A Viral Branching Model for Predicting the Spread of Electronic Word-of-Mouth

Ralf van der Lans, Gerrit van Bruggen, Jehoshua Eliashberg, Berend Wierenga

Forthcoming in *Marketing Science*

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	In a viral marketing campaign an organization develops a marketing message, and stimulates customers to forward this message to their contacts. Despite its increasing popularity, there are no models yet that help marketers to predict how many customers a viral marketing campaign will reach, and how marketers can influence this process through marketing activities. This paper develops such a model using the theory of branching processes. The proposed Viral Branching Model allows customers to participate in a viral marketing campaign by 1) opening a seeding email from the organization, 2) opening a viral email from a friend, and 3) responding to other marketing activities such as banners and offline advertising. The model parameters are estimated using individual-level data that become available in large quantities already in the early stages of viral marketing campaigns. The Viral Branching Model is applied to an actual viral marketing campaign in which over 200,000 customers participated during a six-week period. The results show that the model quickly predicts the actual reach of the campaign. In addition, the model proves to be a valuable tool to evaluate alternative what-if scenarios.
Free Keywords	branching processes, forecasting, Markov processes, online marketing, viral marketing, word of mouth
Availability	The ERIM Report Series is distributed through the following platforms: Academic Repository at Erasmus University (DEAR), DEAR ERIM Series Portal Social Science Research Network (SSRN), SSRN ERIM Series Webpage Research Papers in Economics (REPEC), REPEC ERIM Series Webpage
Classifications	The electronic versions of the papers in the ERIM report Series contain bibliographic metadata by the following classification systems: Library of Congress Classification, (LCC) LCC Webpage Journal of Economic Literature, (JEL), JEL Webpage ACM Computing Classification System CCS Webpage Inspec Classification scheme (ICS), ICS Webpage

A Viral Branching Model for Predicting the Spread of Electronic Word-of-Mouth

Ralf van der Lans
Gerrit van Bruggen
Jehoshua Eliashberg
Berend Wierenga

May 15, 2009

Forthcoming in *Marketing Science*

A Viral Branching Model for Predicting the Spread of Electronic Word-of-Mouth

Abstract

In a viral marketing campaign an organization develops a marketing message, and stimulates customers to forward this message to their contacts. Despite its increasing popularity, there are no models yet that help marketers to predict how many customers a viral marketing campaign will reach, and how marketers can influence this process through marketing activities. This paper develops such a model using the theory of branching processes. The proposed Viral Branching Model allows customers to participate in a viral marketing campaign by 1) opening a seeding email from the organization, 2) opening a viral email from a friend, and 3) responding to other marketing activities such as banners and offline advertising. The model parameters are estimated using individual-level data that become available in large quantities already in the early stages of viral marketing campaigns. The Viral Branching Model is applied to an actual viral marketing campaign in which over 200,000 customers participated during a six-week period. The results show that the model quickly predicts the actual reach of the campaign. In addition, the model proves to be a valuable tool to evaluate alternative what-if scenarios.

Keywords: Branching Processes; Forecasting; Markov Processes; Online Marketing; Viral Marketing; Word of Mouth

1. Introduction

In October 2006, Unilever launched a 75-second *viral* video film 'Dove Evolution'. This campaign generated over 2.3 million views in its first 10 days, and three times more traffic to its website than the 30-second commercial aired during the Super Bowl (van Wyck 2007). More recently, Comic Relief, a British charity organization, achieved 1.16 million participants in the first week after launching their *viral* game 'Let it Flow' that promoted Red Nose Day, their main money-raising event (New Media Age 2007). These two examples illustrate a new way of marketing communication in which organizations encourage customers to send emails to friends

1

containing a marketing message or a link to a commercial website. Because information spreads rapidly on the Internet, viral marketing campaigns have the potential to reach large numbers of customers in a short period of time. Not surprisingly many companies such as Microsoft, Philips, Sony, Ford, BMW, and Procter and Gamble have gone viral. However, not all viral marketing campaigns are successful, and due to competitive clutter, they need to become increasingly sophisticated in order to be effective and successful. It is also important that marketers are able to predict the returns on their expenditures and thus how many customers they will reach. As one marketing agency executive stated: "The move to bring a measure of predictability to the still-unpredictable world of viral marketing is being driven by clients trying to balance the risks inherent in a new marketing medium with the need to prove return on investment" (Morrissey 2007). Despite their importance, no forecasting tools for these purposes are available yet. The aim of this research is to develop a model that predicts how many customers a viral marketing campaign reaches, how this reach evolves, and how it depends on marketing activities.

