Information Systems Research

The Effects of the Social Structure of Digital Networks on Viral Marketing Performance

Mauro Bampo, Michael T. Ewing, Dineli R. Mather, David Stewart, Mark Wallace,

Please scroll down for article—it is on subsequent pages

The Effects of the Social Structure of Digital Networks on Viral Marketing Performance

Mauro Bampo
School of Information Technology, Monash University, Melbourne, Australia,
mauro.bampo@infotech.monash.edu.au

Michael T. Ewing
Department of Marketing, Monash University, Melbourne, Australia,
michael.ewing@buseco.monash.edu.au

Dineli R. Mather
School of Engineering and Information Technology, Deakin University, Melbourne, Australia,
dineli.mather@deakin.edu.au

David Stewart
Department of Marketing, Monash University, Melbourne, Australia,
david.stewart@buseco.monash.edu.au

Mark Wallace
School of Information Technology, Monash University, Melbourne, Australia,
mark.wallace@infotech.monash.edu.au

Viral marketing is a form of peer-to-peer communication in which individuals are encouraged to pass on promotional messages within their social networks. Conventional wisdom holds that the viral marketing process is both random and unmanageable. In this paper, we deconstruct the process and investigate the formation of the activated digital network as distinct from the underlying social network. We then consider the impact of the social structure of digital networks (random, scale free, and small world) and of the transmission behavior of individuals on campaign performance. Specifically, we identify alternative social network models to understand the mediating effects of the social structures of these models on viral marketing campaigns. Next, we analyse an actual viral marketing campaign and use the empirical data to develop and validate a computer simulation model for viral marketing. Finally, we conduct a number of simulation experiments to predict the spread of a viral message within different types of social network structures under different assumptions and scenarios. Our findings confirm that the social structure of digital networks play a critical role in the spread of a viral message. Managers seeking to optimize campaign performance should give consideration to these findings before designing and implementing viral marketing campaigns. We also demonstrate how a simulation model is used to quantify the impact of campaign management inputs and how these learnings can support managerial decision making.

Key words: digital communication; social structure of digital networks; viral marketing
History: Anil Gupta, Senior Editor. This paper was received on July 14, 2006, and was with the authors 5 months for 2 revisions. Published online in *Articles in Advance* June 5, 2008.

1. Introduction

The term "viral marketing" appears to have first been coined by venture capitalist Steve Jurvetson in 1996 to describe the marketing strategy of free e-mail service Hotmail (Kaikati and Kaikati 2004). Since then, contemporary business literature has become somewhat enamoured with the concept. Terms such as "word-of-web," "word-of-mouse," "customer-to-customer" (C2C), or "peer-to-peer" (P2P) communication as well as "buzz marketing" have also been variously associated with the process. The term "buzz," almost by definition, has an ephemeral connotation. We concur with Dobele et al. (2005) and view "buzz" as an output or consequence of viral marketing. The viral metaphor neatly captures the essence of the communications process and draws on a rich body

of literature in medicine and the health sciences (e.g., Mather and Crofts 1999, Mather 2000).

Viral marketing broadly describes any strategy that encourages individuals to propagate a message, thus creating the potential for exponential growth in the message's exposure and influence. Kaikati and Kaikati (2004) view it as "...word of mouth via a digital platform...spreading the message via 'word of mouse' and ensuring that the receivers have the interest to pass along the message to their acquaintances." Similarly, Dobele et al. (2005) describe it as "encouraging individuals to pass on messages received in a hypermedia environment, such as e-mail or other messaging system."

Viral approaches have numerous advantages over more traditional mass media. For example, there is a natural selection process embedded in the way the message is propagated. This reduces redundancy in the sense that communication is more targeted. Other advantages include speed of diffusion and a reduced likelihood for the message to be altered by senders (in other words, a high degree of message integrity). And, if the message has an embedded call to action, then the conversion rate (i.e., behavioral response) is potentially more quantifiable than in other forms of mass communication. Viral communication also affords the marketer a greater degree of creative license through a message delivery medium that is more intimate and personalized, thereby increasing the likelihood of reaching "hard-to-get" audience members.

The viral process can be broadly modelled in terms of three components: the social structure of the digital network through which the message is propagated, the behavioral characteristics of its members that facilitate the propagation of the message, and a seeding strategy that initiates the process. This study is based on the model introduced by Stewart et al. (2004), where the underlying social network is represented by a random graph and network members' behavior is defined by the susceptible-infective-removed (SIR) pattern from epidemic theory (Becker 1989).

The objectives of the study are threefold: first, to identify alternative social network models and to understand the mediating effects of social structures of these models on viral marketing campaigns; second, to develop a process for modelling viral marketing campaigns and empirically validate the ensuing activated digital network model; and third, to conduct a number of simulation experiments to explore the influence of the various controlled and external factors on viral campaigns.

The article is set out as follows. First, we examine the viral marketing process and define various campaign performance metrics. We then model different social structures as digital networks and define parameters which describe the network and control the spread of the viral message. Next, we introduce empirical data from a recent viral marketing campaign carried out by a leading automotive manufacturer, and develop a computer simulation model. We then describe a number of simulations which predict the spread of a viral message within different kinds of social networks under different assumptions about the network itself, the behavioral characteristics of its members, and the seeding strategy that initiates the process. Finally, we draw conclusions about the mediating effect of the social structure of digital networks on campaign performance as well as the extent of the control available to campaign managers.

2. Literature Review

The transition from traditional word-of-mouth networks to digital networks has greatly expanded the opportunities for bidirectional communication (Dellarocas 2003) and, in the process, created a pervasive and intriguing phenomenon (Goldenberg et al. 2001) that has piqued the attention of researchers from diverse disciplinary backgrounds. In reviewing this rich body of cross-disciplinary literature, two emerging streams are discernable: namely, a behavioral stream (incorporating advertising and marketing) and a management science stream—with strong foundations in information systems and operations research.

The behavioral stream has focused on characteristics, motivations, and reported behaviors of customers and the extent to which these might influence the success of viral marketing campaigns. This has included surveys of intended purchasing behaviors (Gruen et al. 2006) as well as examining the interactions between customer and product characteristics (Helm 2000) and their effects on message transmission.

Findings across the board in this research stream suggest that inherent customer heterogeneity warrants highly segmented viral campaigns to address individual customer differences and preferences. Message customization and social network status are two salient antecedents to determining the "spread" of the communication (Phelps et al. 2004, Podoshen 2006). A third key determinant is customer motivations (Gelb and Sundaram 2002, Phelps et al. 2004), be they intrinsic or extrinsic (e.g., reward programs, competitions, coupons).

The behavioral research stream is generally more applied and is aimed at practising managers looking to utilise digital social networks and online word-of-mouth in a more effective manner. Sophisticated targeting strategies are suggested and discussed (Dobele et al. 2005) and much emphasis is placed on accurate initial targeting (seeding) of customers (Phelps et al. 2004).

Despite the early progress made by the behavioral researchers, progress in this area has been somewhat limited by virtue of the character of viral marketing and word-of-mouth networks—there are ethical ramifications regarding consumer privacy (Phelps et al. 2004) if researchers were to accurately track and record data on *all* consumer interactions during a particular viral campaign. Additionally, important information as to *why* consumers propagate viral marketing messages, such as emotional engagement with the message (Dobele et al. 2007) or why consumers seek or provide an opinion for word-of-mouth networks (Goldsmith and Horowitz 2006), is difficult to obtain without directly interviewing or surveying the consumer. To do so invites the possibility of experimenter expectancy effects (Miller and Turnbull 1986, Rosenthal 1994) and potentially influencing the natural activity of passing on the marketing message. Furthermore, given the uncontrollable and "explosive" nature of the spread of viral marketing campaigns and online word-of-mouth networks (Dobele et al. 2005), accurate sampling of the population reached by a viral campaign is problematic.

More formalized studies are also needed to progress beyond extrapolated knowledge gleaned from (often modest) samples of customers (or students) to larger, more heterogeneous, "real world" populations—in other words, to model actual behavior rather than intended or reported behavior (Gelb and Sundaram 2002, Gruen et al. 2006, Helm 2000). The majority of behavioral research has also been limited to snapshot studies (Ba and Pavlou 2002, Gruen et al. 2006), with little opportunity for longitudinal studies to gauge the full extent of viral marketing campaigns on consumers in the natural setting. Small-sized samples (Weinberg and Davis 2005) and constrained populations from which samples have been drawn (Phelps et al. 2004), limitations common to research of this type, also reduce the generalisability of the behavioral research findings in this area. In addition, theoretical explanations for how viral marketing functions (Phelps et al. 2004) would further enhance understandings and applications of this area.

