, Bernd Skiera, and Christophe Van den Bulte (2011), "Referral Programs and Customer Value," *Journal of Marketing*, 75 (January), 46–59.
- Scott, John (2000), *Social Network Analysis: A Handbook*, 2d ed. London: Sage Publications.
- Simmel, Georg (1950), *The Sociology of Georg Simmel*, compiled and translated by Kurt Wolff. Glencoe, IL: The Free Press.
- Steglich, Christian, Tom A.B. Snijders, and Michael Pearson (2010), "Dynamic Networks and Behavior: Separating Selection from Influence," *Sociological Methodology*, 40 (1), 329–93.
- Sundararajan, Arun (2006), "Network Seeding," Extended Abstract for WISE 2006, New York University Stern School of Business.
- Thompson, Clive (2008), "Is the Tipping Point Toast?" *Fast Company*, 122, 74–105, (accessed March 24, 2011), [available at <http://www.fastcompany.com/magazine/122/is-the-tipping-point-toast.html>].
- Toubia, Olivier, Andrew T. Stephen, and Aliza Freud (2010), "Identifying Active Members in Viral Marketing Campaigns," (accessed July 21, 2011), [available at http://sites.google.com/a/andrewstephen.net/andrewresearch/papers/Toubia_Stephen_Freud_viralmarketing_nov2010.pdf].
- Trusov, Michael, Anand Bodapati, and Randolph E. Bucklin (2010), "Determining Influential Users in Internet Social Networks," *Journal of Marketing Research*, 47 (August), 643–58.

- Van den Bulte, Christophe (2010), "Opportunities and Challenges in Studying Consumer Networks," in *The Connected Customer*, S. Wuyts, M.G. Dekimpe, E. Gijsbrechts, and R. Pieters, eds. London: Routledge, 7–35.
- and Yogesh V. Joshi (2007), "New Product Diffusion with Influentials and Imitators," *Marketing Science*, 26 (3), 400–421.
- and Gary L. Lilien (2001), "Medical Innovation Revisited: Social Contagion Versus Marketing Effort," *American Journal of Sociology*, 106 (5), 1409–1435.
- and Stefan Wuyts (2007), *Social Networks and Marketing*. Cambridge, MA: MSI Relevant Knowledge Series.
- Van der Lans, Ralf, Gerrit van Bruggen, Jehoshua Eliashberg, and Berend Wierenga (2010), "A Viral Branching Model for Predicting the Spread of Electronic Word-of-Mouth," *Marketing Science*, 29 (2), 348–65.
- Watts, Duncan J. (2004), "The 'New' Science of Networks," *Annual Review of Sociology*, 30, 243–70.
- and Peter S. Dodds (2007), "Influentials, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34 (4), 441–58.
- and Jonah Peretti (2007), "Viral Marketing for the Real World," *Harvard Business Review*, 85 (5), 22–23.