The structure of this paper is as follows. Section 2 defines viral marketing campaigns and describes how marketers can influence the viral process. Section 3 shows how the flow of communication among customers in viral marketing campaigns follows a branching process, and introduces our Viral Branching Model. Section 4 describes the data of a real-life viral marketing campaign that reached over 200,000 customers after only six weeks. The predictive performance of our model, analyzed using data from this campaign, is presented in Section 5. The final Section discusses implications of our research and suggestions for further research.

2. Viral Marketing Campaigns

In a viral marketing campaign an organization develops an online marketing message and stimulates customers to forward this message to members of their social network. These contacts are subsequently motivated to forward the message to their contacts, and so on. Because

messages from friends are likely to have more impact than advertising and information spreads rapidly over the internet, viral marketing is a powerful marketing communication tool that may reach many customers in a short period of time (De Bruyn and Lilien 2008). Furthermore, the nature of the Internet allows marketers to use many different forms of communication such as videos, games, and interactive websites in their viral campaigns. The term viral marketing may (incorrectly) suggest that information spreads automatically (Watts and Peretti 2007). However, marketers need to actively manage the viral process to facilitate the spread of information (Kalyanam, McIntyre, and Masonis 2007).

2.1 Marketing activities for Managing Viral Marketing Campaigns

In viral marketing campaigns, marketers may use two types of strategies to influence the spread of information. The first focuses on motivating customers to forward marketing messages to their contacts (Chiu, Hsieh, Kao, and Lee 2007; Godes et al. 2005; Phelps, Lewis, Mobilo, Perry, and Raman 2004). As suggested by Godes et al. (2005) motivations to forward messages are either intrinsic or extrinsic. The former can be triggered by the content of the marketing message. Important components of the marketing message are the subject line of the email and the text in the email itself (Bonfrer and Drèze 2009). Furthermore, marketers nowadays develop websites containing videos and games that attract customer attention and interests. These websites usually facilitate the viral process by providing tools to easily forward emails to friends, such as 'Tell a Friend' or 'Share Video' buttons. Examples of extrinsic motivations to forward marketing messages are prizes and other monetary incentives (Biyalogorsky, Gerstner, and Libai 2001).

 Although increasing customers' motivation to forward messages to friends has a strong impact on the reach of the viral campaign, this is usually a difficult and expensive task. In contrast, controlling the number of initial or seeded customers is much more cost effective. In general, marketers can choose from three distinct categories to seed their viral marketing

3

campaign: 1) seeding emails, 2) online advertising, and 3) offline advertising. Seeding emails are usually sent by the company itself or by a specialized marketing agency to customers who have given permission to receive promotional emails (Bonfrer and Drèze 2009). Using this seeding tool, a marketer can target a specific group of customers that are potentially interested in the campaign. The design and content of the emails are crucial since customers easily categorize such emails as spam and quickly delete them. For this reason, seeding emails are expected to be less effective than viral emails that are sent by friends or acquaintances of the recipient.

Online advertising is another important seeding tool that marketers can use to influence the viral process. The effectiveness of online advertising may differ depending on the customers as well as the websites on which the ads are placed. Interestingly, marketers can directly observe when a specific online ad generates a visitor to the viral campaign. Hence, the effectiveness of online advertising can be monitored accurately, and based on its performance marketers can decide to adapt their online advertising strategy. Furthermore, online advertising agencies offer contracts that guarantee a predetermined number of clicks to the campaign website within a certain time window. In such cases organizations usually pay for each click. Because online ads may be perceived as less obtrusive than promotional emails, this seeding tool may be very attractive.

Finally, besides online seeding tools, marketers may still use 'traditional' offline advertising to seed their campaigns. Examples are magazine or TV ads that refer to the website of the viral marketing campaign, and package labels or coupons that try to attract visitors to the campaign website. However, offline seeding is less popular and expected to be less effective because customers cannot directly visit the campaign website by clicking a link. Another disadvantage of offline seeding is that it is more difficult to measure its effectiveness, as marketers cannot directly observe when offline advertising generates a customer to the viral campaign. Possible

solutions for this problem are asking customers on the website how they were informed, or to ask for the barcode of the product or coupon that was used to enter the website.