The management science stream, in contrast, has focused more on the design aspects of specific mechanisms (especially online reputation feedback mechanisms) and on the potential for influencing performance through deliberate structural and design manipulation. Using given systems parameters and different theoretical approaches (such as game theory), this body of work exploits mathematical modelling approaches to studying online communication networks. In particular, the work of Dellarocas (2003, 2005) illustrates how the design of a given system (such as an eBay-like reputation mechanism) can engender, support, and elicit certain responses from customers rather than relying on customer-initiated behavior. Other researchers in this stream have introduced trust into their models, not as a preexisting characteristic of the customer but as a construct engendered by the system itself (Ba and Pavlou 2002, Pavlou and Gefen 2004). Trust is crucial in this context given that the anonymity of online members and lack of actual context (Dellarocas 2003) increases the opportunity for online fraud (Bolton et al. 2004). The building of online reputations has even been modelled as a capital asset that must be maintained and invested in Rob and Fishman (2005).

This body of management science literature potentially provides a proactive approach to afford greater control over the performance of online mechanisms. The mathematical models are elegant and sophisticated in their execution and provide a solid framework (with given assumptions) on which knowledge of online processes can be further developed.

However, the focus of this body of work tends to be on the characteristics of the systems, not the characteristics and behaviors of the customers. As such, many are limited by parameter assumptions that are not replicated in real-world applications. Characteristics and behaviors of the online consumers are often unknown variables within mathematical models. A range of consumer characteristics, for example, intrinsic personal motivations (Bolton et al. 2004) and socio-economic status (Ba and Pavlou 2002, Gruen et al. 2002) are difficult to quantify in models, and behavior that falls outside of model parameters is equally problematic (Dellarocas 2005). Additionally, some models looking at consumer-generated feedback assume that feedback is truthful and not manipulated (Bolton et al. 2004). Even when manipulation is factored as a variable in Internet opinion forum models, there is an underlying assumption that consumers are competent to gauge levels of manipulation by companies (Dellarocas 2006) to adjust their own behavior accordingly. Research in this area has indicated a need to further understand design and parameter implications, for developing responses such as trust in consumers (Pavlou and Gefen 2004), and for controlling the behavior of users, for example, identity fraud (Dellarocas 2003).

A third avenue of inquiry appears to be emerging that has the potential to bridge the aforementioned two literature streams. Mayzlin and her colleagues (Chevalier and Mayzlin 2006, Godes and Mayzlin 2004, Mayzlin 2006) draw on both behavioral and management science traditions and approach customer-generated characteristics and behaviors as constructs which can be utilised as known quantities in mathematical models. Godes and Mayzlin's (2004) work, in particular, applies real-world data to a model to reveal which components of word-of-mouth communication are most effective. Components such as reach (Godes and Mayzlin 2004), quality of networks (Goldenberg et al. 2001), and quality of the message or feedback (Chevalier and Mayzlin 2006) can be modelled. Such information can potentially assist managers in predicting the usefulness of a particular word-of-mouth strategy and the potential for flow-on marketing to continue even after the initial advertising has ended (Goldenberg et al. 2001).

Notwithstanding the emerging bodies of literature in this field, notable gaps still exist. In particular, Dellarocas' (2003) call for more research on feedback mechanism design can be extended more broadly to other aspects of viral marketing/P2P communication, including theory-driven experimental and empirical research that explores the global impact of buyer-seller behavior and a better understanding of how managers must adapt their strategies in online contexts. Our study builds on Mayzlin's foundations and goes some way toward addressing Dellarocas' call. In so doing, it aims to assist firms to develop more formalised and sophisticated approaches to viral marketing (Helm 2000).

3. Deconstructing the Viral Process

We deconstruct the viral marketing process into the following components: underlying population and their social connectivity; the campaign characteristics; the behavioral characteristics of the audience that facilitates the propagation of the message; the size and connectivity of the successfully reached audience; and measures of dynamic campaign progress. Specifically, we model the size and connectivity of the population as a network, taking into account the campaign characteristics. We then simulate the campaign, and the campaign performance measures are reflected as properties of the simulation.

3.1. The Structure of a Digital Network

A network is specified by a set of *nodes* and a set of *edges* linking pairs of nodes. The nodes represent members of the population, or audience, and the edges represent communication links between them that may be used to spread the viral message. The *degree* of a node is the number of edges linking it to other nodes. Two nodes are connected if there is a sequence of edges forming a path from one node to the other. Thus, a node with degree zero is not connected to any other node. The *distance* between two connected nodes is the length of the shortest path connecting them.

Three properties of a network that will be used in this paper are: (i) its number of nodes N, (ii) the average degree of its nodes λ, and (iii) the *average distance between pairs of nodes L*. The parameter λ is a measure of *network connectedness*.

3.1.1. Random Networks. Stewart et al. (2004) introduced a random viral marketing model (RVM) based on the random network model developed by Erdös and Rényi (1959) and described by Albert and Barabási (2002). This digital network is represented by a random graph and network members' behavior is defined by the susceptible-infective-removed sequence (Becker 1989). A random network can be generated by starting with a set of isolated nodes and allowing each of the N nodes to have a probability θ of being connected by an edge to each other node.

As noted by Albert and Barabási (2002), in a random network the degree of its nodes follows a binomial distribution with parameters $N - 1$ and θ. As each node has the potential to connect up to $N - 1$ other nodes, on average we expect each node to be connected to $\lambda = \theta(N - 1)$ other nodes, resulting in an expected total of $\frac{1}{2}\lambda N$ links. In the context of viral marketing, a typical network has large N and small θ resulting in the average degree λ remaining moderate. The degree of a node therefore has an approximate Poisson distribution with mean network connectedness λ. Because the standard deviation is $\sqrt{\lambda}$, it is very unlikely for a node to have degree of size comparable with N. In other words, it is unlikely that any node is directly linked to a significant proportion of the nodes in the network.

3.1.2. Scale-Free Networks. Research into scale-free networks has proliferated since their introduction by Barabási (1999) and Albert and Barabási (2002). These networks provide useful representations of many different self-organizing systems, ranging from the World Wide Web to citation patterns in scientific publications to the electrical power grid of western United States. The defining characteristic of a scale-free network is in the shape of the probability distribution for the degree of each node, which determines the number of communication links or edges emanating from each member. The *degree* is assumed to follow a Power-law distribution, defined by $P(k) \propto k^{-\gamma}$ with $\gamma > 0$, where $P(k)$ denotes the probability that a node is connected to k other nodes. This is a "fat-tailed" distribution where, with increasing k, the probabilities decline at a much slower rate than those of the Poisson distribution which essentially underlies the RVM model. The Power-law distribution allows for a small number of nodes to be directly linked to

a significant proportion of the nodes in the network while most nodes have few connections, thus keeping the mean number of connections comparatively low. These high degree nodes, often called *hub*s, ensure that the average distance L between any two nodes in the network is small (independent of the size of the network).

The scale-free network structure emerges naturally as a consequence of two phenomena: dynamic growth and preferential attachment (Barabási 1999, Albert and Barabási 2002), both important features of social networks. Where a network is created by adding new members over time and these are connected to other members with a probability that is proportional to their connectivity, the resulting distribution for the degree or number of connections per node will exhibit a Power-law distribution. These structures are called scale-free networks because despite their growth, they preserve statistical properties such as the average distance L and the degree distribution.

Some studies (Dorogovtsev and Mendes 2003, Drineas et al. 2004), based on analysis of the traffic on SMTP servers, show that e-mail networks of limited size (involving members of a single university) display properties of a scale-free network. e-mail graphs were constructed in those studies, representing e-mail addresses by nodes and adding a communication link between each pair of nodes where at least one message had passed between them. Both the number of incoming and outgoing links have been shown to follow a Power-law distribution. This feature of e-mail graphs makes scale-free networks particularly interesting from a viral marketing perspective. To create a scale-free graph, nodes are added one by one to the network; every new node is linked by an undirected arc to a preexisting node l with probability

$$P(l) = \frac{\lambda}{2} \cdot \frac{d_l}{\sum_m d_m},$$

d_l being the degree of node l, $\{m\}$ the set of previously added nodes, and λ the desired average degree of the network. Theoretical models of the spread of diseases and the absence of epidemic thresholds in scale-free networks are discussed at length in Boguna et al. (2003), Eguıluz and Klemm (2002), Moreno and Vázquez (2003), and Pastor-Satorras and Vespignani (2001).

