



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Competition for Attention in Online Social Networks: Implications for Seeding Strategies

Sarah Gelper, Ralf van der Lans, Gerrit van Bruggen

To cite this article:

Sarah Gelper, Ralf van der Lans, Gerrit van Bruggen (2021) Competition for Attention in Online Social Networks: Implications for Seeding Strategies. Management Science 67(2):1026-1047. <https://doi.org/10.1287/mnsc.2019.3564>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2020, The Author(s)

Please scroll down for article—it is on subsequent pages





With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Competition for Attention in Online Social Networks: Implications for Seeding Strategies

 Sarah Gelper,^a Ralf van der Lans,^b Gerrit van Bruggen^c

^a School of Industrial Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands; ^b Department of Marketing, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; ^c Department of Marketing, Rotterdam School of Management, Erasmus University, 3000 DR Rotterdam, Netherlands

Contact: s.gelper@tue.nl,  <https://orcid.org/0000-0003-2346-4054> (SG); rrlans@ust.hk,  <https://orcid.org/0000-0002-7726-8238> (RvdL); gbruggen@rsm.nl (GvB)

Received: December 12, 2016

Revised: July 2, 2018; June 12, 2019; October 17, 2019

Accepted: December 2, 2019

Published Online in Articles in Advance: June 30, 2020

<https://doi.org/10.1287/mnsc.2019.3564>

Copyright: © 2020 The Author(s)

Abstract. Many firms try to leverage consumers' interactions on social platforms as part of their communication strategies. However, information on online social networks only propagates if it receives consumers' attention. This paper proposes a seeding strategy to maximize information propagation while accounting for competition for attention. The theory of exchange networks serves as the framework for identifying the optimal seeding strategy and recommends seeding people that have many friends, who, in turn, have only a few friends. There is little competition for the attention of those seeds' friends, and these friends are therefore responsive to the messages they receive. Using a game-theoretic model, we show that it is optimal to seed people with the highest Bonacich centrality. Importantly, in contrast to previous seeding literature that assumed a fixed and non-negative connectivity parameter of the Bonacich measure, we demonstrate that this connectivity parameter is negative and needs to be estimated. Two independent empirical validations using a total of 34 social media campaigns on two different large online social networks show that the proposed seeding strategy can substantially increase a campaign's reach. The second study uses the activity network of messages exchanged to confirm that the effects are driven by competition for attention.

History: Accepted by Anandhi Bharadwaj, information systems.



Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Management Science. Copyright © 2020 The Author(s). <https://doi.org/10.1287/mnsc.2019.3564>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/mnsc.2019.3564>.

Keywords: social networks • information propagation • seeding strategies • networks • graphs • marketing • advertising and media

1. Introduction

In the last decade, a growing number of companies and organizations have initiated marketing campaigns leveraging online social interactions. In 2014, the *First Kiss* video by the clothing brand Wren gathered more than 42 million YouTube views in three days¹ and increased sales by nearly 14,000%.² Other well-known examples of successful campaigns are the 2013 *Dove Real Beauty Sketches* video, which gathered more than 114 million views within one month,³ and the 2012 *Kony* video bringing criminal issues in Africa to the attention of the public. Watching such messages increases people's engagement with products and brands, which, in turn, increases their profitability for the firm (Rishika et al. 2013, Kumar et al. 2016), especially when many people are reached. However, in many situations, information does not spread easily on social networks (Sun et al. 2009, Bakshy et al. 2011, Feng et al. 2015), resulting in only a few campaigns that go truly viral (Watts and Peretti 2007, Goel et al. 2012).

The enormous amount of information that is shared on social networks is an important explanation for why most campaigns do not go viral (Asur et al. 2011, Berger and Milkman 2012, Weng et al. 2012). Users can pay attention to only a subset of all the information they receive, and the more they receive, the less likely it is that they will pay attention to any specific message.

To initiate a campaign that reaches many people, a firm first needs to define a seeding strategy. A seeding strategy involves the identification of a small number of key individuals that maximizes the reach in the social network (Hinz et al. 2011, Aral and Dhillon 2018). Several recent studies deal with identifying these key individuals, in the context of both information propagation (e.g., Yoganarasimhan 2012, Goel et al. 2016, Chen et al. 2017) and new-product adoption (e.g., Bakshy et al. 2009, Goldenberg et al. 2009, Katona et al. 2011, Aral et al. 2013). As argued by Centola (2010), an essential distinction between

information propagation and new-product adoption is that information propagation tends to follow a simple process in which one individual is sufficient to pass information on, typically modeled by using a cascade model (Aral and Dhillon 2018). In contrast, product adoption is more complex because it is influenced by other factors, such as prices and network externalities, and it often requires information of multiple connections to reinforce adoption decisions (Aral and Walker 2011, Iyengar et al. 2011, Aral et al. 2013, Aral and Dhillon 2018). Our research deals with information propagation and thus contributes to the seeding literature on “simple” processes that require limited reinforcement. Empirical studies in this literature highlight the importance of seeding well-connected network members, because they are able to reach many individuals quickly (Hinz et al. 2011, Chen et al. 2017). However, effective seeds not only have many friends, but their friends should also be susceptible to incoming information (Watts and Dodds 2007, Aral and Dhillon 2018). Using randomized experiments on large social networks, Aral and Walker (2012) were able to measure susceptibility and influence of network members and demonstrated their importance in propagation processes. However, they did not consider how this affects optimal seeding strategies and called for future research to examine this. Our research aims to take the next step into this policy-relevant stream of research by deriving the optimal seeding strategy considering competition for attention as a facet of susceptibility. Individuals who receive many messages face stronger competition for attention and are therefore less likely to attend to a specific message and subsequently forward it (Asur et al. 2011, Weng et al. 2012, Iyer and Katona 2016). Moreover, because highly connected individuals receive, on average, more information (Aral and Van Alstyne 2011, Bapna and Umyarov 2015), competition for attention and, thus susceptibility, depends on network position.

To derive the optimal seeding strategy under competition for attention, we build on exchange-network theory that deals with competition in networks (Cook et al. 1983, Markovsky et al. 1988, Yamagishi et al. 1988, Blume et al. 2009). In exchange networks where scarce goods are traded, the most powerful members are those who have many potential trading partners but whose trading partners have only a few alternative trading partners. Analogously, in this paper, we argue that social network members who have many friends⁴ but whose friends have only few friends are able to obtain a high reach. Such network members are effective seeds because there is low competition for their friends’ attention, and these friends therefore have a higher likelihood to further

share the information they receive. Although exchange-network theory explicitly addressed such competition, this notion has been neglected in the seeding literature. We show that competition for attention has strong implications for the effectiveness of seeding strategies.

Our research aims to contribute in three ways. First, although previous research has considered competition for attention in online social networks (Weng et al. 2012), the implications for seeding strategies were not well understood. We are the first to derive an optimal seeding strategy under competition for attention and find that optimal seeding is achieved using the Bonacich centrality measure (Bonacich 1987) in which the connectivity parameter β can be negative. Second, previous empirical research on seeding effectiveness only considered two restricted special cases of Bonacich centrality: (1) degree centrality ($\beta = 0$) and (2) eigenvalue centrality ($\beta =$ inverse of largest eigenvalue of the adjacency matrix). Our paper is the first to propose that β can be negative and that this parameter therefore needs to be estimated. Third, in two empirical applications covering 34 different viral marketing campaigns on two social network platforms, we show that β is indeed negative. Taking into account negative values of β substantially improves seeding effectiveness compared with alternative seeding strategies, including the two special cases of Bonacich centrality that have been applied in the literature (i.e., degree centrality and eigenvalue centrality). Moreover, in the second empirical application, we observe the actual activities of network members, which allows us to test our proposed underlying mechanism of competition for attention. Our empirical results demonstrate the generalizability of our theoretical predictions, which have important practical implications for seeding.

We proceed as follows. First, we introduce our conceptual model and explain how network members who maximize information propagation can be identified. Using a game-theoretic model, we analytically derive the optimal seeding strategy. We validate the strategy in two independent empirical studies. In Study 1, we show that seeds with many friends, who, in turn, have few friends, on average, obtain a higher reach. An out-of-sample comparison demonstrates the substantial gains that can be achieved by applying the optimal seeding strategy derived from the theoretical model. Study 2 generalizes our findings from Study 1 for an additional 33 campaigns on a different social network and illustrates the mechanism of competition for attention. We conclude with a discussion of the main insights our research offers, their implications for seeding social media campaigns, and future research directions on the importance of competition for attention in online social networks.

2. Identifying Effective Seeds

A social media campaign starts with a company communicating a marketing message to (potential) customers, who may subsequently share the message with their friends on the social network, after which a repetitive sharing or viral process evolves (Bampo et al. 2008, De Bruyn and Lilien 2008, van der Lans et al. 2010). A campaign that successfully creates such a viral effect reaches many people after initially seeding only a few individuals in a network (Hinz et al. 2011, Aral and Dhillon 2018). To achieve this goal, it is important to understand which factors influence the information-propagation process, as graphically illustrated in Figure 1. First, the propagation process is driven by a firm's seeding strategy—that is, whom and how many network members to seed. Second, the propagation process is influenced by the properties of the network. Although previous seeding literature mostly focused on the role of network structure, summarized by centrality measures such as degree and eigenvector centrality (e.g., Hinz et al. 2011, Banerjee et al. 2013, Chen et al. 2017), it did not consider network connectivity and how information may compete for attention. We contribute to this literature by explicitly taking such competition for attention into account. We study how it affects sharing and derive implications for optimizing a firm's seeding strategy.

As illustrated in Figure 1, the firm's seeding strategy plays a crucial role in the propagation process of a social media campaign and, as a consequence, in the campaign's reach. Previous seeding research therefore tried to understand which members of a social network are important candidates to target in the seeding strategy (Hinz et al. 2011, Aral et al. 2013, Banerjee et al. 2013, Banerjee et al. 2019, Chen et al. 2017). This research has identified two network properties that contribute to information propagation (Libai et al. 2013). First, network members with many connections are more important because being highly connected enables them to contact many people directly (Goldenberg et al. 2009), and being highly connected

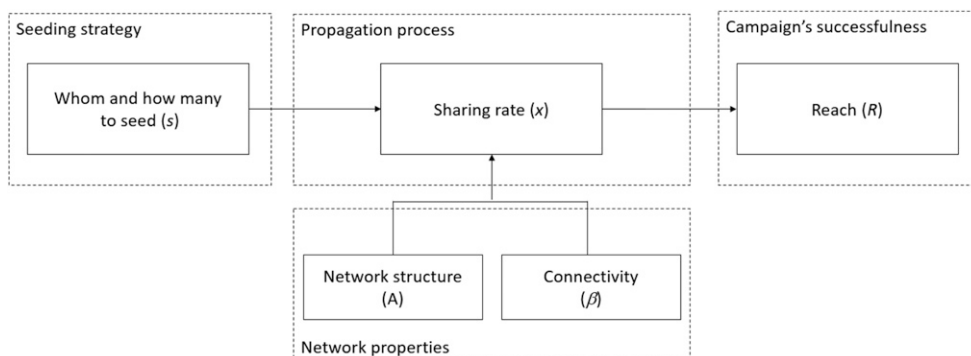
may increase their influence through status (Hu and Van den Bulte 2014, Lanz et al. 2019). Second, network members occupying a strategic network position, such as bridges connecting two subnetworks, are important for spreading information beyond local communities (Granovetter 1973, Burt 1997, Burt 2004, Tucker 2008, Valente 2012). These studies, however, do not investigate whether the responsiveness of receivers of campaign messages depends on their network positions, even though the receivers' responsiveness to new information is an important determinant of information propagation and product-adoption processes (Iyengar et al. 2011, Aral and Walker 2012, Ugander et al. 2012, Aral et al. 2013, Aral and Walker 2014). In this research, we propose a seeding strategy that considers both the connectedness of individuals and the attention and responsiveness of their friends, as we discuss next.

2.1. Competition for Attention in Social Networks

Consumers' attentional resources are limited and have been referred to as "the scarcest resource in today's business" (Pieters and Wedel 2004 p.36). With the growth of shared information on the web and on social networks, competition for attention has greatly increased in the last decade. As a consequence, gaining consumer attention is crucial for the success of marketing campaigns (Pieters et al. 2007, van der Lans et al. 2008). As illustrated by Berger and Milkman (2012), the location of news articles on the *New York Times* web page significantly influences the number of times such articles are shared. News articles that attract more attention, such as the ones presented on the top of a web page, are shared more often—even after controlling for content, complexity, and other article characteristics. Thus, although popularity remains hard to predict (Salganik et al. 2006), gaining consumer attention is crucial for the propagation of information.