As described above, the appropriate strategic decision of the marketing activities at the right moment strongly depends on the spread of the process and the effectiveness of each marketing communication tool. Therefore, marketers need to closely monitor the spread of information in viral marketing campaigns.

2.2 Monitoring Viral Marketing Campaigns

An important feature of viral marketing campaigns is that marketers are able to accurately measure the actions of customers, such as when they open an email (Bonfrer and Drèze 2009), and which pages they visit (Moe 2003). Hence, marketers may obtain large databases containing detailed customer behavior. Monitoring such behavior is not straightforward, and it is therefore important to retain only those variables that are relevant to the viral process.

Figure 1 summarizes the five-stage process that a customer may go through during a viral marketing campaign. In the first stage, a customer receives an invitation at time t_1 from source b, i.e., through a viral email from a friend or through one of the seeding tools of a company. At the end of this stage, the customer decides with probability ϖ_{12}^{b} to go to the second stage and read the invitation at time t_2, or with probability $1 - \varpi_{12}^{b}$ to exit the campaign by deleting or ignoring the invitation. This probability ϖ_{12}^{b} is likely to depend on the source of invitation b, as customers are less likely to open and read a seeding email from a company than a viral email from a friend. After reading the invitation to the viral campaign, a customer decides to accept the invitation with probability ϖ_{23}^{b} by clicking a link to the landing page of the campaign website. After arriving on the landing page at time t_3 (stage 3), a customer decides to participate in the viral

Figure 1: Decision Tree to Participate in Viral Marketing Campaign
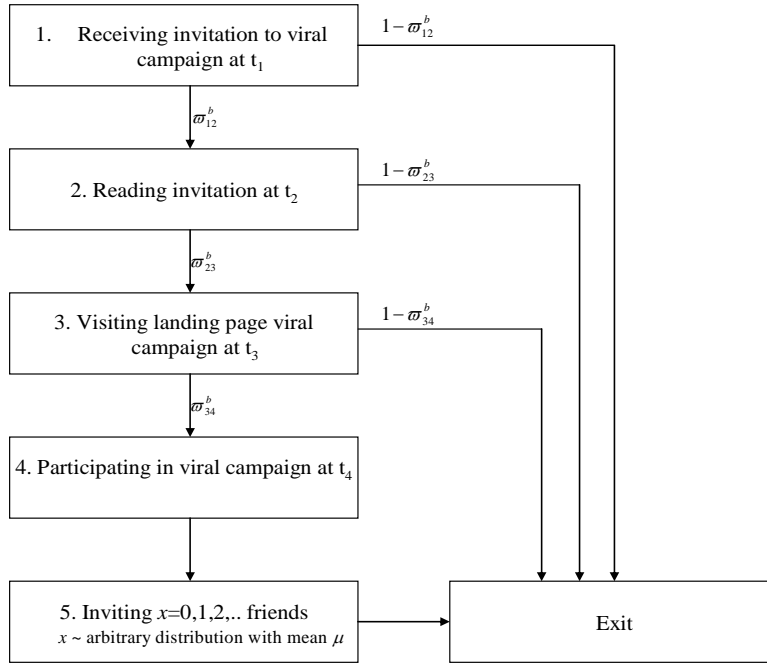
campaign (stage 4) with probability ϖ_{34}^b at time t_4. Participation may consist of watching a video, playing a game, and/or subscribing to a service. Finally, a customer decides to forward the message to x friends.

Figure 1 indicates that the number of customers receiving an email is not necessarily the same as the number of customers who ultimately participate in the viral campaign, as this depends on the probabilities ϖ_{12}^b, ϖ_{23}^b, and ϖ_{34}^b. As described in the previous Section, these probabilities depend on marketing activities such as the attractiveness of the subject line (ϖ_{12}^b), the content of the invitation (ϖ_{23}^b), and the design and content of the website (ϖ_{34}^b). Although the sequence of stages is quite generic for most viral marketing campaigns (De Bruyn and Lilien 2008), we recognize that it does not necessarily hold for all viral marketing campaigns. For instance, participation may consist of several stages (activities) such as watching a video, subscribing to a newsletter, and/or playing a game. In addition, it is possible that customers forward the message before participation, i.e. in cases where customers can only participate when they invite a certain

number of friends. Therefore, marketers should adapt Figure 1 depending on the specific structure of their campaign. For the campaign of interest in our empirical application, Figure 1 accurately matches its structure. However, the agency executing our campaign did not store data for stages 2 and 3. Hence, for each participant we observed the transition from stage 1 to 4, which occurred with probability $\varpi_{12}^b \varpi_{23}^b \varpi_{34}^b$. Adapting our model (Section 3) to an alternative structure of a viral marketing campaign is straightforward.