3.1.3. Small World Networks.

Small world graphs were first introduced by Watts and Strogatz (1998) to model a class of social networks characterized by high clustering and short average distance between nodes. Clustering is a local property of the network and is a measure of the connectivity of a neighbourhood. The *clustering coefficient C* of a node is defined as the fraction of the node's neighbours that are linked to each other. High clustering and long average distances are typical features of lattice networks (Dorogovtsev and Mendes 2003), where nodes can be thought of as points in a multidimensional space and nearby points are linked by edges. In contrast, small world networks have short average distances between nodes.

Small world networks can be constructed from lattice networks by applying a rewiring procedure: arcs connecting neighbours (within the clusters) are removed from the graph with probability *rewiring probability r* and substituted by random links (making connections outside of the cluster). As *r* increases, the average distance *L* decreases very quickly (Watts and Strogatz 1998), producing a graph structure characterized by low node separation typical of random networks and strongly connected neighbourhoods of regular networks. With increasing *r*, the graph starts to become more like a random network. The transition, however, is smooth and the evolution of the average distance and level of clustering is also influenced by *N* (Barthélémy and Amaral 1999).

Small world networks are also potentially applicable to viral marketing because they capture the connections generated through physical proximity. Tightly linked neighbourhoods reflect social structures based on friendship or professional relationships which are likely to form among people who interact within a confined physical environment. For example, Albert and Barabási (2002) refer to a social system where people are well-connected with their neighbours and work colleagues but also have a much smaller number of connections with people who live far away, in another state or country. Random links represent the distant acquaintances and are useful in representing connections between local networks. A higher level of rewiring makes the viral message spread faster and thus saturates the network sooner.

3.2. Campaign Characteristics

The impact of a viral marketing campaign can be influenced by the message attractiveness, the campaign design, and any intervention strategies. The attractiveness or perceived value of a viral message as well as offering an incentive (if any) play an important role in determining a recipient's propensity to forward the communication as well as which communication links to activate from within their digital network connections. The campaign manager determines the number of seeds used, with seeding typically taking place at the start of a campaign. Once a campaign is in progress, there are a number of ways in which a campaign manager can track its progress. For example, if the campaign includes a call to action such as an online coupon or uses a Web interface as a registration process, it is possible to identify the signs of a flagging campaign and take corrective action to resuscitate it.

3.3. Modelling the Propagation of a Viral Message

The behavior of network nodes determines the propagation of the message through the network. Network propagation is modelled on a discrete time basis. Propagation along network edges occurs simultaneously at each time instant. Using the SIR sequence nodes can be in three states: (i) S—susceptible; (ii) I—infective; and (iii) R—removed (or "immune") and at any given time, the total number of nodes $N = S + I + R$. Each node is in a "susceptible" state before receiving the message. On receiving the message, a node becomes infected and remains "infective" for one time period when it may propagate (forward) the message along any of its edges according to a probability p_F (sampled from a probability distribution) which we refer to as the *forwarding parameter*. This is analogous to the *contagion parameter* in epidemic theory (Becker 1989), as in a digital context the contagion parameter refers to the probability of forwarding a message. After that time period, the node becomes "immune" to the message (removed) and takes no further part in the propagation process. Thus, we assume that any further messages reaching the node are ignored in the SIR sequence, as described by Moreno and Vázquez (2003).

We employ the concept of a *generation G* (Stewart et al. 2004) to identify the nodes reached by the message at each time instant, with the seeds forming the

initial (zeroth) generation of the network members reached. Hence, the generation can be used to index the sequence of transmissions. Under the SIR pattern, nodes forming each generation are the only infective ones.

3.4. The Activated Digital Network

Digital networks are best understood by considering the formation of digital connections as illustrated in Figure 1. We start with the *complete social network* which encapsulates all types of connections between the nodes; for example, family ties or social or professional connections. Some of these connections are digital or electronic and we refer to this "digital subset" of the complete network as the *underlying digital social network*. When an electronic message is received by an individual (node) within the underlying digital social network, he or she is faced with two decisions: first, whether or not to forward the message, according to the forwarding parameter p_F; and second, if he or she decides to forward the message, then choosing which existing digital links to activate. This latter process results in the creation of the *activated digital network* and is captured by an additional measure which we define as the *activation parameter p_A*. The parameters p_F and p_A are treated as stochastic in nature. This process is represented in Figure 1.

3.5. Performance Measures of a Campaign

Stewart et al. (2004) defines three output measures for viral campaigns: the process duration T_i; the number of network members reached at the tth generation $X_i(t)$; and the cumulative number reached up to and including the tth generation $Y_i(t)$. The index i denotes the number of seeds ($1 \leq i < N$) used in the campaign.

A campaign is said to naturally terminate when there are no longer any infectives. In this study, we use two performance measures for a viral campaign: the final *reach* or penetration, i.e., the proportion of the target audience that has received the communication by the time the process dies, and the *length of the campaign*, i.e., the number of generations required to reach a predetermined proportion of the target audience.

Stewart et al. (2004) show that the main parameter influencing the spread of the viral message is the *epidemic threshold parameter (ETP) μ* (Becker 1989), which measures the growth rate of infectives. The *ETP* is defined as the product of the network connectedness λ, the activation parameter p_A, and the forwarding parameter p_F (i.e., $\mu = p_F p_A \lambda$). In the early generations, the growth of the digital network is governed by the size of μ. When μ is significantly greater than one, the message is being forwarded, on average, to more than one individual and, hence, the number of infectives $X_i(t)$ grows at an exponential rate during the earlier generations. Borrowing terminology from the theory of branching processes (Becker 1989), we say the digital network exhibits *supercritical growth*. As the network becomes saturated (the proportion of removals increases), the growth rate is reduced to the point where it transmutes into an exponential decay. An important property of a supercritical network is that the eventual reach (penetration) rapidly approaches 100% of connected members and this is governed primarily by the *ETP*. In this case, the number of seeds used is important to the extent of ensuring that the propagation process does not terminate in the initial generations, but beyond that plays no role in determining the eventual reach. On the other hand, when μ is not significantly greater than one, the

Figure 1 Formation of an Activated Digital Network for Generations 1–2 Initiated by Four Seeds

Complete social network   Underlying digital network   Activated digital network
                                  $(N, \theta; \lambda)$ (N', p_F, p_A)
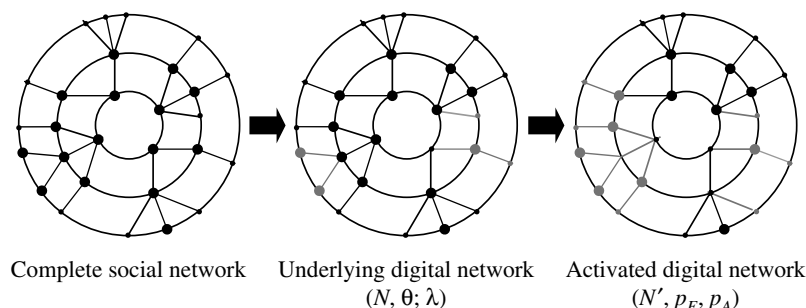
network exhibits *subcritical growth*, where the number of infectives $X_i(t)$ decays at an exponential rate almost from the beginning of the campaign. In the subcritical case, the number of seeds play an important role as a higher penetration can be achieved by seeding a wider proportion of the audience.

4. Design of the Simulation Model and Empirical Validation

4.1. Results from an Empirical Campaign

We analyse a viral marketing campaign conducted by a leading automotive manufacturer, General Motors Holden in Australia, which was considered to be highly successful. In its use of seeding and the propagation of the marketing message through digital links within the population, it was a typical digital marketing campaign. We simulate this campaign using alternative network models and match their results against the real campaign data. These results then enable us to validate the simulation model and infer characteristics of the real campaign. The knowledge of such characteristics can be used to help manage future campaigns.

In this digital automotive campaign, almost 39,000 self-selected target audience members were seeded with promotional information about a new product in the form of an e-brochure and were invited to provide e-mail addresses of contacts who might be interested in receiving the brochure themselves. The

company used the prize of a holiday with the likelihood of winning linked to the number of contacts nominated as an incentive. The campaign eventually reached an additional 43,000 people. Although the estimated target market (N) for the particular automotive model is approximately 171,000, it is unrealistic to assume that the campaign remained within the bounds of this group and, hence, the percentage reach that is implied by the numbers is very unlikely to be accurate.