Using an agent-based model, Weng et al. (2012) illustrated that heterogeneity in the virality of different

Figure 1. Theoretical Framework



messages on Twitter can be explained by only two factors: (1) competition for our limited attention and (2) the structure of the social network. Other factors, such as the appeal of the message content, the persuasiveness of an individual, and external events, were not necessary to derive the observed empirical patterns on Twitter. Competition for attention also explains the fact that a small proportion of individuals is responsible for most of the information shared on social networks (Iyer and Katona 2016) and the power-law distribution of trending topics on Twitter (Asur et al. 2011). Similarly, competition for attention is an important explanatory factor for the popularity of stories on digg.com (Wu and Huberman 2007).

All these studies stress the importance of taking competition for attention into account when determining which individuals are critical for information propagation on social networks. Individuals with many friends are likely to receive more information, and hence, there is more competition for their attention (Feng et al. 2015). Thus, for seeding decisions, it is important to take into account how many friends a potential seed has and how many friends these friends have. The latter has a direct impact on the attention of the seeds' friends and thus on the competition for it. These two aspects of information propagation are related to *negatively* connected networks, as proposed in exchange-network theory. Exchange-network theory allows us to combine the ideas of obtaining a high reach as a result of having many friends and being surrounded by responsive friends because they receive relatively little competing information.

2.2. Positively and Negatively Connected Social Networks

According to exchange-network theory (Cook 1982, Cook et al. 1983), networks are either positively or negatively connected depending on whether exchange in one relationship affects exchange in other relationships positively or negatively (Yamagishi et al. 1988). In a positively connected network, exchange in one relationship is contingent on exchange in another relationship. Networks of brokerage are an example of positively connected networks. In such networks, exchange between buyers and brokers is contingent on exchange between brokers and sellers. However, as argued by Cook et al. (1983), networks with only positive connections are probably rare. In a negatively connected network, exchange in one relationship is contingent on *nonexchange* in another relationship. There is competition between the contacts of each network member. A dating network, for example, is strongly negatively connected (Bearman et al. 2004)

because exchange in one relationship inhibits exchange in another relationship.

Although we study information propagation rather than exchange, online social networks can also be positively or negatively connected. Whether they are positively or negatively connected has not been addressed in the literature yet and remains an empirical question. People share many types of information with their friends on online social networks, such as status updates, pictures, links to external web pages, and marketing messages. We expect that online social networks are negatively connected because people have limited mental resources and bandwidth to process information (Aral and Van Alstyne 2011). Hence, messages are competing for attention. Next, we describe how the framework of positively versus negatively connected networks can be applied in the context of information propagation in order to identify effective seeds who can obtain a high reach.

2.3. Who to Seed in Positively vs. Negatively Connected Networks?

To identify effective seeds as a function of network structure and connectivity, we develop a network game. Network games are powerful tools to model strategic behavior and to identify the most important network members (Jackson 2008, Lobel et al. 2016). To identify which network members can obtain the highest reach in both positively and negatively connected networks, we build on the network game of Ballester et al. (2006). Our network game contains N individuals connected in a social network that is represented by adjacency matrix A . This is an $N \times N$ matrix,⁵ with $a_{ij} = 1$ if the information that individual i shares is received by individual j —that is, j “follows” i —and $a_{ij} = 0$ otherwise or when $i = j$. Similar to members of social networks such as Facebook, LinkedIn, Instagram, Twitter, and Weibo, individuals in the network share content with their friends (undirected networks) or followers (directed networks). The shared content consists of both newly generated messages (e.g., a family picture) and passing on existing messages (e.g., a campaign message). The more active network members are—that is, the more messages they share—the more information will propagate. Each network member i derives utility u_i from sharing depending on his or her sharing rate, represented by x_i . The sharing rate can be interpreted as the number of messages shared within a given time interval. Following Ballester et al. (2006), we define the utility of sharing as follows:

$$u_i(x_1, \dots, x_N) = \alpha x_i - \frac{1}{2} x_i^2 + \beta \sum_{j=1}^N a_{ji} x_j x_i. \quad (1)$$

In Equation (1), we assume that $\alpha > 0$ to capture decreasing marginal returns of sharing. To capture competition for attention, we follow Aral and Van Alstyne (2011) and allow that individuals have a capacity constraint on listening to and sharing information. Given the capacity constraint, the more messages an individual receives, the less likely he or she is able to attend to any specific message, process it, and, subsequently share it. This effect can be captured by negative cross-effects ($\beta < 0$) between the received messages $\sum_{j=1}^n a_{ji}x_j$ and the shared messages x_i . By contrast, if cross-effects are positive ($\beta > 0$), someone’s capacity to process information increases as the number of messages received increases. Such complementarity effects may occur when sharing messages is more enjoyable if someone’s friends or the people he or she follows also actively share messages (Lin and Lu 2011).

Because of the linear-quadratic specification of the utility function, network members have a unique sharing rate that maximizes their utility. In matrix notation, the first-order condition of the game is given by

$$\alpha \mathbf{1}_N - I_N x + \beta A^T x = \mathbf{0}, \tag{2}$$

where $\mathbf{1}_N$ is an N -dimensional vector of ones, I_N is the $N \times N$ identity matrix, A^T is the transpose of the adjacency matrix, and x is an N -dimensional vector with sharing rates (x_1, x_2, \dots, x_N) . Solving for x , the equilibrium sharing rates x^* are given by

$$x^* = \alpha (I_N - \beta A^T)^{-1} \mathbf{1}_N. \tag{3}$$

In equilibrium, it holds that

$$x_i^* = \alpha + \beta \sum_{j=1}^N a_{ji} x_j^*. \tag{4}$$

Equation (4) shows that the equilibrium sharing rate of i is a linear function of the sum of the sharing rates of everyone from whom i receives messages—that is, i ’s friends in an undirected network or the people who i follows in a directed network. In negatively (positively) connected networks in which $\beta < 0$ ($\beta > 0$), i ’s optimal sharing rate decreases (increases) with the sharing rates of i ’s friends (undirected network) or the people i follows (directed network).

In a viral marketing campaign, the goal of the firm is to seed those network members who are instrumental in maximizing the campaign’s reach (Bampo et al. 2008, Kane et al. 2012). Hence, marketers aim at choosing seeds who trigger interest in the campaign among their friends or followers such that they will, in turn, share the campaign with their friends or followers. To determine the optimal seeding strategy, we extend the network game as follows. First, we introduce the N -dimensional vector s , with $s_i = 1$ if

network member i is seeded and $s_i = 0$ otherwise. This vector captures the unilateral seeding decision of the firm. Second, seeded individuals receive one more message, corresponding to the campaign message. Third, following previous literature (Tang et al. 2014, Aral and Dhillon 2018), we “assume that seeding is ‘successful’ at some basic level” (Aral et al. 2013, p. 148), which implies that seeds share the campaign message. In addition, we assume for now that only one network member k is seeded such that $s_k = 1$ and $s_i = 0$ for all $i \neq k$. Incorporating these assumptions in Equation (1) leads to the following new utility function for seeded network member k :

$$u_k(x_1, \dots, x_N, s) = \alpha(x_k + s_k) - \frac{1}{2}(x_k + s_k)^2 + \beta \sum_{i=1}^N a_{ik} x_i (x_k + s_k) + \beta s_k (x_k + s_k). \tag{5}$$

As seed k shares the campaign and other messages, his or her sharing rate consists of sharing noncampaign messages and the campaign and thus can be written as $x_k + s_k$. The last term of Equation (5), $\beta s_k (x_k + s_k)$, captures how the sharing rate of the seed changes in response to receiving the campaign from the firm. In a positively connected network ($\beta > 0$), the seed derives more utility from own sharing as he or she receives additional information from the firm. In a negatively connected network ($\beta < 0$), however, the seed derives less utility from own sharing because he or she now receives information from the firm in addition to the messages received from network connections. Because the seed has limited capacity, his or her optimal sharing rate in equilibrium will drop. The adjusted sharing rate of the seed will, in turn, affect the sharing rate of other people in the network. In particular, for all network members j who are not seeded but who might be connected to seed k ($j \neq k$), we extend Equation (1) as follows:

$$u_j(x_1, \dots, x_N, s) = \alpha x_j - \frac{1}{2} x_j^2 + \beta \sum_{i=1}^N a_{ij} (x_i + s_i) x_j. \tag{6}$$

Equation (6) is equivalent to Equation (1), except that when j is connected to seed k , j will receive the campaign message, as captured by s_i , which equals 1 for $i = k$. Nonseeded network members choose their sharing rates in response to the sharing rates of their friends, treating noncampaign and campaign messages equally.

We can now combine utility functions for the seed (Equation (5)) and nonseeds (Equation (6)) to arrive at the utility function for any network member i .

$$u_i(x_1, \dots, x_N, s) = \alpha(x_i + s_i) - \frac{1}{2}(x_i + s_i)^2 + \beta \sum_{j=1}^N a_{ji} (x_j + s_j) (x_i + s_i) + \beta s_i (x_i + s_i). \tag{7}$$

Equation (7) holds for both seeds and nonseeds and for any seeding strategy s , also when seeding more than one network member. The firm’s seeding strategy disturbs the equilibrium derived in Equation (3). The first-order conditions of the new equilibrium are given by

$$\alpha \mathbf{1}_N - I_N(x + s) + \beta A^T(x + s) + \beta s = \mathbf{0}. \quad (8)$$

The solution to this equation leads to the equilibrium sharing rates of the network members under seeding strategy s (see Appendix A).

$$x^*(s) = \underbrace{x^*(\mathbf{0})}_{\text{Equilibrium without seeding}} - \underbrace{(1 - \beta)s}_{\text{Direct seeding effect}} + \underbrace{\beta^2(I_N - \beta A^T)^{-1}A^T s}_{\text{Indirect seeding effect}}. \quad (9)$$

The new equilibrium in Equation (9) consists of three components. First, all network members adjust their previous equilibrium sharing rate without seeding $x^*(\mathbf{0})$ defined in Equation (3). Second, seeded network members reduce their sharing rate for noncampaign messages by $(1 - \beta)$. Because of the capacity constraint on listening and sharing, the sharing of one noncampaign message is replaced by the campaign message. In a negatively (positively) connected network, this reduction is enhanced (attenuated) by β because the campaign message competes for attention (complements sharing) with noncampaign messages. Third, sharing rates are affected indirectly in response to the adjustment of the seeds’ sharing rates. This indirectly affects all network members, both seeded and nonseeded, who adjust their sharing efforts by $\beta^2(I_N - \beta A^T)^{-1}A^T s$. Under competition for attention, this is a consequence of the reduction of the sharing rates of seeds, leading to lower levels of competition for the attention of nonseeded network members.

The goal of the firm is to optimize the campaign’s reach by choosing seeding strategy s such that the total sharing in the network is maximized. Because the direct seeding effect does not depend on network structure A , a firm only needs to consider the indirect seeding effect ($\beta^2(I_N - \beta A^T)^{-1}A^T s$) when deciding whom to seed. Given a predetermined seed size $|s|$, the optimal seeding strategy s corresponds to maximizing the sum of indirect seeding effects across all network members.

$$\max_s \mathbf{1}_N^T \beta^2 (I_N - \beta A^T)^{-1} A^T s \quad \text{subject to} \quad \sum_{i=1}^N s_i = |s|. \quad (10)$$

In Equation (10), $\mathbf{1}_N^T$ corresponds to an N -row vector (i.e., the transpose of $\mathbf{1}_N$). Interestingly, the maximization objective in (10) equals (see Appendix A)

$$\mathbf{1}_N^T \beta^2 (I_N - \beta A^T)^{-1} A^T s = s^T \beta^2 B(A, \beta), \quad (11)$$

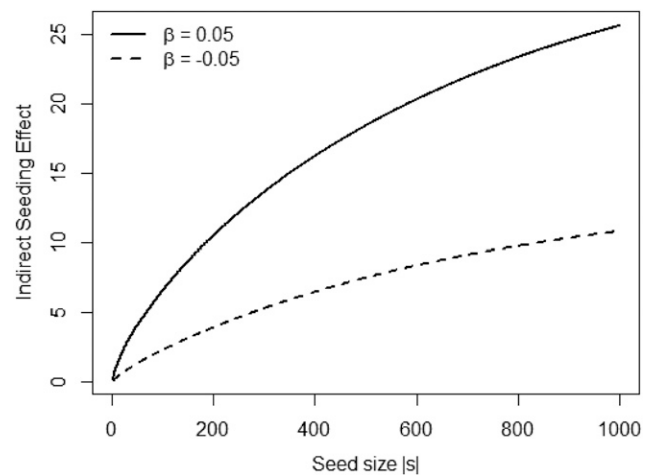
with $B(A, \beta)$ representing the vector of Bonacich centralities for each network member (Bonacich 1987).

$$B(A, \beta) = (I_N - \beta A)^{-1} A \mathbf{1}_N = A \mathbf{1}_N + \beta A^2 \mathbf{1}_N + \beta^2 A^3 \mathbf{1}_N + \beta^3 A^4 \mathbf{1}_N + \dots \quad (12)$$

Hence, the optimal seeding strategy is obtained when firms sequentially—either using a roll-out strategy or by selecting the seed size $|s|$ a priori—target the seeds with the highest Bonacich centrality.