In order to manage viral marketing campaigns, marketers need to monitor the stages represented in Figure 1 for each individual customer. Specifically, they should register the following variables: 1) the source of the invitation, 2) if and when a customer arrives at each stage, and 3) how many friends a customer invites. This leads to a dynamic database in which each row represents a customer and in which corresponding variables are updated when a customer switches to the next stage. New rows are added when new customers are invited. Such a database can be automatically generated in real time during the process of a viral marketing campaign.

In summary, viral marketing is an effective online marketing communication tool that may reach many customers in a short period of time. The reach of a viral marketing campaign is a function of seeding activities and the number of forwarded viral emails. While the seeding activities are under the direct control of marketers, they can only influence the number of forwarded emails through incentives. To reach the campaign's goals, it is important for marketers to be able to forecast the reach of a viral marketing campaign as early as possible, and to determine how this reach depends on marketing activities. Because tools for supporting these forecasts do not yet exist, we have developed such a forecasting model in the next Section.

3. Modeling the Viral Marketing Process

Insights from epidemics about the spread of viruses are useful to understand and model the spread of marketing messages in viral marketing campaigns. In epidemics, both aggregate and disaggregate level models have been developed to describe the spread of viruses (Bartlett 1960). Aggregate level or diffusion models assume an underlying infection process, and the corresponding model parameters are inferred from the total number of infected individuals over time. Based on these insights, Bass (1969) developed his famous diffusion model and assumed adoption to depend on two forces: one that is independent of previous adoptions and one that depends positively on previous adoptions. As the number of customers in viral marketing campaigns (i.e. adopters) is also influenced by these two forces, the Bass model should be able to describe the spread of information during viral marketing campaigns. However, there are two important reasons why the Bass model does not optimally describe the viral marketing process. First, it assumes a specific process, but does not include actual information on this process at the individual level. Such information becomes readily available in viral marketing campaigns and can be used to describe the process accurately at the aggregate level. Second, the Bass model assumes that every customer who has adopted the product increases the probability of others adopting in each time period after adoption. However, in viral marketing campaigns customers only influence each other right after participation when they invite their friends.

Disaggregate level or branching process models (Athreya and Ney 1972; Dorman, Sinsheimer, and Lange 2004; Harris 1963) may alleviate these two limitations as parameters are estimated based on individual-level information, and they assume that customers only influence each other right after participation by infecting a fixed number of others. Although branching process models have proven to be very useful in describing the spread of viruses theoretically, they have so far, to our knowledge, not been applied to real empirical process data. The reason

for this is that, similar to the diffusion of products, the process of the actual spread of viruses is typically not observed. Interestingly, in viral marketing campaigns, marketers can observe the actual spread of information across customers, and branching processes might therefore be a promising tool to describe and predict the reach of these campaigns. Furthermore, since standard branching models and their extensions are not capable of describing viral marketing campaigns, another contribution of our research is to extend the standard branching model. In order to do so, we now first explain the standard branching process.

3.1 Viral Marketing as a Branching Process

Branching or Galton-Watson processes were originally developed at the end of the 19th century to derive the probability of extinction of families (Athreya and Ney 1972; Dorman et al. 2004; Harris 1963). Generalizations of these processes, of which the birth-and-death process is an example, have been applied to model phenomena in physics, biology, and in epidemiology to describe the spread of viruses in populations. Figure 2 graphically demonstrates the spread of information according to a standard branching process. The process represents T generations of customers that all invite $x = 2$ other customers. In the branching literature, x is crucial and has an arbitrary probability distribution with mean μ, which is called the infection or reproduction rate of the process. In Figure 2, the first generation (represented by stars) consists of an initial seed of n 'infected' customers that forward the message to a second generation of customers that subsequently forward the message to a third generation, etc. Therefore, the total number of customers $V(t)$ in generation t equals nx^{t-1} and the total reach of the campaign at generation T equals $n\sum_{t=1}^{T} x^{t-1}$. In situations where the infection rate is greater than 1, it is sufficient for marketers to seed only a few initial customers to start the viral process, after which the whole

population will ultimately be infected. However, unlike in an epidemic, the infection rate in viral marketing campaigns is generally smaller than 1 (Watts and Peretti 2007), which means that the spread of information dies out quickly as each customer generates on average less than one new customer. In such situations, marketers should influence the viral process by: 1) increasing the campaign's infection rate μ, or 2) increasing the number of seeded customers *n*.