The initial viral seeding (i.e., generation 0) was to 38,668 potential customers. Of these, 10,244 registered and generated at least 1 self-initiated outbound e-mail, and the total number of e-mails sent to the next generation (generation 1) was 26,548. Of these, 3,091 registered online and generated 9,089 new contacts (generation 2) in total. And of these, 1,221 went on to register and generated 3,858 new e-mails (i.e., generation 3). This process continued to the 13th generation. Table 1 presents the campaign statistics by generation t: the number of infectives who registered (n_t); calculated parameter values for the estimated probability of forwarding (that is, registering on the campaign Web site and providing e-mail addresses of their contacts) ($p_{F,t}$), the estimated average number of activated contacts per registered infective ($p_{A,t}\lambda_t$), and the growth rate of infectives (μ_t) at each generation. The number of people receiving the message in the later generations (7–13) was insufficient to make reasonable estimates.

Table 1   Campaign Statistics by Generation (t)

Gen. t	No. of susceptibles S_t	No. of infectives I_t	Cumulative no. of removals R_t	No. of infectives who registered n_t	Estimated prob. of forwarding $p_{F,t} = n_t/I_t$ (%)	Estimated average no. of activated contacts per registered infective $p_{A,t}\lambda_t = I_{t+1}/n_t$	Estimated growth rate of infectives $\mu_t = I_{t+1}/I_t$
0	132,339	38,661	0	10,244	26.5	2.6	0.7
1	105,791	26,548	38,661	3,091	11.6	2.9	0.3
2	96,702	9,089	65,209	1,221	13.4	3.2	0.4
3	92,844	3,858	74,298	564	14.6	3.2	0.5
4	91,039	1,805	78,156	279	15.5	3.1	0.5
5	90,188	851	79,961	147	17.3	3.0	0.5
6	89,745	443	80,812	72	16.3	3.0	0.5
7	89,527	218	81,255	39			
8	89,406	121	81,473	20			
9	89,326	80	81,594	12			
10	89,294	32	81,674	6			
11	89,279	15	81,706	2			
12	89,271	8	81,721	2			
13	89,265	6	81,729				

4.2. Network Model of the Campaign

The campaign data contains the communication links that were activated as the message was transmitted through the digital network. While it is not possible to identify the totality of the underlying social or digital network, the realised or activated digital network links are identified uniquely by the data. Closer analysis of this activated network shows while most of the individuals have moderate to low (less than 10) number of contacts, a small number of the audience members (less than 0.03%) forwarded the message to a significantly large number of people (more than 20 contacts, with a very small number forwarding to over 100). In other words, there is some semblance of a small number of large hubs (suggestive of scale-freeness). The data also shows that a majority of message forwarding takes place within the same state, but a significant minority (10%) of links are activated across different Australian states. While this provides some preliminary evidence of small world characteristics with a high level of local clustering and a small proportion of "long distance" connections, given the large geographic regions covered by the Australian states, inferring local clustering based on being in the same state may be questionable.

A generation-by-generation analysis of the campaign shows that the behavior of the seeds is very different to that of the subsequent generations where the observed parameters are reasonably consistent. This variability is not surprising given that the seeds were a self-selected group who had registered beforehand on the company's Web site to receive product news and promotional information and, therefore, had already manifested some interest in the category of products. (This would not be unusual for viral marketing campaigns where typically mailing lists are used to seed campaigns.) In contrast, the subsequent generations are less aligned to the promotion and likely to be more homogenous in their behavior. Surprisingly, the campaign data also shows that the seeds, while being more likely to pass the message on, on average nominate fewer contacts than those in subsequent generations.

We observe another anomaly in the empirical data in the distribution of the number of contacts provided by those who register for the campaign. As the Web interface was designed to display five textboxes at a time (with a button to request a further page for listing five more contacts), the distribution of the number of contacts is a periodic U-shaped distribution that cycles on multiples of five. Hence, the Web interface in the General Motors campaign discouraged users from forwarding a message to a large number of people. Therefore, forwarding to 5 people was easy but forwarding to 500 would have been an extremely long and tedious operation.

Again, the seeds have a different distribution (mean and shape), reinforcing the argument presented above that the seeds behave differently to the latter generations. Closer examination of the underlying distribution for the number of contacts across the campaign appears to provide a reasonable fit to a Power-law distribution, signifying that the associated digital network will display scale freeness. However, analysis of the data shows that there are a very small number of hubs in total and, given the decay in the reach, there are very few hubs beyond the first two generations. In light of the above discussion, the activated digital network of the campaign appears to exhibit mixed characteristics (random, small world, scale free) which perhaps on reflection is not unexpected.

4.3. Simulation of the Campaign

Given that the empirical data displays some evidence of social structure (both small world and scale-free characteristics) in order to determine the model of best fit, we consider all three network models discussed in the previous section. We develop a computer simulation that enables us to replicate this digital marketing campaign within each of the social network models. The computer program simulates the following processes (illustrated in Figure 2):

1. The creation of the underlying digital network based on the population size, the network model, and connectivity parameters (stochastic) used;

2. The seeding of the campaign with the message based on campaign strategy used; and

3. The transmission of the message through the digital network, generation by generation, as potential communication links are activated based on the probability of transmission and the distribution of the number of links activated (both stochastic).

**Figure 2   The Processes Used in Simulating a Viral Marketing Campaign**
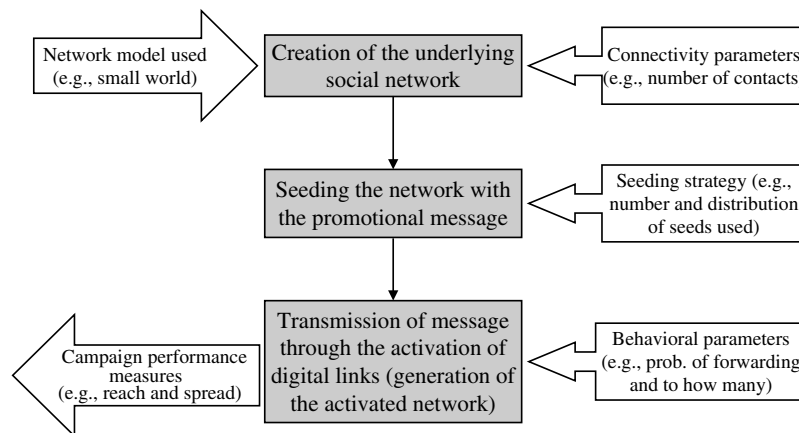
When running test simulations, we encountered a significant number of redundant communications where some people receive the message from more than one source, resulting in early termination of the simulated campaign. Unfortunately, as redundant e-mails were not recorded in the actual campaign, we are not able to validate this fact. However, we were able to confirm mathematically that it would not have been possible for a viral communication to spread through a population of the magnitude 171,000 without having generated a significant number of redundancies. As mentioned in §4.1, it is unrealistic to assume that the campaign remained within the bounds of the target population. Hence, it is likely that the actual digital network through which this promotional message was transmitted was much larger. In the following simulations, we use a larger network of size one million in order to represent a more realistic (larger) target audience, generate a simulation corresponding to the length of the actual campaign, and maintain a low proportion of simulated redundancies (5%). A reason for limiting it to one million was a pragmatic consideration, taking into account computer memory and time limitations.

The simulation model uses weighted averages of the observed transmission parameters, the estimated target audience, and number of seeds used in the campaign within each of three network structures. The simulation model is validated using the empirical data.

The initial results (base case) generated by the computer simulation model for each of the network

models versus observed campaign data are shown in Figure 3. The variance of the simulation outputs is low for all the networks (coefficient of variation less than 5%), particularly in the early stages of the campaign. These initial simulations also show that the scale-free network produces significantly different results to those observed in the actual campaign, particularly with respect to a higher growth in generation 2. As discussed in §4.2, the small number of hubs significantly diminishes the scale freeness of the activated digital network, and this would explain the lack of fit with a simulated scale-free model. In the simulation, hubs that were not seeded are much more likely to be connected to the seeds (given their high connectivity). This results in a high proportion of hubs in the simulated first generation, which in turn creates a surge in growth in the next generation.