To illustrate the optimal seeding strategy, we determined the optimal seeding strategy in a simulated undirected network of size $N = 1,000$. To ensure that the simulated network has real-world properties, such as a scale-free degree distribution, clustering, and degree assortativity, we followed the method of Sendiña-Nadal et al. (2016).⁶ We compared a positively and a negatively connected network ($\beta = 0.05$ and $\beta = -0.05$) and calculated the sum of the indirect seeding effects in Equation (11) for the optimal seeding strategy of targeting network members with the highest Bonacich centrality. Figure 2 shows the indirect seeding effect for different seed sizes, ranging from seeding only one network member with the highest Bonacich centrality to seeding all network members. In both positively and negatively connected networks, the indirect seeding effect increases in seed size. Importantly, for a given seed size, the indirect seeding effect is always larger in a positively than in a negatively connected network. Firms operating on a negatively connected network thus have to increase their efforts in terms of seed size to achieve the same network activation as firms operating on a positively connected network. This is in line with our proposed mechanism of competition for attention hindering information sharing in a negatively connected network.

Figure 2. Indirect Seeding Effect as a Function of Seed Size in Positively Connected ($\beta = 0.05$) and Negatively Connected ($\beta = -0.05$) Networks



2.4. Bonacich Centrality as a Measure for Seed Selection

As derived in the preceding section, optimal seeding consists of selecting network members with the highest Bonacich centrality, in both positively and negatively connected networks. This centrality measure depends on the connectivity parameter β of the social network (Bonacich 1987). As can be seen in Equation (12), for both positively and negatively connected networks, $B(A, \beta)$ increases in the number of friends a network member has (i.e., $A1_N$). The difference between Bonacich centrality in both types of networks is in how connected someone’s friends are. In a positively connected network, someone’s Bonacich centrality is high if his or her friends also have many friends (i.e., A^21_N is large⁷), and this quickly leads to a higher reach of network members. In a negatively connected network, someone’s Bonacich centrality is high if his or her many friends have only a few friends (i.e., A^21_N is small), meaning that there is little competition for the attention of these friends, which in this case will facilitate reach. To illustrate this, consider the undirected network presented in Figure 3. In this network, we highlighted four individuals, A–D. These four individuals have an equal number of friends, but their friends (i.e., $A1, A2, B1, B2$, etc.) differ with respect to the number of their friends. Suppose that a marketer is initiating a social media campaign and considers individual A, B, C, or D as a potential seed. If the network is positively connected, the marketer should consider people who are connected to as many other people as possible in just a few steps. In such a case, seeding individual A would be the best option—the message could then quickly spread to more network

members than if it was seeded to B, C, or D. However, in a negatively connected network, the friends of A are more prone to information overload because they potentially receive more information than the friends of B, C, and D. Individual A therefore might not be the most effective one to seed. Individual A’s friends may be receiving many competing messages and thus may be less likely to pay attention to and share the message received from A. In this situation, individual B, C, or D may be a better candidates for seeding.

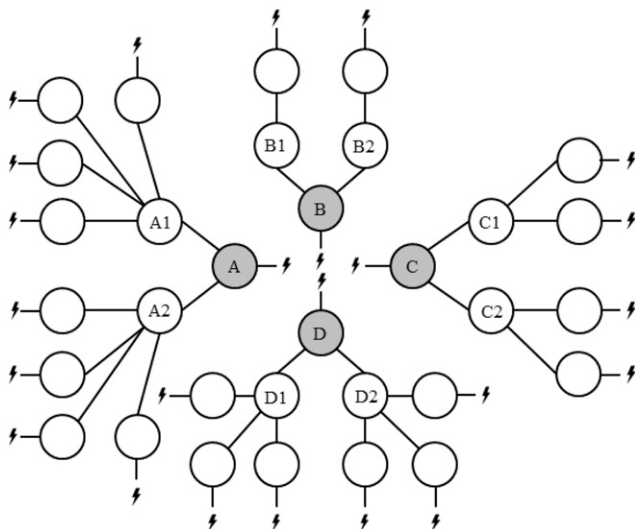
Although special cases of the Bonacich centrality measure have appeared in the recent seeding literature, previous research, to the best of our knowledge, determined the value of β a priori, and no research has considered negative values. If $\beta = 0$, Equation (12) corresponds to (out)degree centrality, which is the most frequently used centrality measure in research in marketing on social networks (Tucker 2008, Goldenberg et al. 2009, Lee et al. 2010, Trusov et al. 2010, Ansari et al. 2011, Aral and Walker 2011, Braun and Bonfrer 2011, Hinz et al. 2011, Iyengar et al. 2011, Katona et al. 2011, Zubcsek and Sarvary 2011, Yoganasimhan 2012). Furthermore, if β is set equal to the inverse of the largest eigenvalue of adjacency matrix A , it corresponds to eigenvector centrality, a centrality measure in a positively connected network. Tucker (2008) and Chen et al. (2017) have examined eigenvector centrality and concluded that it performs worse than degree centrality in explaining technology adoption and information propagation, respectively.

Although a seeding strategy based on Bonacich centrality is theoretically optimal, a practical limitation of this measure is that it requires observing the entire social network (see Equation (12)). This is infeasible in many situations because businesses running social media campaigns usually do not observe the entire network. We therefore propose a truncated version of the Bonacich centrality measure for practical seeding purposes, which is defined as follows:

$$TB(A, \beta) = A1_N + \beta A^21_N. \tag{13}$$

This approximation captures the idea that both own and friends’ degrees matter. Moreover, because β is typically a very small number (Bonacich 1987), the higher-order terms in Equation (12) get a very small weight and thus are less important. The approximation in Equation (13) has two important advantages over the original definition. First, it has more practical use because it does not require observing the complete network. Given an initial set of network members—for example, based on Facebook likes or Twitter followers—data on first and second degree can easily be obtained by navigating the network (Kane et al. 2012, van Dam and van de Velden 2015).

Figure 3. Illustrative Artificial Social Network



Note. The ⚡ sign indicates that there may exist more network connections than presented here.

Second, the implementation in Equation (13) allows for a more straightforward estimation procedure of β because Equation (12) involves an infinite sum or a large matrix inversion, which is generally more difficult to estimate.⁸

3. Study 1: Optimal Seeding, an Empirical Validation Based on a Social Media Game Campaign

We validate our proposed optimal seeding strategy by analyzing a real-life social media campaign on a large online social network platform. This campaign involved an online bowling game and was initiated by an entertainment company to promote the launch of an animated movie. The game was developed specifically for this purpose and was similar to Angry Birds.⁹ In the game, the gamer shoots a ball and aims at hitting bowling skittles. To seed the social media campaign, the entertainment company did not select specific network members strategically but posted a banner visible to all members between May 25 and 31, 2009. Network members who clicked on the banner connected to the campaign website, where they could play the game. After playing the game, participants were offered the opportunity to select friends with whom to share the campaign by challenging them also to play the game. After sharing, the receivers could click on the link in the invitation received from their friends, which also connected them to the campaign website, where they could play the game. These participants could then, in turn, select with whom to share the campaign among their friends and so forth. The company recorded time-stamped information on

who accessed the campaign website and who shared with whom. By visiting the campaign website, participants permitted the entertainment company to access information on their own and on their friends' social network profiles. In our analysis, we identify initial participants who clicked on the banner as seeds.

3.1. Network, Campaign, and Seed Descriptive Statistics

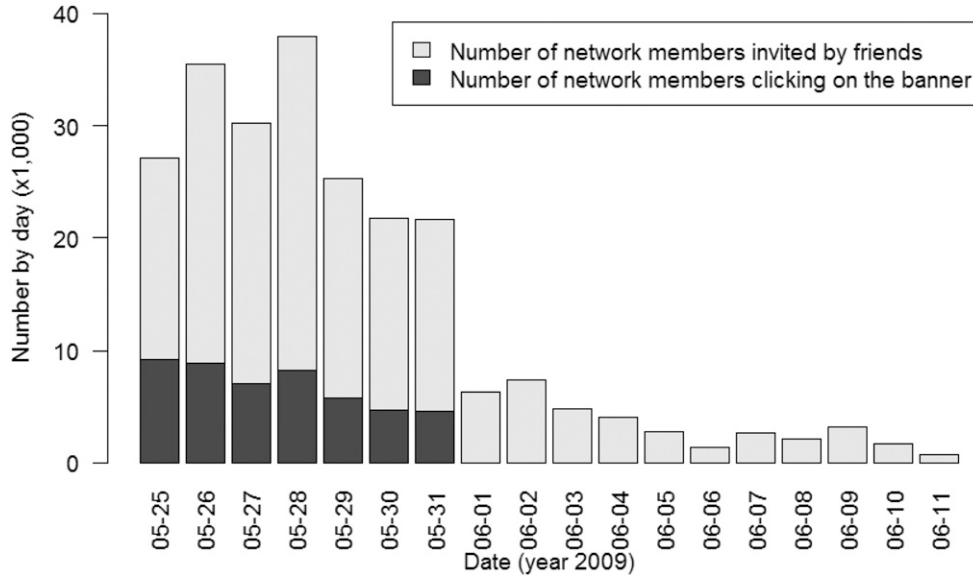
Summary statistics of the undirected network, the campaign, and the seeds are presented in Table 1. We have profile information for more than 4 million network members, which constitute more than half the total estimated 7 million social network members at the time of the campaign. The observed network members had a strongly right-skewed degree distribution (mean degree = 158.3, standard deviation (SD) = 398.7), consisted of relatively young members (mean age = 26.0, SD = 16.0), with slightly more women (56%) than men (44%). Several of the observed network members (20%) had not disclosed personal information on either age or gender. We recorded this in a missingness dummy, which takes the value one if the information on age or gender is missing and zero otherwise.