Although the standard branching model is useful to understand the underlying process in viral marketing campaigns, a more detailed model is needed to accurately describe and predict the actual spread of information. Therefore, we have extended this standard model as follows. First, while the standard branching model is a Markov process with fixed transition times, we allow customers to participate at any moment in time leading to a Markov process with stochastic transition times. Second, we incorporate two different types of marketing seeding activities; the first type allows seeding via sources *Q* such as banners and traditional advertising, while the

Figure 2: Spread of a Message in a Viral Marketing Campaign as a Branching Process
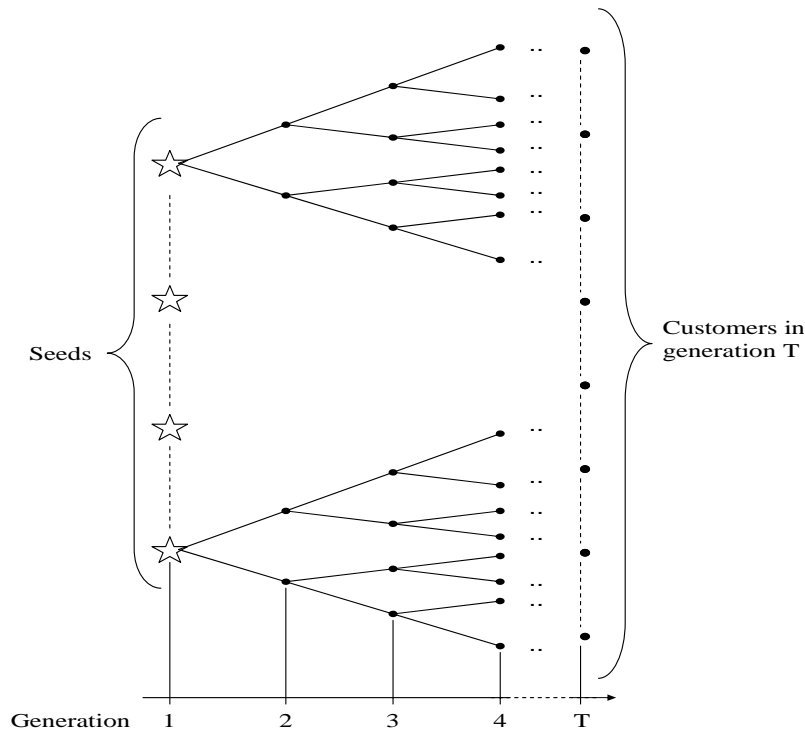
second type allows seeding through emails. To incorporate this second type, we add the dimension $M(t)$ to the standard branching process, which represents the number of unopened seeding emails at time t. Third, while branching models typically count the number of 'infected' customers $V(t)$ (i.e. customers who received emails and did not participate or delete the email yet), we also count the cumulative number of customers who actually participate by introducing a third dimension $N(t)$ to the branching model. Fourth, standard branching processes assume parameters to be constant over time. However, it is likely that new invitations become less effective during the course of the campaign, because invitations may be sent to customers who already received one or already participated in the campaign. Interestingly, invitations by seeding activities are less likely to be affected by this, because companies observe participants and invitations in real time during viral marketing campaigns. Hence, if a company carefully selects email addresses, seeding emails should be sent to customers that did not receive an invitation yet. Furthermore, as discussed in Section 2.1, online marketing agencies frequently offer banner contracts generating a pre-specified number of clicks. Also, these clicks are likely to come from new customers that did not participate yet. However, the probability that a participant invites a friend who already received an invitation or already participated increases as a function of the number of participants and sent invitations. In this research, we explicitly model this dynamic phenomenon, by allowing μ to decrease as a function of $N(t)$ and already invited customers. Next, we describe how the three processes $M(t)$, $V(t)$, and $N(t)$ interact in our Viral Branching Model.