In contrast to the scale-free network, the simulated campaigns using the random and small world networks have a good likeness to the actual campaign,

Figure 3   Results from the Base Case Simulation for Each Network Structure Compared to the Actual Data
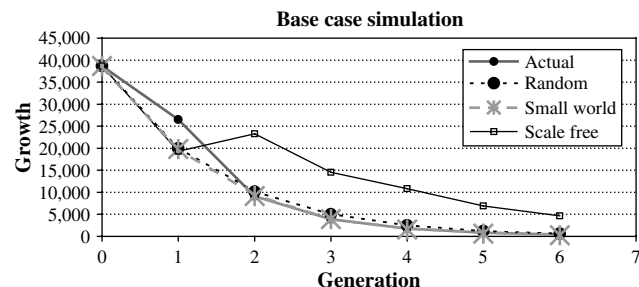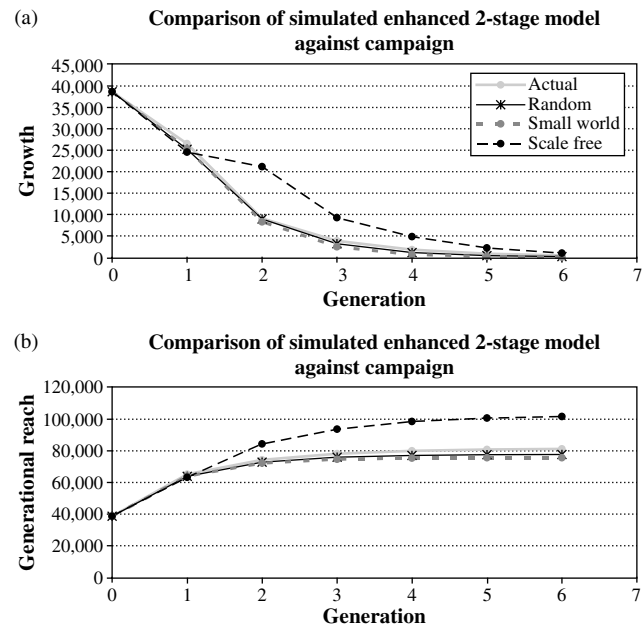
but these results differ from the outputs observed in the actual campaign in two aspects: first, there is a significant difference in the number of contacts the message was sent to by the seeds (the reach at generation 1), where the real campaign registered a higher value than the simulated cases; and second, there is a difference in the eventual reach, where the simulated campaigns' performances are not as good as the actual campaign.

4.4. Model Enhancements to Improve the Fit to Actual Campaign Results

A limitation of the base case model is that it does not take into account the different transmission behavior of the seeds, as discussed in §4.2 and as is also evident from the results shown in Figure 3. In recognition of the different behavior of the seeds, we modify our model by dividing the campaign into two "stages," each with its own transmission parameters (different forwarding probability and activation parameter but same network structure). The first stage consists of just the seeds and the contacts they send to (first generation), and the second stage starts with the first generation and extends to the end of the campaign. We estimate the transmission parameters for each stage from the empirical data (shown in Table 2).

As shown in Figures 4(a) and 4(b), the enhanced (2-stage) simulation models produce a very good match with the outputs from the actual campaign, particularly in the early generations. The simulated model that best fits the campaign data is the random network (lowest mean square error). However, the small world network also provides a very good fit to the results of the General Motors campaign. This is not surprising as the rewiring probability used for the small world network estimated from the campaign data is relatively high and, therefore, this network displays similar characteristics to a random network. Again, contrary to earlier expectations, the scale-free network produces the least fit. Further analysis of the

Table 2   Estimated Transmission Parameters

	Est. p_F	Est. $p_A\lambda$	Est. μ
Whole campaign	0.192	2.82	0.527
Stage 1 (generation 0–1)	0.265	2.59	0.687
Stage 2 (generation 1–13)	0.126	3.02	0.382

Figure 4   Generational Growth and Reach for the Enhanced 2-Stage Model (Random, Small World, and Scale Free)

(a) Comparison of simulated enhanced 2-stage model against campaign

(b) Comparison of simulated enhanced 2-stage model against campaign

Notes. Parameter values: $N = 1{,}000{,}000$, $i = 38{,}661$, $r = 0.1$, p_F (seeds) = 0.265, p_F (nonseeds) = 0.126, $p_A = 0.1$; mean number of people sent to by seeds = 2.59; mean number of people sent to by nonseeds = 3.02.

simulated scale-free network shows that it has a disproportionately high number of large hubs in comparison to the actual campaign. These large hubs are responsible for creating a surge in the reach midway through the campaign, which results in higher overall reach. In contrast, the reach within the random and small world networks falls short of the actual campaign (the eventual reach of the random model is within 5% and the small world model is within 7.5%). These results would indicate that the activated digital network of the General Motors Holden viral marketing campaign is best captured by a random network. This is not unexpected given that the earlier analysis of the structure of the network suggested mixed characteristics.

The simulation output of the enhanced model also has low variance (coefficient of variance less than 5%) in the early stages of the campaign and becomes a bit more variable in the latter generations (coefficient of variance up to 15%), as expected. In general, it is expected that the more complex networks (with more parameters) would generate more variable simulated output, and the results from this study are consistent

with that premise. With the small world network, the strong clustering tendency makes simulation results especially sensitive to the progress in the earlier generations of the campaign, as the number of random links that are activated by the early generations can influence how quickly the viral message spreads in the system. In scale-free networks, the inclusion or exclusion of hubs at each generation can make a significant difference in the simulated growth.

5. Sensitivity Analysis

The input parameters a campaign manager can influence are: (i) the number of seeds used; (ii) the probability of forwarding a message; and (iii) the average number of activated contacts (from the total set of contacts). We vary each of these parameters one at a time (while keeping all other parameters constant and in line with the actual campaign), and simulate the spread of the campaign within each theoretical network model. We vary the parameters within a range on either side of the observed value from the actual campaign.

5.1. Varying the Number of Seeds

Figure 5(a) shows the sensitivity of the reach (at each generation) to varying the number of seeds used from 10,000 to 50,000 to 100,000 (with a target population of 1 million) within all network models. In general, the change in the average eventual reach is proportional to the change in the number of seeds used, suggesting the relationship between the number of seeds and reach is approximately linear. At first glance, this may appear counterintuitive given the potentially exponential growth of viral propagation. However, with campaigns exhibiting subcritical growth rate of transmission (as in the General Motors campaign), this underscores the need to maximize initial seeding. The relative ranking of the three network models is consistent with what was observed in earlier simulations, with the curve for the scale-free network positioned clearly above the other two network models. This may be explained by the role of hubs created by the preferential attachment. When the number of seeds used is low, there is very little separation between the average reach achieved within the random and small world networks. The reasons for this are twofold: first, as discussed earlier, the relatively high value

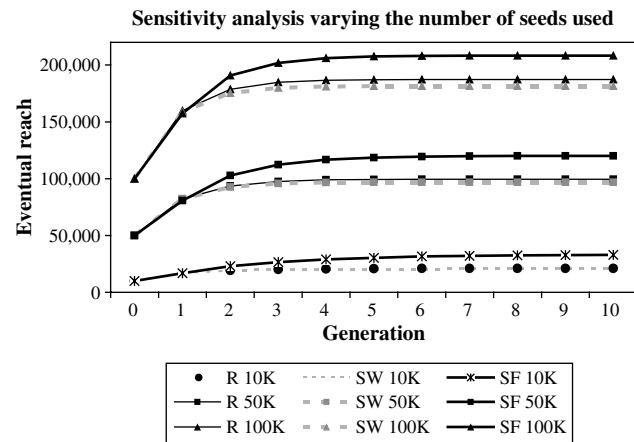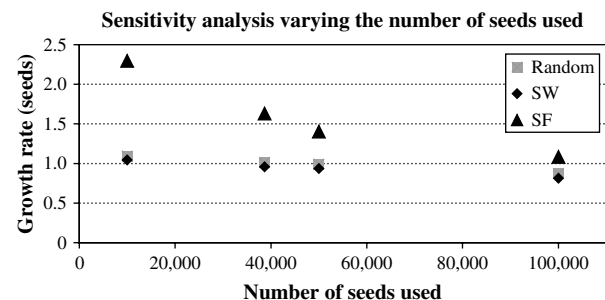**Figure 5(a)   Generational Reach with Varying Numbers of Seeds (All Networks)**

Sensitivity analysis varying the number of seeds used

Figure 5(b)   Growth Rate with Varying Numbers of Seeds (Random, Small World, Scale Free)

Sensitivity analysis varying the number of seeds used

Notes. Parameter values: $N = 1{,}000{,}000$, $r = 0.1$, p_F (seeds) $= 0.265$, p_F (nonseeds) $= 0.126$, $p_A = 0.1$; mean number of people sent to by seeds $= 2.59$; mean number of people sent to by nonseeds $= 3.02$.

of the rewiring probability has resulted in the small world network tending to mimic a random network; and second, combined with the relatively low values used for the transmission parameters (forwarding probability and number sent to), the differences in the network structure play a less significant role.