Figure 4 summarizes the spread of the social media campaign over time. The banner was available during the first seven days of the campaign. We observed a sharp drop in the number of campaign participants once the banner was removed. The sharing process continued for 11 more days, during which the number of shares gradually declined. Throughout the forwarding chains, the campaign reached a total of

Table 1. Study 1 Descriptive Statistics

	Total	Mean	Standard deviation
Network statistics			
Number of network members observed	4,002,033		
Degree		158.3	398.7
Age (based on complete observations)		26.0	16.0
Gender (male = 1, female = 0, based on complete observations)		0.44	0.50
Missingness dummy		0.20	0.40
Degree assortativity	0.04		
Campaign statistics			
Time period of seeding (days)	7		
Number of days active spreading	18		
Total number of people reached by the campaign	188,303		
Seed statistics			
Number of seeds	71,501		
Number of seeds sharing the campaign	5,028		
Number of friends shared with sharing		19.1	35.4
Reach sharing		32.4	72.3
Degree		145.2	193.7
Second degree		252,669	509,220
Age (based on complete observations)		21.9	25.3
Gender (male = 1, female = 0, based on complete observations)		0.53	0.49
Missingness dummy		0.15	0.36

Figure 4. Study 1: Spread of the Viral Marketing Campaign over Time

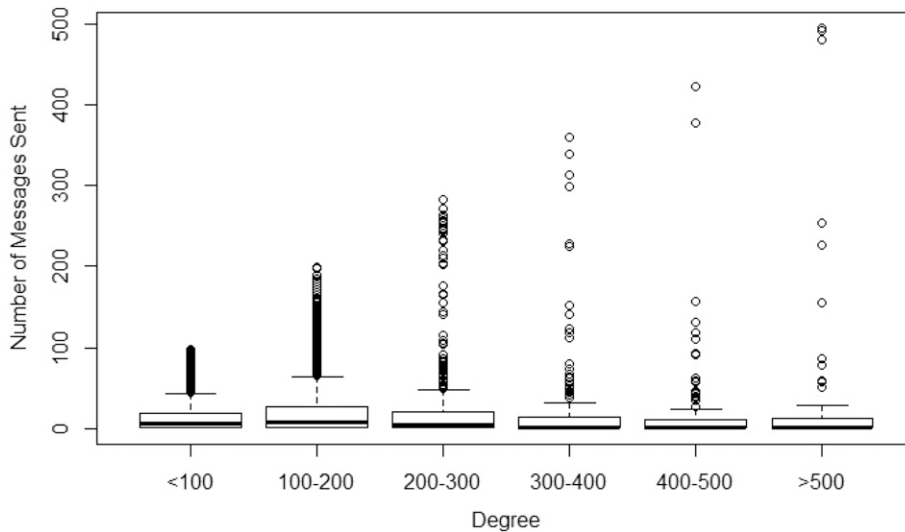


188,303 participants. Among these, 71,501 network members clicked on the banner—we call these the seeds—and 116,802 were invited by friends.

Of the 71,501 seeds, a total of 5,028 shared the campaign with their friends, corresponding to 7.03% sharing. The average number of friends shared with (M_i), conditional on sharing, was 19.1 (SD = 35.4). However, as illustrated in Figure 5, the distribution of M_i is heavily skewed. To compute the reach (R_i) that a seed i obtained, we counted the number of network members in the cascade that he or she initiated. The average reach conditional on sharing is 32.4, but the distribution is again heavily skewed (SD = 72.3). The seeds have an average degree of 145.2 (SD = 193.7)

and an average second-order degree of 252,669 (SD = 509,220). The high second-order degree relative to the first-order degree is in line with the friendship paradox (Feld 1991), which states that most people have fewer friends than their friends have. Regarding demographics, the seeds are, on average, 21.9 years old (SD = 25.3) and are about equally divided among men (53%) and women (47%). About 15% of the seeds opted not to disclose age or gender information. For our model estimation, we have mean imputed age for those profiles, where the mean is computed based on the complete observations, and we have used an effect coding for gender: -1 for female, 1 for male, and 0 if the information on gender is missing.

Figure 5. Study 1: Number of Messages Sent by Seeds as a Function of Their Degree



3.2. Model Formulation

In the first stage of a social media campaign, a company communicates a marketing message to social network members. When confronted with this message, a seeded network member decides whether to share the message with friends. Once a decision to share is made, he or she needs to choose how many friends with whom to share. In the campaign we analyzed, the vast majority of seeded network members who received a message from the company did not share, as will most likely be the case for most social media campaigns (Goel et al. 2012). Therefore, in modeling the sharing decision and reach, we used a hurdle model, which accounts for the excess zeroes in the data. This modeling approach is based on Hinz et al. (2011). However, Hinz et al. (2011) used independent models for the sharing, the number of friends shared with, and the reach obtained and thus assumed that these are independent decisions. We extend their approach and model these decisions simultaneously by allowing for a correlated error structure. Our approach controls for sample selection because someone may decide to share a message only if he or she believes that the information is useful for his or her friends (Berger and Milkman 2012) and, hence, obtains a higher reach. If not properly accounted for, sample selection may lead to biased parameter estimates.

For each seed i , let D_i denote whether i shares or not—that is, $D_i = 1$ if i shares and $D_i = 0$ otherwise. We model the decision variable D_i using a probit model with latent variable v_i such that

$$v_i = a_1 + b_1 \text{Degree}_i + \gamma_1 \mathbf{Z}_i + \varepsilon_{1i}, \quad (14)$$

and $D_i = 1$ if and only if $v_i > 0$. The vector \mathbf{Z}_i contains the control variables age and gender and the missingness dummy.

Following exchange-network theory and the network game, network members with high degree tend to receive more messages, and there is thus more competition for their attention. Hence, we expect a negative value for b_1 . This is also in line with the assumptions of Bakshy et al. (2011), who suggest that seeding high-degree individuals is costlier because it is more difficult to convince these individuals to share information. In contrast, Hinz et al. (2011) find that high-degree individuals have a higher probability to share a marketing message. However, in their study, individuals received a monetary incentive to share information—that is, free airtime for a mobile service. Because degree centrality in their study was measured by the number of phone calls to other individuals, customers with high-degree centrality derive higher benefits from sharing (free airtime).

After deciding to share, seed i chooses how many friends with whom to share, denoted by M_i , which

follows a zero-truncated negative binomial distribution given by

$$M_i \sim \text{TruncNB}(\lambda_i, q), \quad (15)$$

where q is the overdispersion parameter, and λ_i is the expectation of M_i conditional on the covariates

$$\lambda_i = \exp(a_2 + b_2 \text{Degree}_i + \gamma_2 \mathbf{Z}_i + \varepsilon_{2i}). \quad (16)$$

The zero-truncated negative binomial distribution accounts for overdispersion and for the fact that the number of shared messages is always positive, conditional on the sharing decision in Equation (14). We expect a positive value of b_2 because high-degree individuals generally have more friends with whom to share the message.

The effectiveness of the seeds is measured by their reach R_i , defined as the number of network members who receive the message in the cascade initiated by seed i . We use the number of friends shared with M_i and the truncated Bonacich centrality $TB_i(A, \beta)$ as predictors of reach. We model reach using a zero-truncated negative binomial distribution

$$R_i \sim \text{TruncNB}(\mu_i, r), \quad (17)$$

where r is the overdispersion parameter, and μ_i is the expectation of R_i conditional on the covariates

$$\mu_i = \exp(a_3 + b_3 TB_i(A, \beta) + d_3 M_i + \gamma_3 \mathbf{Z}_i + \varepsilon_{3i}). \quad (18)$$

Following our theoretical model, we expect a positive effect of Bonacich centrality on reach after controlling for the number of shares M_i , which corresponds to $b_3 > 0$. The more friends someone shares with, the higher is the expected reach, so we expect that $d_3 > 0$. In the estimation, we recast Equation (18) as

$$\mu_i = \exp(a_3 + b_3 \text{Degree}_i + b_4 \text{SecondDegree}_i + d_3 M_i + \gamma_3 \mathbf{Z}_i + \varepsilon_{3i}), \quad (19)$$

with $\beta = b_4/b_3$.

Because the seeds decide themselves whether to share the campaign message, we need to account for self-selection in Equations (15) and (17). Therefore, we use a correlated error structure between the error terms of Equations (14), (16), and (19).

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \right). \quad (20)$$

In Equation (20), we set $\sigma_{11} = 1$ for identification purposes of the probit part of the model.

3.3. Model Estimation

Because the error terms of the three model equations (Equations (14), (16), and (19)) are correlated, we use a

joint maximum-likelihood procedure that simultaneously estimates all model parameters. To derive the model likelihood, we first partition Σ as follows for notational convenience:

$$\Sigma = \begin{pmatrix} 1 & | & \sigma_{12} & \sigma_{13} \\ - & | & - & - \\ \sigma_{12} & | & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & | & \sigma_{23} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} 1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \quad (21)$$

The conditional distribution of ε_{1i} , given ε_{2i} and ε_{3i} , is then given by

$$\varepsilon_{1i} | \varepsilon_{2i}, \varepsilon_{3i} \sim N\left(m_i = \Sigma_{12} \Sigma_{22}^{-1} \begin{pmatrix} \varepsilon_{2i} \\ \varepsilon_{3i} \end{pmatrix}, V = 1 - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}\right). \quad (22)$$

The conditional probability of sharing can be written as

$$\Pr(D_i = 1 | \varepsilon_{2i}, \varepsilon_{3i}) = \Phi\left(\frac{a_1 + b_1 \text{Degree}_i + \gamma_1 Z_i + m_i}{\sqrt{V}}\right) = \Phi_i^*, \quad (23)$$

where $\Phi(\cdot)$ is the standard normal probability function. The conditional likelihood contribution of observation i is given by

$$\mathcal{L}_i(\boldsymbol{\theta} | \varepsilon_{2i}, \varepsilon_{3i}) = \left(1 - \Phi_i^*\right)^{(1-D_i)} \times \left(\Phi_i^* f_{TNB}(M_i; \lambda_i, q) f_{TNB}(R_i; \mu_i, r)\right)^{D_i}, \quad (24)$$

where $\boldsymbol{\theta}$ is the parameter vector collecting $\{a_1, a_2, a_3, b_1, b_2, b_3, b_4, d_3, \gamma_1, \gamma_2, \gamma_3, q, r, \Sigma\}$, $f_{TNB}(M_i; \lambda_i, q)$ is the density function of truncated negative binomial distribution

$$f_{TNB}(M_i; \lambda_i, q) = \frac{\lambda_i^{M_i}}{M_i!} \frac{\Gamma(q + M_i)}{\Gamma(q)(q + \lambda_i)^{M_i}} \frac{1}{\left(1 + \frac{\lambda_i}{q}\right)^q - 1}, \quad (25)$$

and similarly for $f_{TNB}(R_i; \mu_i, r)$. The unconditional likelihood contribution is given by

$$\mathcal{L}_i(\boldsymbol{\theta}) = \iint \mathcal{L}_i(\boldsymbol{\theta} | \varepsilon_{2i}, \varepsilon_{3i}) g(\varepsilon_{2i}, \varepsilon_{3i}) d\varepsilon_{2i} d\varepsilon_{3i}, \quad (26)$$

where $g(\varepsilon_{2i}, \varepsilon_{3i})$ is the joint density function of ε_{2i} and ε_{3i} .

Because the likelihood function in Equation (26) does not have a closed-form solution for $\boldsymbol{\theta}$, we apply numerical integration. To reduce computational costs in evaluating the double integrant, we use sparse grids, as proposed by Heiss and Winschel (2008). Likelihood approximation based on sparse grids is computationally less demanding than a simulated maximum-likelihood approach. Confidence intervals

of all elements of $\boldsymbol{\theta}$ are obtained by using 1,000 bootstrap samples (Efron 1985).

3.4. Results

In addition to the full model, in which we estimated all parameters, including the network connectivity parameter β and the full covariance matrix Σ , we also estimated four benchmark models. The first benchmark model neglects network structure and only includes the demographic variables age and gender and the control variable for missingness. In the second benchmark model, we include degree centrality, which is the most common network measure in the previous literature. This is equivalent to setting the network connectivity parameter β equal to zero and does not take competition for attention into account. The third benchmark model includes a proxy for eigenvector centrality,¹⁰ the centrality measure in a positively connected network. The fourth benchmark model includes all the covariates of the full model but does not account for sample selection bias. We computed the approximated log-likelihood and Bayesian information criterion (BIC) for all five models to determine which model best describes the underlying process of the social media campaign (Table 2).

Table 2 presents the estimation results and model-fit statistics (log likelihood and BIC) of the four benchmark models and the full model. Compared with the model without network information (benchmark 1), including degree centrality (benchmark 2) significantly improves model fit, which is in line with previous research (Hinz et al. 2011). However, assuming a positively connected network reduces model fit because this measure is not related to reach (benchmark 3). Moreover, controlling for selection bias is important because restricting covariances between the three equations to zero significantly reduces model fit (benchmark 4). The estimation results show that our full model outperforms all benchmarks in terms of both the approximated log-likelihood and BIC.

Because signs of parameters do not differ across the five models, we interpret only the results of the full model (last two columns in Table 2) in the remainder of this section. A key finding of this research is that the estimated network connectivity parameter β is negative ($\hat{\beta} = \hat{b}_4 / \hat{b}_3 = -0.927 \times 10^{-5}$) and significant at the 95% level across the bootstrap samples. Thus, as we argued earlier, network members who have many friends ($\hat{b}_3 = 1.200 \times 10^{-3}$) but whose friends have only few friends ($\hat{b}_4 = -1.112 \times 10^{-8}$) are able to obtain the highest reach.¹¹ In addition, following the expectations in a negatively connected network, we find in Equation (14) that degree has a significant negative effect on the probability of sharing ($\hat{b}_1 = -0.437 \times 10^{-3}$). Although high-degree network members have a lower probability of sharing, as we expected, these

Table 2. Study 1 Estimation Results

	Benchmark 1: Control only		Benchmark 2: Control + degree		Benchmark 3: Control + eigen		Benchmark 4: No error variance		Full model	
	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.
Sharing (Equation (14))										
Intercept	-1.550	***	-1.381	***	-1.303	***	-1.404	***	-1.735	***
Degree $\times 10^{-3}$			-0.689	***	-0.571	***	-0.604	***	-0.437	***
Age	0.012	**	0.010	**	0.009	**	0.011	***	0.016	**
Male	-0.143	***	-0.133	***	-0.115	***	-0.132	***	-0.114	***
Missingness dummy	-0.090	***	-0.094	***	-0.105	***	-0.090	***	-0.140	***
Number of friends shared with (M_i) (Equation (19))										
Intercept	1.174	***	1.224	***	2.973	***	2.630	***	2.508	***
Degree $\times 10^{-3}$			0.272	***	0.182	***	2.077	***	0.281	***
Age	0.001		0.001		0.016	**	0.001	**	0.002	*
Male	-0.013	**	-0.010	**	-0.007	**	-0.082	***	-0.011	*
Missingness dummy	-0.130	***	-0.158	***	-0.295	***	-0.123	*	-0.212	***
Reach (R_i) (Equation (16))										
Intercept	1.590	***	1.632	***	2.504	***	1.963	***	1.756	***
Degree $\times 10^{-3}$			0.178	**			1.602	***	1.200	**
Second degree $\times 10^{-8}$							-1.758	***	-1.112	**
Eigenvector centrality					-0.001					
$M_i \times 10^{-2}$	4.501	***	4.691	***	3.529	***	3.788	***	0.049	***
Age	0.012	**	0.013	**	0.023	***	0.013	***	0.004	**
Male	-0.073	**	-0.130	**	-0.195	***	-0.090	***	-0.055	*
Missingness dummy	-0.107	***	-0.072	***	-0.054	**	-0.136	**	-0.063	***
Network connectivity ($\beta \times 10^{-5}$)			0	Fixed	0.927	Fixed	-1.097	***	-0.927	**
Covariance matrix (Equation (20))										
Sigma 12	0.025		0.009		-0.733				-0.094	
Sigma 13	-0.276	*	-0.190	*	-0.260	*			-0.256	*
Sigma 23	0.191	**	0.082	**	0.042	**			0.003	*
Sigma 22	0.033	***	0.008	***	0.005	***			0.010	***
Sigma 33	1.123	***	0.850	***	0.975	***			1.720	***
LL	-55,084		-54,962		-55,022		-57,635		-51,842	
BIC	110,370		110,159		110,146		115,460		103,930	

Note. Sig., significance level.

*The 90% confidence interval does not contain zero; **the 95% confidence interval does not contain zero; ***the 99% confidence interval does not contain zero.

individuals share significantly more messages with their friends once they do decide to share ($\hat{b}_2 = 0.281 \times 10^{-3}$). These results support the mechanism that social media messages compete for the attention of social network members. First, network members with many connections are less likely to respond to messages received from an advertiser. Second, these network members are also less likely to respond to messages received from their friends, as indicated by the negative network connection parameter β .

In addition to the strong effects of degree and second-order degree, we find that older people are more likely to share, share with more friends, and obtain a higher reach. We also find that for this specific campaign, men are less likely to share, and if they share, they share with fewer friends and obtain a lower reach. This is different from earlier findings by Hinz et al. (2011), who report that in the context of mobile phone subscriptions, men are more likely to share a campaign message. However, in general, we

can expect gender differences in sharing behavior to be campaign specific (Phillip and Suri 2004). Finally, individuals with missing profile data are less likely to share, tend to share with fewer friends, and obtain a lower reach when they share. This confirms earlier findings on privacy concerns by Goldfarb and Tucker (2011) or indicates that these network members are simply less active in general.

3.5. Out-of-Sample Counterfactual Comparison of Seeding Strategies

To compare the potential reach of different seeding strategies, we conducted an out-of-sample comparison. All the seeding strategies that we compare include personal characteristics of the network members, which are typically observed by firms. In particular, we compare the expected reach of the campaign when using seeding strategies based on personal characteristics (i.e., age, gender, and missing profile information; i.e., benchmark 1) and seeding strategies that take into

account network information (i.e., first and second degree and eigenvector centrality; i.e., benchmarks 2–4 and the full model).

We first randomly split the seeds in two samples, a training sample that we used for estimating the model parameters and a holdout sample that we used for comparing the effectiveness of seeding strategies. Based on the parameters estimated by using the training sample, we ranked all seeds in the holdout sample according to their expected reach, as generated by each model. We repeated this procedure 100 times with different random splits. We then computed the actual cumulative reach of the seeds as a function of their rank in the holdout samples. The results, averaged across the 100 random holdout samples, are presented in Figure 6. The 45-degree line in this figure represents the results for a random seeding strategy, where each initial participant has the same expected reach. As expected, the analytically derived optimal seeding strategy within the network game based on the estimated Bonacich centrality led to the highest expected reach compared with all benchmark models. The vertical dashed line in Figure 6 compares the reach that is obtained by seeding the 10% initial participants with the highest expected reach following different seeding strategies. The reach obtained by the optimal seeding strategy is 4.2 times the reach obtained by random seeding, 2.2 times the reach of benchmark 1 that only uses control variables, 1.7 times the reach of benchmark 2 that uses control variables and degree centrality, 1.7 times the reach of benchmark 3 that uses control and eigenvector centrality, and 1.8 times the reach based on benchmark 4 that does not control for the correlated error structure.

The total area under the seeding curve (AUC)¹² of the full model is 0.70, whereas it is only 0.65, 0.66, 0.64, and 0.63, respectively, for benchmarks 1–4 (Table 3). To further evaluate the predictive power of the models, we also computed two out-of-sample prediction accuracy measures, the mean absolute prediction error (MAPE) and the root-mean-squared prediction error (RMSPE), of the reach of each seed in the holdout sample. The MAPE of the full model (3.3) is lower than that of all benchmark models (respectively, 4.5, 3.6, 5.2, and 3.5 for benchmarks 1–4). Also, the RMSPE of the full model (17.7) is lower than that of all benchmark models (respectively, 19.0, 18.8, 18.5, and 18.6 for benchmarks 1–4).

In sum, our first empirical study on a large-scale real-life social media campaign finds that the social network is negatively connected and that the proposed optimal seeding strategy, which accounts for competition for attention, outperforms benchmark seeding strategies by up to 70% depending on the seed size. Although these empirical results confirm the analytical results of Section 2, a number of concerns exist. First, the data of Study 1 cover only a single campaign on a single social network platform, limiting the generalizability of our findings. Second, because of the lack of full network information, we used a truncated version of the Bonacich centrality measure instead of the full measure in Equation (12). Finally, although the theoretical setup suggests that competition for attention is the driver of our results, the data of Study 1 do not allow us to test this idea directly. To address these concerns, we executed a second study. First, the data in Study 2 cover 33 campaigns (versus one campaign in Study 1) with

Figure 6. Study 1: Out-of-Sample Seeding Comparison

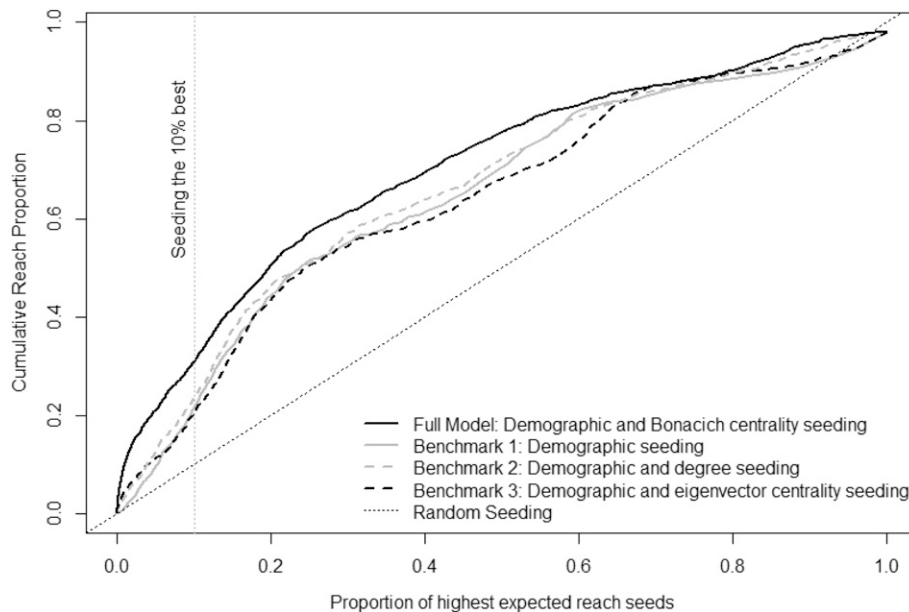


Table 3. Study 1 Out-of-Sample Model Comparisons

	Benchmark models				Full model
	(1) Control	(2) Degree	(3) Eigenvalue	(4) Uncorrelated	
Area under the curve	0.65	0.66	0.64	0.63	0.70
Mean absolute prediction error	4.5	3.6	5.2	3.5	3.3
Root-mean-squared prediction error	19.0	18.8	18.5	18.6	17.7

video content (versus an online game in Study 1), in which the seeds broadcast to *all* their friends (versus *selective* sharing, as in Study 1) on a different large social network platform. Second, because we observe the full network in Study 2, we can estimate the effect of the untruncated Bonacich centrality measure. Finally, because we observe how many messages people exchange in the network, we can provide empirical support for the mechanism of competition for attention.

4. Study 2: Empirical Validation Based on Multiple Social Media Campaigns

We obtained data from a social network of undergraduate students of a major university in the United States (Chen et al. 2017). The data were collected during the 2010 Super Bowl, a time when many brands launched new advertising campaigns. During this event, people can share these advertising campaigns with their friends in their social network. These friends may further share the campaign with their friends, and so on. In contrast to Study 1, in which people selectively shared the campaign with a selected number of friends, in Study 2, people broadcasted to all their friends. For 33 Super Bowl campaigns, we identified seeds and the cascades that they generated following previous research (Bakshy et al. 2011). A seed is identified as someone who initiated sharing a campaign without having received the campaign on their own social network before. The cascade initiated by the seed is then identified by following the chain of shares of the campaign throughout the friendship network. We observe all network connections between individuals, which allows us to compute the full Bonacich centrality, as defined in Equation (12). Importantly, similar to the face-to-face network of activities studied by Iyer and Katona (2016), we also observe how many messages were exchanged between friends on the network. These exchanges were measured over a two-month time period prior to the Super Bowl and allow us to obtain a more direct measure of how information competes for attention.

4.1. Network, Campaign, and Seed Descriptive Statistics

Summary statistics of the undirected network, the campaigns, and the seeds are presented in Table 4. We observe 42,858 network members with an average degree of 79.5 (SD = 75.4) and an average second degree of 12,009.6 (SD = 14,653.9). We also computed *k*-core centrality, a network measure that differentiates the periphery of the network (low *k*-core) from the inner core (high *k*-core) and is known as a good predictor of reach (Kitsak et al. 2010). The average *k*-core in the network is 40.5 (SD = 21.7). The network members are on average 19.0 (SD = 1.4) years old, are male in 55% of the cases, and are members of the social network for, on average, 4.70 years (SD = 1.22). The network has a positive degree assortativity of 0.23.

In total, we observe the spread of 33 Super Bowl campaigns. The average number of seeds per campaign is 109.6 (SD = 231.1), and the average total reach per campaign is 13,935.4 (SD = 38,439.6). The cascades are initiated by a total of 3,618 seeds, who obtained an average reach of 127.1 (SD = 330.1). Although the seeds' average degree (67.9; SD = 66.2), average second degree (9,908.2; SD = 12,761.3), and average *k*-core (37.2, SD = 21.5) are somewhat lower than the network averages, the seeds are very similar in terms of their demographic characteristics: average age of 18.8 years (SD = 1.4), 53% men, and average membership duration of 4.54 years (SD = 0.97).