Figure 5(b) shows the average growth rate of the infectives as the number of seeds is increased within each network model. For example, when 500 seeds are used in a scale-free network model, the growth rate starts at 2.9 (500 seeds pass the message onto a further 1,452 people, achieving a total eventual reach of 1,952 for the campaign). Hence, as expected, as the number of seeds used increases, the impact of each seed (on the eventual reach) decreases. Further, the simulated campaigns show that when the number of

seeds used is high (10% of the total population), there is little difference between the network models as the message spreads very quickly (and easily) at the early stages of the campaign and the intricacies of the network structure have less opportunity to play a role.

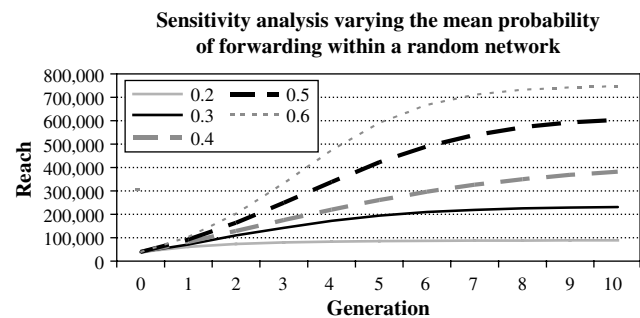## 5.2. Varying the Probability of Forwarding the Message

Figures 6(a)–6(c) show the sensitivity of the eventual reach to variation in the mean probability of forwarding within different network models. This parameter has the effect of a "switch," where an increased likelihood of forwarding results in all the activated contacts being added versus the incremental increase resulting from changing the average number of contacts. Hence, the impact of an increase in this parameter is significant. As shown in the simulated results, an increase in the mean probability from 0.5 to 0.6 results in an increase in reach of over 10,000. Comparison of Figures 6(a)–6(c) also shows that the increase in the probability of forwarding has much more of an impact at the earlier generations of the campaign in a scale-free network.

Figure 7 illustrates the change in the simulated eventual reach across all networks as the probability of forwarding is varied. This analysis shows that when the probability of forwarding is low, the scale-free network produces the best reach, but when the probability of forwarding is high, this network structure does not perform as well as the other two networks. This is likely to be caused by the higher number of isolated nodes in a scale-free network, where the increased probability of forwarding can not counteract the lack of links. Up to generation 5, the networks show similar sensitivity to forwarding probabilities (see Figures 6(a)–6(c)), but thereafter the impact of the isolated nodes or groups of nodes shows up and the scale-free sensitivity graph flattens out. As shown in Figure 7, when the forwarding probability is high, the small world network produces the best reach.

## 5.3. Varying the Number of Activated Links

Figures 8(a) and 8(b) show the effect of a variation in the number of activated links. As expected, an increased number of links leads to higher reach and once again there is a marked difference when

Figure 6(a) **Generational Reach with Varying Mean Probability of Forwarding (Random Network)**

Notes. Parameter values: $N = 1,000,000$, $i = 38,661$, $p_A = 0.1$; mean number of people sent to by seeds $= 2.59$; mean number of people sent to by nonseeds $= 3.02$.

Figure 6(b) **Generational Reach with Varying Mean Probability of Forwarding (Scale-Free Network)**
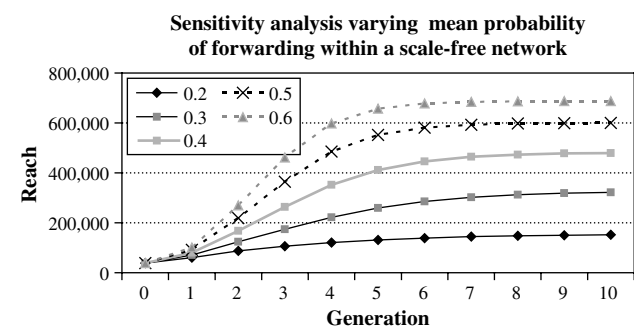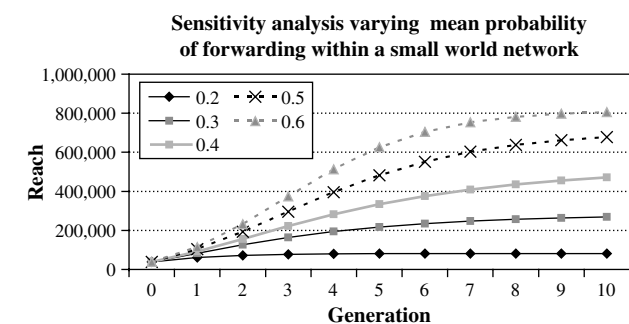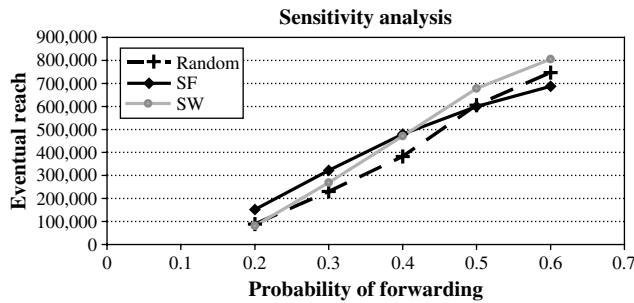

Notes. Parameter values: $N = 1,000,000$, $i = 38,661$, $p_A = 0.1$; mean number of people sent to by seeds $= 2.59$; mean number of people sent to by nonseeds $= 3.02$.

Figure 6(c) **Generational Reach with Varying Mean Probability of Forwarding (Small World Network)**

Notes. Parameter values: $N = 1,000,000$, $i = 38,661$, $r = 0.1$, $p_A = 0.1$; mean number of people sent to by seeds $= 2.59$; mean number of people sent to by nonseeds $= 3.02$.

the structure of the network is scale-free, especially at the higher end of the range of values used as shown in Figure 8(a). This can be explained by

Figure 7   Simulated Eventual Reach with Varying the Probability of Forwarding (Random, Small World, Scale-Free Networks)
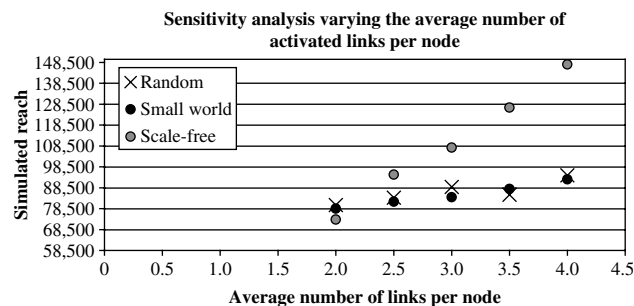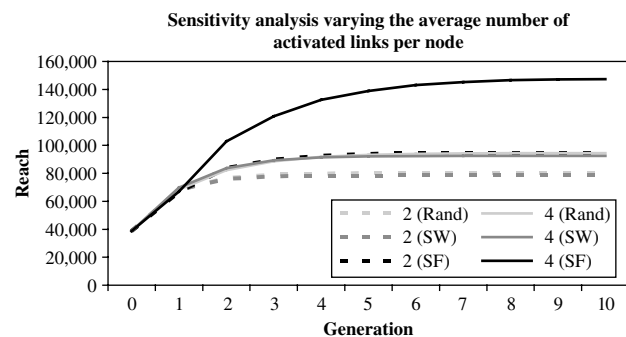

Notes. Parameter values: $N = 1{,}000{,}000$, $i = 38{,}661$, $r = 0.1$, $p_A = 0.1$; mean number of people sent to by seeds $= 2.59$; mean number of people sent to by nonseeds $= 3.02$.

the large hubs existing within a simulated scale-free network. This latter point is illustrated more clearly in Figure 8(b), where the reach is presented

Figure 8(a)   Simulated Eventual Reach with Varying Numbers of Activated Links Per Node (Random, Small World, Scale-Free Networks)

Notes. Parameter values: $N = 1{,}000{,}000$, $i = 38{,}661$, $r = 0.1$, $p_A = 0.1$, p_F (seeds) $= 0.265$, p_F (nonseeds) $= 0.126$.