4.2. Model Formulation and Estimation

Different from Study 1, *all* seeds share the campaign with *all* friends in their network. Therefore, we directly model the reach of each seed and do not model the decision to share and with how many friends the campaign is shared. Similar to Study 1 (Equations (17) and (18)), we model the reach R_i of seed i as a function of Bonacich centrality and control variables Z_i using a negative binomial regression model.

$$R_i \sim NB(\mu_i, r), \tag{27}$$

$$\mu_i = \exp(a_0 + bB(A, \beta_i) + \gamma Z_i). \tag{28}$$

Table 4. Study 2 Descriptive Statistics

	Total	Mean	Standard deviation
Network statistics			
Number of network members	42,858		
Degree		79.5	75.4
Second degree		12,009.6	14,653.9
<i>k</i> -Core		40.5	21.7
Age		19.0	1.4
Gender (male = 1, female = 0)		0.55	0.5
Membership duration (years)		4.70	1.22
Degree assortativity	0.23		
Campaign statistics			
Number of campaigns	33		
Number of seeds per campaign		109.6	231.1
Total number of people reached per campaign		13,935.4	38,439.6
Seed statistics			
Number of seeds	3,618		
Reach		127.1	330.1
Degree		67.9	66.2
Second degree		9,908.2	12,761.3
<i>k</i> -Core		37.2	21.5
Age		18.8	1.4
Gender (male = 1, female = 0)		0.53	0.51
Membership duration (years)		4.54	0.97

The parameters of interest are b and β , and we expect $b > 0$ and $\beta < 0$, indicating a positive effect of the Bonacich centrality on reach and a negatively connected network. As control variables Z_i , we include age, gender, and membership duration. We also control for *k*-core (Seidman 1983), which indicates whether seeds are located in connected regions of the network. Kitsak et al. (2010) found that a seed's degree centrality becomes unimportant after controlling for *k*-core. In contrast, Aral et al. (2013) did not find any additional value of seeding dense network regions compared with degree centrality. They did not, however, consider the possibility that networks may be negatively connected. To estimate the model, we used maximum likelihood with 1,000 bootstrap samples to obtain significance levels (Efron 1985).

4.3 Results

We estimated two full models. Similar to Study 1, the first full model used the truncated Bonacich centrality measure, as discussed in Equation (13). The second full model used Bonacich centrality obtained from the entire network (Equation (12)). In addition to the two full models, we also estimated three benchmark models. The first benchmark model only includes control variables: *k*-core, age, gender, and membership duration. In the second benchmark model, we added degree centrality to the model with control variables, which corresponds to assuming that the connectivity parameter β is equal to zero. In the third benchmark model, we fixed the connectivity parameter to the inverse of the largest eigenvalue of the

adjacency matrix (5.268×10^{-3}), which corresponds to eigenvector centrality and assumes a positively connected network.

Table 5 presents the estimation results of all five models. First, benchmark model 1 shows that *k*-core is a significant predictor of a seed's reach, confirming previous research (e.g., Kitsak et al. 2010, Harush and Barzel 2017, Lokhov and Saad 2017). The other control variables (age, gender, and membership duration of seeds) are not related to reach. Second, similar to Study 1, adding degree centrality significantly improves model fit (BIC = 35,716 versus 36,194, respectively, for a model with and without degree centrality). As expected, seeds with higher degree obtain a higher reach. Third, although eigenvector centrality is positively related to reach, the relationship is weaker than degree centrality (BIC = 36,127), corroborating Study 1. Fourth, and most important, both our full models, which include Bonacich centrality (truncated and nontruncated), fit the data significantly better than all three benchmark models. Interestingly, our full model with truncated Bonacich centrality describes the data slightly better according to BIC (35,472 versus 35,596, respectively, for the truncated and nontruncated Bonacich centrality). Moreover, the estimated network connectivity is negative and significant in both models, and the estimates are very similar ($\hat{\beta} = -2.982 \times 10^{-3}$ and $\hat{\beta} = -2.803 \times 10^{-3}$, respectively, for the model with truncated and nontruncated Bonacich centrality). The similarity between these two estimates provides support for use of the truncated Bonacich centrality measure, which is likely to be useful in practice. The absolute value of

Table 5. Study 2 Estimation Results

	Benchmark 1: Control only		Benchmark 2: Control + degree		Benchmark 3: Control + eigen		Truncated Bonacich		Bonacich	
	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.
Intercept	2.020	***	2.316	***	2.001	***	2.233	***	2.327	***
Degree			0.006	***			0.016	***		
Second degree $\times 10^{-4}$							-0.467	***		
Bonacich centrality									0.010	***
Eigenvector centrality					1.835	***				
<i>k</i> -Core	0.060	***	0.042	***	0.056	***	0.038	***	0.039	***
Age	-0.003		-0.006		-0.003		-0.011		-0.008	
Gender	-0.022		0.007		-0.018		0.022		0.014	
Membership duration	-0.007		-0.010		-0.051		0.013		-0.005	
Network connectivity (beta) $\times 10^{-3}$			0	Fixed	5.268	Fixed	-2.982	***	-2.803	***
LL	-18,072		-17,829		-18,035		-17,703		-17,769	
BIC	36,194		35,716		36,127		35,472		35,596	

Notes. Dependent variable is the reach of a seed. Sig., significance level.

***The 99% confidence interval does not contain zero.

the connectivity parameter is bounded by the inverse of the largest eigenvalue of the adjacency matrix (Bonacich 1987). The absolute value of our estimated coefficient is 0.53 times the bound of 5.268×10^{-3} , indicating an effect of real importance. Furthermore, both models illustrate that seeds with high Bonacich centrality obtain a higher reach. In line with Study 1, the full model with truncated Bonacich centrality finds a positive effect of degree ($b = 0.016$) and a negative effect of second-order degree ($b = -0.467 \times 10^{-4}$).¹³ Similarly, the full model with nontruncated Bonacich centrality finds a significantly positive effect of this measure on reach ($b = 0.010$).

4.4. Generalizability: The Underlying Mechanism and Heterogeneity

In line with Study 1, our estimation results show that the network in Study 2 is also negatively connected and that Bonacich centrality is a powerful predictor of a seed’s reach. In this section, we will further explore the generalizability of our findings. First, we demonstrate that competition for attention is indeed the underlying mechanism of our findings. Second, we explore the heterogeneity of the connectivity parameter across campaigns and across network members.

4.4.1. The Underlying Mechanism: Competition for Attention. Our results imply that seeds who have many friends but whose friends have only few friends are able to obtain the highest reach. Although we explain this effect through competition for attention, so far we have not directly shown that this is indeed the underlying mechanism. To support the explanation of competition for attention, we use the actual number of messages exchanged in the network and perform two separate analyses. In our first analysis, instead of predicting reach using the seeds’ degree and

second degree (the truncated Bonacich model in Table 5), we use the seed’s degree and the number of messages that the seed’s friends receive. The latter serves as a measure of the competition for the attention of the seed’s friends. Thus, if competition for attention indeed drives the effect, we expect a negative effect for the number of received messages: The more messages the seeds’ friends receive, the more competition there is for their attention, and the lower are the expected reach. Table 6 presents the results of this analysis. Consistent with the idea of competition for attention, we find a negative effect of the number of messages received by the seed’s friends on reach (estimated coefficient -0.489×10^{-3}).

In our second analysis, we test competition for attention more directly by studying the receivers in the cascades rather than the seeds. For all 455,868 instances in which a network member receives a campaign message from a friend (either from a seed or from someone further down the cascade), we observe whether this network member decides to share that message further. To explain the sharing decisions,

Table 6. Study 2 Estimation Results with Number of Messages as Measure of Competition for Attention

	Estimate	Sig.
Intercept	2.626	***
Degree	0.013	***
Number of messages received by friends $\times 10^{-3}$	-0.489	***
<i>k</i> -Core	0.039	***
Age	-0.031	
Gender	-0.001	
Membership duration	0.009	
LL	-17,713	
BIC	35,492	

Notes. Dependent variable is the reach of a seed. Sig., significance level.

***The 99% confidence interval does not contain zero.

we collected the following explanatory variables: (1) degree of the receiver, (2) the number of messages or posts received by the receiver in the two months prior to the Super Bowl, (3) the number of messages or posts sent by the receiver in the same period, and (4) other control variables: age, gender, and membership duration. If competition for attention explains the negative connectivity of networks, the number of messages received by a network member should mediate the effect of degree on the sharing decision. First, in line with these expectations, degree and the number of messages received are positively correlated ($r = 0.69$, $p < 0.001$). Second, we estimated five probit models of the sharing decision of receivers. To control for unobserved heterogeneity in sharing probabilities, we included a random receiver effect. Table 7 reports the results.

First, model 1 illustrates that degree is negatively related to the sharing decision, which is in line with a negatively connected network. Second, model 2 shows that the effect of degree centrality on the sharing decision becomes insignificant after controlling for the number of messages that the network member received. Hence, the number of messages received fully mediates the effect of degree, which supports our proposed mechanism of competition for attention. Interestingly, if we control for the number of messages sent by the network member, as suggested by Bakshy et al. (2011) (model 3), the effect of degree centrality remains significant and negative. Finally, the effect of the number of messages received on the sharing decision remains robust if we control for the number of messages sent (model 4), as well as for other control variables (model 5).

The underlying mechanism contributes to the generalizability of our findings and reduces the concerns about selection bias (Pearl 1995).¹⁴ However, so far we assumed a homogeneous connectivity parameter (β) across campaigns and individuals. Next, we will explore the heterogeneity of this parameter across campaigns, as well as across network members.

4.4.2. Heterogeneity Across Campaigns. To further investigate the generalizability of our findings, we separately estimated our full model with *nontruncated* Bonacich centrality for each of the 17 campaigns in which we observed at least 20 cascades. Consistent with our main findings reported in Table 5, we find that the network is negatively connected for *all* of these 17 campaigns. Figure 7 illustrates these results and plots the point estimates of the connectivity parameter β and their 95% bootstrap confidence intervals for each campaign. For only 4 of 17 campaigns, this parameter is not significant at the 5% level. Moreover, point estimates vary between -0.003 and -0.001 and are very similar across campaigns. Finally, consistent with the results in Table 5, the estimated effect of Bonacich centrality on reach is positive and strongly significant for all campaigns. This provides further evidence for the generalizability of our findings to different campaigns and networks.

4.4.3. Heterogeneity Across Network Members. In all previous analyses, we assumed a homogeneous connectivity parameter (β) for the entire network. This is an advantage from a managerial perspective because it suffices to know whether β is positive or negative to find the seeds who are able to obtain the highest reach. However, it is possible that the connectivity is heterogeneous across the network members. Some people might be more negatively connected than others—that is, there is stronger competition for their attention—whereas some might even be positively connected. To investigate this, we analyzed heterogeneity as a function of observed network characteristics¹⁵ for both Studies 1 and 2 (see the online appendix). Although we only found weak evidence for heterogeneity as a function of age in Study 2, the connectivity parameter was always negative and did not vary much across age levels. These results were also corroborated by a reduction in model fit (BIC) for both studies, suggesting that a homogeneous connectivity parameter captured the underlying process

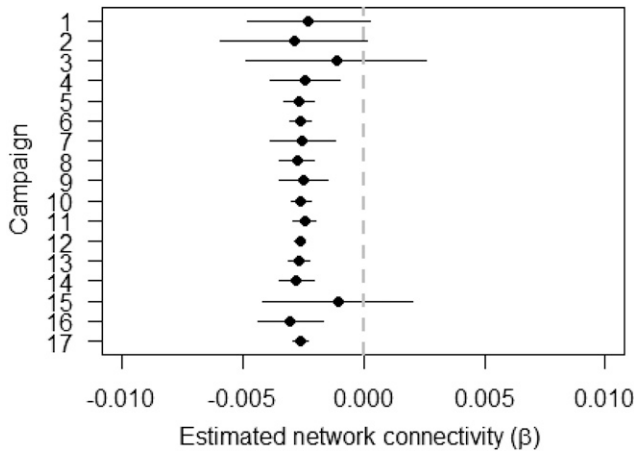
Table 7. Probit Model for the Forwarding Decision of Receivers

	Model 1		Model 2		Model 3		Model 4		Model 5	
	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.	Estimate	Sig.
Intercept	-1.048	***	-1.180	***	-1.041	***	-1.180	***	0.093	
Degree $\times 10^{-5}$	-0.057	***	-0.012		-0.059	***	-0.013		-0.007	
Number of messages received			-0.010	***			-0.011	***	-0.014	***
Number of messages sent					0.028	***	0.034	***	0.030	***
Age									-0.061	***
Gender									-0.064	***
Membership duration									-0.019	
Standard deviation random receiver effect	0.063		0.079		0.061		0.058		0.058	

Note. Sig., significance level.

***The 99% confidence interval does not contain zero.

Figure 7. Study 2 Generalizability: Estimated Effect (Dot) and 95% Confidence Interval (Line) of the Estimated Network Connectivity Across 17 Campaigns



better and that this parameter was significantly negative across network members with different characteristics.

In sum, the results of both studies reveal that networks are negatively connected and that this result generalizes across campaigns, network platforms, and network members. Moreover, our analyses using messages exchanged between network members support our proposed mechanism that information competes for attention. These findings have important theoretical and managerial implications, which we discuss in the next section.

5. Discussion

5.1. Main Findings and Implications

Online social networks are attractive platforms for the execution of marketing campaigns. However, people’s position in and communication behavior on these networks will determine whether and how marketing messages spread and to what extent they go viral. In this paper, we have developed new theoretical and empirical insights into the dissemination of information on online social networks. In particular, we provide evidence that online social networks may be negatively connected and that different messages compete for the attention of network members. The more intense the competition for the attention of a network member is, the less likely it is that this network member will respond to a message received from a friend. This insight has important implications for seeding. More specifically, we identified that seeding according to the Bonacich centrality with a negative connectivity parameter β is optimal when faced with competition for attention in the network. Network members with many friends who, in turn, have few other friends—so that there is little competition for these friends’ attention—are able to obtain the highest reach. The previous literature has

neglected the possibility of a negatively connected network or has implicitly assumed a positively connected network. In particular, Tucker (2008) and Chen et al. (2017) concluded that eigenvector centrality, a centrality measure in a positively connected network, performs worse than degree centrality in explaining technology adoption and information diffusion, respectively. Our results reconfirm this finding in the context of information diffusion. This is explained by our finding that online social networks are not positively but rather negatively connected.

Watts and Peretti (2007) suggested using “big seed marketing” to make social media campaigns successful in an environment where messages do not spread easily. In their approach, large numbers of people are randomly selected for targeting purposes. Our theoretical and empirical results, however, show that firms can substantially improve the reach of their campaigns by selectively targeting customers instead of randomly targeting them. In our research, this resulted in an expected reach that was more than three times as high when compared with a targeting strategy that does not account for network structure. When network structure is taken into account, we show that both first- and second-order degree matter.

A practical concern of applying the proposed strategy of seeding network members with many friends who themselves have few friends might be that these people are rare in real-world networks. This is because the friends of network members with many friends typically also have many friends, indicating positive degree assortativity. The degree assortativity in the networks of Studies 1 and 2 are 0.04 and 0.23, respectively, numbers that are in line with what the literature reported for other social networks (Newman 2003, Ugander et al. 2012). Importantly, we were still able to identify effective seeds in both our empirical studies despite the networks’ positive degree assortativity.

5.2. Future Research Directions

We have shown that competition for attention plays an important role in information-propagation processes and that this has strong implications for seeding strategies. Because our results hold across many campaigns involving different content, different sharing processes, and different platforms, we expect competition for attention to be important for information propagation on social media in general. Given the presence of competition for attention, future research could explore other factors that might influence its strength, such as relational, timing, and campaign-content factors. For example, information received from close friends might compete less for attention than information received from acquaintances because of the nature of the relationship. Similarly, competition for attention might be weaker during

certain times of the day or days of the week. As richer data become available—for example, on time-stamped interactions between friends in a social network—future research might extend our seeding strategy with such relational and timing information. Campaign content may play an essential role in breaking through the clutter and mitigating the negative impact of information overload. Although previous research has studied how to keep people engaged when watching online videos (e.g., Teixeira et al. 2012), it remains unclear which content grabs attention on social media in the first place.

Last, it remains an open question whether competition for attention is also at play in social contagion processes more generally, such as new-product adoption or opinion formation (Ugander et al. 2012). Whereas forwarding a message is an active decision of a single person, social contagion may happen more passively and typically requires reinforcement (Hosanagar et al. 2010). Such contagion processes may also have a limited capacity problem. For example, adoption of the product of one company can impede adoption of products offered by competing companies. So far this phenomenon has only been studied by using a game-theoretic perspective (e.g., Borodin et al. 2010, Goyal and Kearns 2012). Combining game-theoretical and empirical insights, as our research shows, may be a promising road for future research to study competition for attention and contagion processes in social media research.

Acknowledgments

The authors thank the editorial team for their invaluable role in the review process of this paper, including constructive and knowledgeable feedback and insightful suggestions. The authors also thank Xi Chen and Tuan Q. Phan for help with

the data used in Study 2 and Klaas Weima for supporting their research.

Appendix A. Proofs of Equations (9) and (11)

Proof of Equation (9). The first-order conditions in Equation (8) can be rewritten as follows:

$$\begin{aligned} \alpha \mathbf{1}_N - I_N(x+s) + \beta A^T(x+s) + \beta s &= \mathbf{0}, \\ (I_N - \beta A^T)x &= \alpha \mathbf{1}_N - (I_N - \beta A^T)s + \beta s, \\ x &= x^*(\mathbf{0}) - s + \beta(I_N - \beta A^T)^{-1}s, \\ x &= x^*(\mathbf{0}) + (\beta - 1)s + \beta^2(I_N - \beta A^T)^{-1}A^T s. \end{aligned}$$

The proof of the fourth equality is as follows:

$$\begin{aligned} (I_n - \beta A^T)^{-1} \mathbf{1}_n &= \mathbf{1}_n + \beta(I_n - \beta A^T)^{-1}A^T \mathbf{1}_n, \\ \mathbf{1}_n &= (I_n - \beta A^T) \mathbf{1}_n + \beta A^T \mathbf{1}_n, \\ \mathbf{1}_n &= \mathbf{1}_n - \beta A^T \mathbf{1}_n + \beta A^T \mathbf{1}_n, \end{aligned}$$

which proves Equality (9). □

Proof of Equation (11). Note that $(I_N - \beta A^T)^{-1} = I_N + \sum_{i=1}^{\infty} \beta^i (A^T)^i$; hence $(I_N - \beta A^T)^{-1}A^T = A^T(I_N - \beta A^T)^{-1}$. Therefore,

$$\begin{aligned} \mathbf{1}_N^T \beta^2 (I_N - \beta A^T)^{-1} A^T s &= \mathbf{1}_N^T \beta^2 A^T (I_N - \beta A^T)^{-1} s \\ &= s^T \beta^2 (I_N - \beta A)^{-1} A \mathbf{1}_N, \end{aligned}$$

which proves Equality (11). □

Appendix B. Simulation Study on the Accuracy of the Truncated Bonacich Centrality Measure

Bonacich centrality is defined by the following infinite sum (Equation (12)):

$$B(A, \beta) = A \mathbf{1}_n + \beta A^2 \mathbf{1}_n + \beta^2 A^3 \mathbf{1}_n + \beta^3 A^4 \mathbf{1}_n + \dots$$

Bonacich (1987) showed that if $|\beta|$ is smaller than the inverse of the largest eigenvalue of the adjacency matrix (A), the higher-order terms are less important in the sum. Thus, for sufficiently small values of $|\beta|$, most of the variation in the

Table B.1. Correlation Between Nontruncated and Truncated Bonacich Centrality

Network size (N)	β as a fraction of maximum bound	Truncation order (k)			
		$k = 2$	$k = 3$	$k = 5$	$k = 10$
1,000	0.9	0.866	0.910	0.959	0.991
1,000	0.5	0.993	0.998	1.000	1.000
1,000	-0.5	0.992	0.997	1.000	1.000
1,000	-0.9	0.839	0.843	0.915	0.983
5,000	0.9	0.828	0.883	0.944	0.988
5,000	0.5	0.992	0.998	1.000	1.000
5,000	-0.5	0.994	0.998	1.000	1.000
5,000	-0.9	0.838	0.851	0.919	0.983
10,000	0.9	0.813	0.871	0.937	0.986
10,000	0.5	0.993	0.998	1.000	1.000
10,000	-0.5	0.995	0.998	1.000	1.000
10,000	-0.9	0.850	0.864	0.926	0.984

Bonacich centrality measure is captured by the lower-order terms degree ($A1_n$) and second-order degree (A^21_n).

To verify the accuracy of the truncation at the second-order degree, we ran a simulation study. We simulated undirected networks of size $N \in \{1,000; 5,000; 10,000\}$ using preferential attachment (Barabási and Albert 1999). In each network, we computed the untruncated value of Bonacich centrality (Equation (11)) and the truncated version up to k th-order-degree friends (in our empirical application, we use $k = 2$; see Equation (12)). Because $|\beta|$ is bounded by the inverse of the largest eigenvalue of the adjacency matrix, we set β equal to $\{0.9, 0.5, -0.5, -0.9\}$ times its upper bound. Table B.1 reports the correlation between the truncated and the nontruncated Bonacich centralities, averaged over 100 simulated networks, for each network size N . From this simulation, we conclude that our truncation provides an accurate approximation of the Bonacich centrality measure, even when β is close to its bound. Moreover, adding higher-order terms ($k = 3, k = 5, \text{ or } k = 10$) only leads to small improvements in the accuracy of the truncated measure. We believe that this provides important support for use of the truncated Bonacich measure.

Endnotes

¹ See <http://www.nytimes.com/2014/03/14/business/media/a-kiss-is-just-a-kiss-unless-its-an-ad-for-a-clothing-company.html>; accessed October 11, 2019.

² See <http://www.businessinsider.com/wren-first-kiss-viral-success-2014-3>; accessed October 11, 2019.

³ See <https://www.businessinsider.com/how-doves-real-beauty-sketches-became-the-most-viral-ad-video-of-all-time-2013-5?IR=T>; accessed October 11, 2019.

⁴ We use the term *friends* to refer to all connections in the egocentric network of an individual, which may also involve acquaintances, colleagues, or anyone else to whom someone is connected.

⁵ We allow the adjacency matrix to be either symmetric (undirected networks) or asymmetric (directed networks) to maintain the generalizability of our findings, whereas our empirical studies use undirected networks only.

⁶ Following the notation of Sendiña-Nadal et al. (2016), we set the simulation parameters as follows: penetration depth $l = 1$, fraction of potential leaders $p = 0.5$, and each new node connecting to $m = 3$ existing nodes and 10 initialization nodes. This leads to a degree assortativity of 0.20, which is realistic in real-world networks (Newman 2003, Ugander et al. 2012).

⁷ Element a_{ij}^2 of matrix A^2 counts the number of paths of length 2 between individuals i and j and thus captures second-order connections. If i and j are not directly connected (i.e., $a_{ij} = 0$), but have one common friend, then $a_{ij}^2 = 1$.

⁸ We have included a simulation study on the accuracy of the approximation in Appendix B.

⁹ See <http://www.angrybirds.com/>; accessed June 3, 2019.

¹⁰ Because we do not observe the full network in Study 1, we could not compute eigenvector centrality exactly. Therefore, we approximate it by setting the connectivity parameter equal to the absolute value of the estimated parameter in the full model. A direct comparison with eigenvalue centrality is given in Study 2.

¹¹ Although these estimates seem small, they are economically significant because they are multiplied by large numbers: average degree of seeds = 145.2 (SD = 193.7, minimum (min) = 1, maximum (max) = 27,363) and average second degree = 252,669 (SD = 509,220, min = 11, max = 23,496,245).

¹² The area under the curve (AUC) is the proportion of the unit square that is located under the curve for each targeting strategy. The AUC takes values between zero and one, and higher AUC values are better. See Müller et al. (2016) for an example of AUC in big-data analytics for information systems research.

¹³ Similar to Study 1, these estimates are economically significant because they are multiplied by large numbers: average degree of seeds = 67.9 (SD = 66.2, min = 1, max = 600) and average second degree = 9,908.2 (SD = 12,761.3, min = 8, max = 107,102).

¹⁴ If there is a selection bias in Study 2, there would exist a missing variable that explains both the decision to share information and the reach of a campaign. However, as illustrated by Pearl (1995), for the selection bias to hold, this variable should also relate to the number of messages that friends of seeds receive after controlling for seeds' network position and reach.