Figure 8(b)   Generational Reach with Varying Numbers of Activated Links Per Node (Random, Small World, Scale-Free Networks)

Notes. Parameter values: $N = 1{,}000{,}000$, $i = 38{,}661$, $r = 0.1$, $p_A = 0.1$, p_F (seeds) $= 0.265$, p_F (nonseeds) $= 0.126$.

generation by generation. This set of results also shows that an average of four activated links per node are needed within random and small world networks to achieve the reach achieved with an average of two activated links within a scale-free network.

6.   Theory Building and Managerial Implications

6.1.   Implications for Campaigns in General

Campaign managers need guidance in adapting their marketing strategies in online contexts (Dellarocas 2003). They generally have no way of knowing what kind of social or digital network structures they are working with. Our findings show that social network structures have a significant impact on campaign performance. In particular, we show that scale-free networks are very efficient for viral campaigns and thus encourage campaign managers to try to capture scale-free properties in their target audience—possibly through identifying and seeding influential customers who might then function as hubs. Further, we detect little differences between small world and random networks (even with rewiring parameter $r = 0.1$). It appears that clustered networks are not particularly efficient. Small world networks present a more difficult scenario for the campaign manager because high clustering generally tempers the spread of the message. Future research should consider mechanisms for managing areas with poor spread by reinforcement seeding.

Building on Godes and Mayzlin (2004), we find that in general the reach is proportional to the number of seeds used. When using a high initial number of seeds, the structure of the digital network is less important (see Figure 5(b)), but at lower levels of initial seeding, the network structure has a marked impact (see Figure 5(a)). We find that an increase of one activated contact per person has an appreciable impact on the campaign—this is especially so with scale-free network (see Figures 8(a) and 8(b)). Our sensitivity analysis shows that the reach is quite sensitive to changes in the number of activated contacts. This is particularly true with scale-free networks. In fact, we find that if the empirical campaign managed to increase the average number of activated contacts per person

by one, from the observed value of 2.8, the related incremental increase in reach would be over 30,000. In contrast, given the low growth rate of seeds in this campaign, to achieve the same increase in reach, the company would have needed to seed between 15,000–20,000 more people. In practical terms, the impact of a viral marketing campaign is due to messages being received from friends and acquaintances and not from mass marketing. The initial seeds are not the target of a viral campaign—their buying intentions may not be strongly influenced by a campaign because they have not received the message from a friend or acquaintance. In short, the target of the campaign manager is strong growth, not massive seeding. There is also a cost trade-off between the acquisition costs of additional seeds versus the costs associated with whatever incentives one embeds in the campaign. As inferred from the simulated results, an incentive that increased the mean probability of forwarding in the actual campaign from 0.5 to 0.6 would have resulted in an increase in reach of over 10,000.

The characteristics of the message and creative execution play an equally important role in determining a recipient's propensity to forward the communication, the average number of activated connections, and, hence, the average number of transmissions. For example, a humorous advertisement is likely to be transmitted in much the same manner as jokes are e-mailed within social networks. The other way to increase the success of a viral campaign is to introduce a tangible promotional incentive and link it to behaviors that increase p and λ. However, one could also argue that incentives have the potential for the campaign to extend outside the desired target market. While this is not necessarily a bad thing per se and there is no financial wastage involved, it has the potential to inflate campaign performance statistics and thus overstate the success of the campaign. There is evidence of this in the empirical data.

### 6.2. Implications for Specific Campaigns

To this point, we have simulated three different network structures and identified both random and small world networks as providing an adequate fit to the actual campaign. In addition to using the simulation to deconstruct campaigns and develop insights, this methodology holds the promise of providing predictive modelling. For example, a campaign manager

can use the first few generations of a campaign as a learning platform to decode the underlying network structure and estimate the transmission behavior of the target audience, and then use this knowledge to intervene or reshape the campaign strategy for the later generations. Alternatively, a test campaign could be run to identify the appropriate network model, calibrate its parameters and forecast actual campaign performance, and then modify campaign strategy accordingly. For example, if the campaign manager wanted to achieve a better penetration of the target audience than predicted, the strategic options available would include increasing the number of seeds used and/or modifying the reward to influence the transmission behavior.

The General Motors campaign had two distinct phases of seeding with the initial phase starting in September and the subsequent, larger phase (comprising 86% of the total number of seeds who participated in the campaign) commencing a few months later. Hence, the "test campaign" approach is one that could have been used by the campaign manager to assess the performance of the second phase on the basis of learnings from the first phase. For example, if the predicted reach was deemed too small, the number of seeds could be increased or actions could be taken to encourage more people to forward the message and/or forward the message to more people. In addition, if this approach had been used for the General Motors campaign, the analysis could have revealed the limiting impact that the Web interface appears to have had on their campaign as discussed §4.2.

Figure 9 compares the simulated second phase for the campaign based on a random network to what

Figure 9 Using the Learnings from the First Phase of the General Motors Campaign to Forecast the Second Phase of the Campaign

Learning from the early part of the campaign
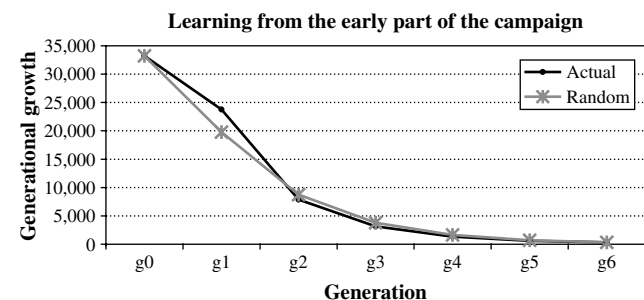
actually transpired. As shown, the fit is generally good and the total reach forecast by the simulated campaign is only 3% below the actual. The major contributor to this disparity is at the first generation. Analysis of the second phase of the campaign shows that the group of seeds that initiated this part of the campaign were even more active (higher forwarding probability and connectivity) than those in the first phase. This behavior could be explained by the fact that the deadline for the competition (and associated reward) was much closer.

7. Conclusions and Directions for Future Research

We began this study with three research objectives in mind: first, to understand the mediating effects of differing social network structures on viral marketing campaign performance; second, to develop a process for modelling viral marketing campaigns and then to validate the different models using empirical data; and third, to conduct a number of simulation experiments to predict the spread of a viral message within different kinds of social network structures under different assumptions and scenarios for the empirical campaign, and show how a campaign manager can build and apply a learning platform based upon early performance of the campaign.

In exploring the impact that social network structures have on campaign dynamics, we have provided managers with useful approaches for optimising the success of a viral campaign. Specifically, our contributions are threefold. First, we propose a conceptual framework for digital social networks that differentiates between the underlying social network and the activated digital network. Second, we illustrate the impact that network structure, connectivity, and campaign design have on campaign performance. In particular, we demonstrate the effect of varying the number of seeds, the probability of forwarding, and the number of contacts the message is forwarded to. Third, and most importantly, the models in this article provide a basis for quantifying the impact of campaign management inputs and how the analysis can be used as learnings for managerial decision making. The subtle differences between the network models also provide the essential basis for monitoring

a campaign and determining whether it is performing as expected or whether further input is needed. The marketing challenge is to achieve enough seeds and a high enough "epidemic threshold" (which is impacted through the combined effect of the activation and forwarding) to achieve campaign objectives without the unnecessary expense and possibly negative impact of flooding the target population (mass marketing). The models developed here provide a sound basis for campaign managers to meet this challenge.

These models and simulations provide the first solid method for measuring the impact of viral promotional activities on the campaign audience's behavior. With the tools introduced in this paper, it will become possible to analyse the results of campaigns and to produce a mathematically supported measure of the actual forwarding probability of audience members. This provides a basis for scientifically relating promotional activities to the audience's probabilities of forwarding and, thus, reaching a balance between promotion costs and the size of audience reached by a campaign.