¹⁵ Because we do not observe sufficient repeated observations in sharing decisions across individuals, we are unable to estimate unobserved heterogeneity of the connectivity parameter.

References

- Ansari A, Koenigsberg O, Stahl F (2011) Modeling multiple relationships in social networks. *J. Marketing Res.* 48(4):713–728.
- Aral S, Dhillon PS (2018) Social influence maximization under empirical influence models. *Nature Humam Behav.* 2(6):375–382.
- Aral S, Van Alstyne M (2011) The diversity-bandwidth trade-off. *Amer. J. Sociol.* 117(1):90–171.
- Aral S, Walker D (2011) 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 (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341.
- Aral S, Walker D (2014) Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Sci.* 60(6):1352–1370.
- Aral S, Muchnik L, Sundararajan A (2013) Engineering social contagions: Optimal network seeding in the presence of homophily. *Network Sci.* 1(2):125–153.
- Asur S, Huberman BA, Szabo G, Wang C (2011) Trends in social media: Persistence and decay. Nicolov N, Shanahan JG, eds. *Proc. 5th Internat. AAAI Conf. Weblogs Soc. Media* (Association for the Advancement of Artificial Intelligence, Palo Alto, CA), 434–437.
- Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: Quantifying influence on Twitter. *Proc. 4th ACM Internat. Conf. Web Search and Data Mining* (ACM, New York), 65–74.
- Bakshy E, Karrer B, Adamic LA (2009) Social influence and the diffusion of user-created content. *EC'09 Proc. 10th ACM Conf. Electronic Commerce* (ACM, New York), 325–334.
- Ballester C, Calvó-Armengol A, Zenou Y (2006) Who's who in networks? Wanted: The key player. *Econometrica* 74(5):1403–1417.
- Bampo M, Ewing MT, Mather DR, Stewart D, Wallace M (2008) The effects of the social structure of digital networks on viral marketing performance. *Inform. Systems Res.* 19(3):273–290.
- Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO (2013) The diffusion of microfinance. *Science* 341(6144):1236498.
- Banerjee A, Chandrasekhar AG, Dulfo E, Jackson MO (2019) Using gossips to spread information: Theory and evidence from a randomized controlled trial. *Rev. Econom. Stud.* 86(6): 2453–2490.
- Bapna R, Umyarov A (2015) Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Management Sci.* 61(8):1902–1920.
- Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512.

- Bearman PS, Moody J, Stovel K (2004) Chains of affection: The structure of adolescent romantic and sexual networks. *Amer. J. Sociol.* 110(1):44–91.
- Berger J, Milkman KL (2012) What makes online content viral? *J. Marketing Res.* 49(2):192–205.
- Blume LE, Easley D, Kleinberg J, Tardos E (2009) Trading networks with price-setting agents. *Games Econ. Behav.* 67(1):36–50.
- Bonacich P (1987) Power and centrality: A family of measures. *Amer. J. Sociol.* 92(5):1170–1182.
- Borodin A, Filmus Y, Oren J (2010) Threshold models for competitive influence in social networks. Saberi A, ed. *Internet and Network Economics WINE 2010*, Lecture Notes in Computer Science, vol. 6484 (Springer, Berlin), 539–550.
- Braun M, Bonfrer A (2011) Scalable inference of customer similarities from interactions data using Dirichlet processes. *Marketing Sci.* 30(May–June):513–531.
- Burt RS (1997) The contingent value of social capital. *Admin. Sci. Quart.* 42(2):339–365.
- Burt RS (2004) Structural holes and good ideas. *Amer. J. Sociol.* 110(2):349–399.
- Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197.
- Chen X, van der Lans R, Phan TQ (2017) Uncovering the importance of relationship characteristics: Implications for seeding strategies. *J. Marketing Res.* 54(2):187–201.
- Cook KS (1982) Network structures from an exchange perspective. Marsden P, Lin N, eds. *Social Structure and Network Analysis* (Sage, Thousand Oaks, CA), 177–199.
- Cook KS, Emerson RM, Gillmore MR, Yamagishi T (1983) The distribution of power in exchange networks: Theory and experimental results. *Amer. J. Sociol.* 89(2):275–305.
- De Bruyn A, Lilien GL (2008) A multi-stage model of word of mouth through viral marketing. *Internat. J. Res. Marketing* 25(3):151–163.
- Efron B (1985) Bootstrap confidence intervals for a class of parametric problems. *Biometrika* 72(1):45–58.
- Feld SL (1991) Why your friends have more friends than you do. *Amer. J. Sociol.* 96(6):1464–1477.
- Feng L, Hu Y, Li B, Stanley HE, Havlin S, Braunstein LA (2015) Competing for attention in social media under information overload conditions. *PLoS One* 10(7):e0126090.
- Goel S, Watts DJ, Goldstein DG (2012) The structure of online diffusion networks. *Proc. 13th ACM Conf. Electronic Commerce (EC'12)* (ACM, New York), 623–638.
- Goel S, Anderson A, Hofman J, Watts DJ (2016) The structural virality of online diffusion. *Management Sci.* 62(1):180–196.
- Goldenberg J, Han S, Lehmann DR, Hong JW (2009) The role of hubs in the adoption process. *J. Marketing* 73(2):1–13.
- Goldfarb A, Tucker C (2011) Online display advertising: Targeting and obtrusiveness. *Marketing Sci.* 30(3):389–404.
- Goyal S, Kearns M (2012) Competitive contagion in networks. *Proc. 44th Sympos. Theory Comput.* (ACM, New York), 759–774.
- Granovetter MS (1973) The strength of weak ties. *Amer. J. Sociol.* 78(6):1360–1380.
- Harush U, Barzel B (2017) Dynamic patterns of information flow in complex networks. *Nature Comm.* 8(1):2181.
- Heiss F, Winschel V (2008) Likelihood approximation by numerical integration on sparse grids. *J. Econometrics* 144(1):62–80.
- Hinz O, Skiera B, Barrot C, Becker JU (2011) Seeding strategies for viral marketing: An empirical comparison. *J. Marketing* 75-(November):55–71.
- Hosanagar K, Han P, Tan Y (2010) Diffusion models for peer-to-peer (P@P) media distribution: On the impact of decentralized, constrained supply. *Inform. Systems Res.* 21(2):271–287.
- Hu Y, Van den Bulte C (2014) Nonmonotonic status effects in new product adoption. *Marketing Sci.* 33(4):509–533.
- Iyer G, Katona S (2016) Competing for attention in social communication markets. *Management Sci.* 62(8):2304–2320.
- Iyengar R, Van den Bulte C, Valente TW (2011) Opinion leadership and social contagion in new product diffusion. *Marketing Sci.* 30(12):195–212.
- Jackson MO (2008) *Social and Economic Networks*, vol. 3 (Princeton University Press, Princeton, NJ).
- Kane GC, Alavi M, Labianca GJ, Borgatti S (2012) What's different about social media networks? A framework and research agenda. *MIS Quart.* 38(1):275–304.
- Katona Z, Zubcsek PP, Sarvary M (2011) Network effects and personal influences: The diffusion of an online social network. *J. Marketing Res.* 48(3):425–443.
- Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. *Nature Phys.* 6(11):888.
- Kumar A, Bezawada R, Rishika R, Janakiraman R, Kannan PK (2016) From social to sale: The effects of firm-created content in social media on customer behavior. *J. Marketing* 80(1):7–25.
- Lanz A, Goldenberg J, Shapira D, Stahl F (2019) Climb or jump: Status-based seeding in user-generated content networks. *J. Marketing Res.* 56(3):361–378.
- Lee SH, Cotte J, Noseworthy TJ (2010) The role of network centrality in the flow of consumer influence. *J. Consumer Psych.* 20(1):66–77.
- Libai B, Muller E, Peres R (2013) Decomposing the value of word-of-mouth seeding programs: Acceleration vs. expansion. *J. Marketing Res.* 50(2):161–176.
- Lin KY, Lu HP (2011) Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Comput. Human Behav.* 27(3):1152–1161.
- Lobel I, Sadler E, Varshney LR (2016) Customer referral incentives and social media. *Management Sci.* 63(10):3514–3529.
- Lokhov AY, Saad D (2017) Optimal deployment of resources for maximizing impact in spreading processes. *Proc. Natl. Acad. Sci. USA* 114(39):E8138–E8146.
- Markovsky B, Willer D, Patton T (1988) Power relations in exchange networks. *Amer. Sociol. Rev.* 53(2):220–236.
- Müller O, Junglas I, Brocke Jv, Debortoli S (2016) Utilizing big data analytics for information systems research: Challenges, promises and guidelines. *Eur. J. Inform. Systems* 25(4):289–302.
- Newman ME (2003) Mixing patterns in networks. *Phys. Rev. E* 67(2):026126.
- Pearl J (1995) Causal diagrams for empirical research. *Biometrika* 82(4):669–710.
- Phillip MV, Suri R (2004) Impact of gender differences on the evaluation of promotional emails. *J. Advertising Res.* 44(4):360–368.
- Pieters R, Wedel M (2004) Attention capture and transfer in advertising: Brand, pictorial, and text-size effects. *J. Marketing* 68(2):36–50.
- Pieters R, Wedel M, Zhang J (2007) Optimal feature advertising design under competitive clutter. *Management Sci.* 53(11):1815–1828.
- Rishika R, Kumar A, Janakiraman R, Bezawada R (2013) The effect of customers' social media participation on customer visit frequency and profitability: An empirical investigation. *Inform. Systems Res.* 24(1):108–127.
- Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.
- Seidman SB (1983) Network structure and minimum degree. *Soc. Networks* 5(3):269–287.

- Sendiña-Nadal I, Danziger MM, Wang Z, Havlin S, Boccaletti S (2016) Assortativity and leadership emerge from anti-preferential attachment in heterogeneous networks. *Sci. Rep.* 6:21297.
- Sun E, Rosenn I, Marlow CA, Lento TM (2009) Gesundheit! Modeling contagion through Facebook news feed. *Proc. Third Internat. Conf. Weblogs Soc. Media* (Association for the Advancement of Artificial Intelligence, Palo Alto, CA), 146–153.
- Tang Y, Xiao X, Shi Y (2014) Influence maximization: Near-optimal time complexity meets practical efficiency. *Proc. 2014 ACM SIGMOD Internat. Conf. Management Data* (ACM, New York), 75–86.
- Teixeira T, Wedel M, Pieters R (2012) Emotion-induced engagement in Internet video advertisements. *J. Marketing Res.* 49(2):144–159.
- Trusov M, Bodapati A, Bucklin R (2010) Determining influential users in Internet social networks. *J. Marketing Res.* 47(4):643–658.
- Tucker C (2008) Identifying formal and informal influence in technology adoption with network externalities. *Management Sci.* 54(12):2024–2038.
- Ugander J, Backstrom L, Marlow C, Kleinberg J (2012) Structural diversity in social contagion. *Proc. Natl. Acad. Sci. USA* 109(16):5962–5966.
- Valente TW (2012) Network interventions. *Science* 337(6090):49–53.
- van Dam J-W, van de Velden M (2015) Online profiling and clustering of Facebook users. *Decision Support Systems* 70:60–72.
- van der Lans R, Pieters R, Wedel M (2008) Competitive brand salience. *Marketing Sci.* 27(5):922–931.
- van der Lans R, van Bruggen GH, Eliashberg J, Wierenga B (2010) A viral branching model for predicting the spread of electronic word-of-mouth. *Marketing Sci.* 29(2):348–365.
- Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *J. Consumer Res.* 34(December):441–458.
- Watts DJ, Peretti J (2007) Viral marketing for the real world. *Harvard Bus. Rev.* (May):22–23.
- Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. *Nature Sci. Rep.* 2(335):1–9.
- Wu F, Huberman BA (2007) Novelty and collective attention. *Proc. Natl. Acad. Sci. USA* 104(45):17599–17601.
- Yamagishi T, Gillmore MR, Cook KS (1988) Network connections and the distribution of power in exchange networks. *Amer. J. Sociol.* 93(4):833–851.
- Yoganarasimhan H (2012) Impact of social network structure on content propagation: A study using YouTube data. *Quant. Marketing Econom.* 10(1):111–150.
- Zubcsek P, Sarvary M (2011) Advertising to a social network. *Quant. Marketing Econom.* 9(1):71–107.