In addition to our already-stated theoretical and managerial contributions, this paper also presents considerable opportunities for future research. The set of models we have chosen to test is obviously not exhaustive. It is possible, for example, to reproduce the viral campaign with a model based on a random network and susceptible-immune-susceptible (SIS) behavior (in such a model, when members receive the message, they move into a temporary state of immunity from further communications but become susceptible again at a later time). Initial investigations of a wider set of models show that modelling a campaign is no simple task. With a wider set of models, it is possible to devise quite different models which match the same set of campaign data. It will, therefore, become increasingly important to measure statistics that distinguish between different types of model, for example, between SIR and SIS behavior. Further, it is conceivable that a social network could be a hybrid of connectivity models; for example, the underlying social network structure may have random connectivity while the activated digital network may display structured (small world or scale freeness) connectivity. In this study, we used

the concept of a "generation" as a temporal measure. It would be useful to understand the relationship between time (duration) and generation, as a time-based analysis would offer further insight from a managerial perspective.

Further investigation on the mutual influence of the network and behavior models presented here is also required. Our current models do not take into account the notion of receptivity: All individuals reached by the viral message are assumed to assimilate it. Is this a valid assumption? Marketing studies have pointed out the existence of special people, called "infuentials" (Keller and Berry 2003), or "salesmen" (Gladwell 2000) who are extraordinarily effective in persuading other people to adopt an idea. How can their effect on viral process' dynamics be modelled? What precise role do they play in message diffusion? How much effort should marketers spend trying to locate them and factor them into campaign plans?

In closing, we offer three specific directions for further research. First, we see a need for more sophisticated and targeted seeding experimentation. In particular, a better understanding of the role of hubs in seeding strategies is needed, as these special individuals have considerably higher connectivity than others (Gladwell 2000, Granovetter 1983) and, through identification and targeting as campaign seeds, can be successfully exploited to increase the success of the campaign. Second, a related avenue for further enquiry is to consider the effect of possible managerial interventions during a campaign. Viral marketing has hitherto been portrayed as a random, ground-up phenomenon over which marketers have little control (Dobele et al. 2005). We disagree with this contention and believe that further empirical and experimental research with real campaigns will unearth opportunities for astute managers to proactively resurrect underperforming campaigns. Conversely, there may even be occasions where a manager needs to restrict an overly successful viral campaign, for example, one linked to an expensive sales promotion such as free samples or coupons. Finally, there is a need for further research into the aesthetic, creative, and technical components of campaign design. A viral campaign can be designed in one of two ways: Either the message is forwarded directly between audience members using e-mail, for example, or via a centralised system where an e-mail embedded link channels recipients through a Web interface. The empirical campaign discussed in this study is an example of the latter strategy. An inherent benefit of this strategy is that it enables the manager to monitor the campaign's progress and control the process' dynamics. An additional benefit of this two-stage approach is that recipients effectively self-screen, thereby reducing wastage (e.g., generating unsolicited spam). This system also enables the manager to revive a flagging campaign. A key benefit of employing a Web interface is that it can be used to produce an image of the underlying social network of the target audience. As shown in this study, the network structure plays a key role in how a campaign should be managed.

Acknowledgments

References
Albert, R., A. Barabási. 2002. Statistical mechanics of complex networks. *Rev. Modern Phys.* **74**(1) 47–97.

Ba, S., P. A. Pavlou. 2002. Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quart.* **26**(3) 243–268.

Barabási, A. R. 1999. Emergence of scaling in random networks. *Science* **286** 509–512.

Barthélémy, M., L. A. N. Amaral. 1999. Small-world networks: Evidence for a crossover picture. *Physical Rev. Lett.* **82**(15) 3180–3183.

Becker, N. 1989. *Analysis of Infectious Disease Data*. Chapman & Hall, New York.

Boguna, M., R. Pastor-Satorras, A. Vespignani. 2003. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.* **90**(2) 028701.

Bolton, G. E., E. Katok, A. Ockenfels. 2004. How effective are electronic reputation mechanisms? An experimental investigation. *Management Sci.* **50**(11) 1587–1602.

Chevalier, J. A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* **43**(3) 345–354.

Dellarocas, C. 2003. The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms. *Management Sci.* **49**(10) 1407–1424.

Dellarocas, C. 2005. Reputation mechanism design in online trading environments with pure moral hazard. *Inform. Systems Res.* **16**(2) 209–230.

Dellarocas, C. 2006. Strategic manipulation of Internet opinion forums: Implications for consumers and firms. *Management Sci.* **52**(10) 1577–1593.

Dobele, A., D. Toleman, M. Beverland. 2005. Controlled infection: Spreading the brand message through viral marketing. *Bus. Horizons* **48**(2) 143–149.

Dobele, A., A. Lindgreen, M. Beverland, J. l. Vanhamme, R. van Wijk. 2007. Why pass on viral messages? Because they connect emotionally. *Bus. Horizons* **50**(4) 291–304.

Dorogovtsev, S. N., J. Mendes. 2003. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford, UK.

Drineas, P., M. S. Krishnamoorthy, M. D. Sofka, B. Yener. 2004. Studying e-mail graphs for intelligence monitoring and analysis in the absence of semantic information. *IEEE Internat. Conf. on Intelligence and Security Informatics*. IEEE, Washington, D.C., 297–306.

Eguıluz, V. M., K. Klemm. 2002. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.* **89** 108701.

Erdös, P., A. Rényi. 1959. On random graphs. *Pub. Math. Debrecen* **6** 290–297.

Gelb, B. D., S. Sundaram. 2002. Adapting to "word of mouse." *Bus. Horizons* **45**(4) 21–25.

Gladwell, M. 2000. *The Tipping Point*. Little, Brown, and Company, Boston.

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

Goldenberg, J., B. Libai, E. Muller. 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Lett.* **12**(3) 211–223.

Goldsmith, R. E., D. Horowitz. 2006. Measuring motivations for online opinion seeking. *J. Interactive Advertising* **6**(2) 1–16.

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

Gruen, T. W., T. Osmonbekov, A. J. Czaplewski. 2006. eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *J. Bus. Res.* **59**(4) 449–456.

Helm, S. 2000. Viral marketing—Establishing customer relationships by "word-of-mouse." *Electronic Markets* **10**(3) 158–161.

Kaikati, A., J. Kaikati. 2004. Stealth marketing: How to reach consumers surreptitiously. *California Management Rev.* **46**(4) 6–22.

Keller, E., J. Berry. 2003. *The Influentials*. Free Press, New York.

Mather, D. R. 2000. A simulation model of the spread of Hepatitis C within a closed cohort. *J. Oper. Res. Soc.* **51** 656–665.

Mather, D., N. Crofts. 1999. A computer model of the spread of Hepatitis C Virus among injecting drug users. *Eur. J. Epidemiology* **15** 5–10.

Mayzlin, D. 2006. Promotional chat on the Internet. *Marketing Sci.* **25**(2) 155–163.

Miller, D. T., W. Turnbull. 1986. Expectancies and interpersonal processes. *Annual Rev. Psych.* **37** 233–256.

Moreno, Y., A. Vázquez. 2003. Disease spreading in structured scale-free networks. *Eur. Physical J. B—Condensed Matter* **31**(2) 265–271.

Pastor-Satorras, R., A. Vespignani. 2001. Epidemic spreading in scale-free networks. *Physical Rev. Lett.* **86**(14) 3200–3203.

Pavlou, P. A., D. Gefen. 2004. Building effective online marketplaces with institution-based trust. *Inform. Systems Res.* **15**(1) 37–59.

Phelps, J. E., R. Lewis, L. Mobilio, D. Perry, N. Raman. 2004. Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email. *J. Advertising Res.* **44**(4) 333–348.

Podoshen, J. S. 2006. Word of mouth, brand loyalty, acculturation and the American Jewish consumer. *J. Consumer Marketing* **23**(4/5) 266–282.

Rob, R., A. Fishman. 2005. Is bigger better? Customer base expansion through word-of-mouth reputation. *J. Political Econom.* **113**(5) 1146–1162.

Rosenthal, R. 1994. Interpersonal expectancy effects: A 30-year perspective. *Current Directions Psych. Sci.* **3**(6) 176–179.

Stewart, D., M. Ewing, D. Mather. 2004. e-Audience estimation: Modelling the spread of viral advertising using branching theory. Annual meeting, *Institute for Operations Research and the Management Sciences*, Denver, CO, 24–27.

Watts, D. J., S. H. Strogatz. 1998. Collective dynamics of "small-world" networks. *Nature* **393** 440–442.

Weinberg, B. D., L. Davis. 2005. Exploring the WOW in online-auction feedback. *J. Bus. Res.* **58**(11) 1609–1621.