3.2 The Viral Branching Model

In this study, we decided, without loss of generality, to count those customers who participated in the viral campaign as the reach metric (Stage 4 in Figure 1). Before introducing our model, we

present its notations. Let:

$t \in [0,..,T]$ denote continuous time, with $t = 0$ the start and $t = T$ the end of the campaign;

$N(t)$ denote the cumulative number of participants in the viral campaign at time t;

$V(t)$ denote the number of customers who received a viral email from a friend and who did not participate or delete this email yet;

$M(t)$ denote the number of customers who received a seeding email from an organization and who did not participate or delete this email yet;

$Z(t)$ denote the vector $\{M(t), V(t), N(t)\}$;

$q \in Q$ denote the set of seeding sources excluding seeding emails (i.e. banners, advertising);

b denote the index over all invitation sources, i.e. $b \in \{\text{viral mail, seeding mail}, Q\}$;

μ^* denote the average number of invited contacts, given participation;

θ denote the average proportion of invited contacts that have already been invited or already participated in the campaign;

μ denote the average number of invited contacts who have not been invited or participated in the campaign, given participation, hence $\mu = \mu^* \cdot (1-\theta)$ [1];

π_b denote the probability of participation upon receiving an invitation by source b (i.e. $\pi_b = \varpi_{12}^b \varpi_{23}^b \varpi_{34}^{b}{}^2$);

$1/\lambda_v$ denote the average time between receiving a viral email and participating;

$1/\lambda_m$ denote the average time between receiving a seeding email and participating;

β_q denote the rate with which customers are invited by seeding tool q.

Figure 3 summarizes our Viral Branching Model and shows how $Z(t)$ changes over time. It

shows how customers are invited to participate in the viral campaign by 1) receiving a seeding

---

[1] Without loss of generality, in the derivations of the viral branching model in paragraphs 3.2 and 3.3, we express the processes $Z(t)$ as a function of μ. In Section 3.4 we show how μ^* and θ are incorporated.

[2] To count the number of customers in another stage of Figure 1, it is sufficient to change the definition of π_b and μ. For instance, to count the number of participants in stage 2, π_b becomes equal to ϖ_{12}^b, and μ needs to be multiplied by $\varpi_{23}^b \varpi_{34}^b$.

email from a company, 2) another seeding source $q \in Q$ such as a banner or traditional

advertising, or 3) receiving a viral email from a friend. When a customer participates in the viral

campaign at time t, the number of participants $N(t)$ increases by 1 and $M(t)$ or $V(t)$

decreases by 1 if this participant was invited by a seeding or viral email respectively.

Furthermore, customers may invite y friends, of which w friends are already invited to or already

participated in the viral campaign. Hence, the number of customers that has an invitation by viral

email increases by $x = y - w$. Because each participant may decide to invite a different number

of friends, we assume that $y \in \{0, 1, 2, 3, ...\}$ comes from an arbitrary distribution with mean $\mu*$.

Furthermore, we assume that $w \in \{0, 1, 2, .., y\}$ is an arbitrarily distributed proportion θ of y.

Hence, x comes from an arbitrary distribution with mean $\mu = \mu*(1-\theta)$. As shown in Figure 3,

every time t a customer decides to participate, the process variables $M(t)$, $V(t)$, and $N(t)$

change to new values. These process variables only depend on the parameters β_q, π_b,

$\mu = \mu*(1-\theta)$. Finally, to incorporate the speed at which people open viral and seeding emails,

we assume the time between receiving an invitation and participation to be exponentially

distributed with means $1/\lambda_v$ and $1/\lambda_m$ for viral and seeding emails respectively. Although other

distributions may fit better, the exponential distribution for the time between receiving an email

and participation is a reasonable approximation (Bonfrer and Drèze 2009). In addition, the

exponential distribution is the only distribution that leads to mathematically tractable solutions

(Dorman et al. 2004).

   Based on the flow diagram in Figure 3, Figure 4 illustrates one possible realization of the

stochastic process that is generated by our Viral Branching Model. In this Figure, we assume for

simplicity that only a single customer is seeded by an email from a company to customer A at

Figure 3: Flow Diagram of a Viral Marketing Campaign
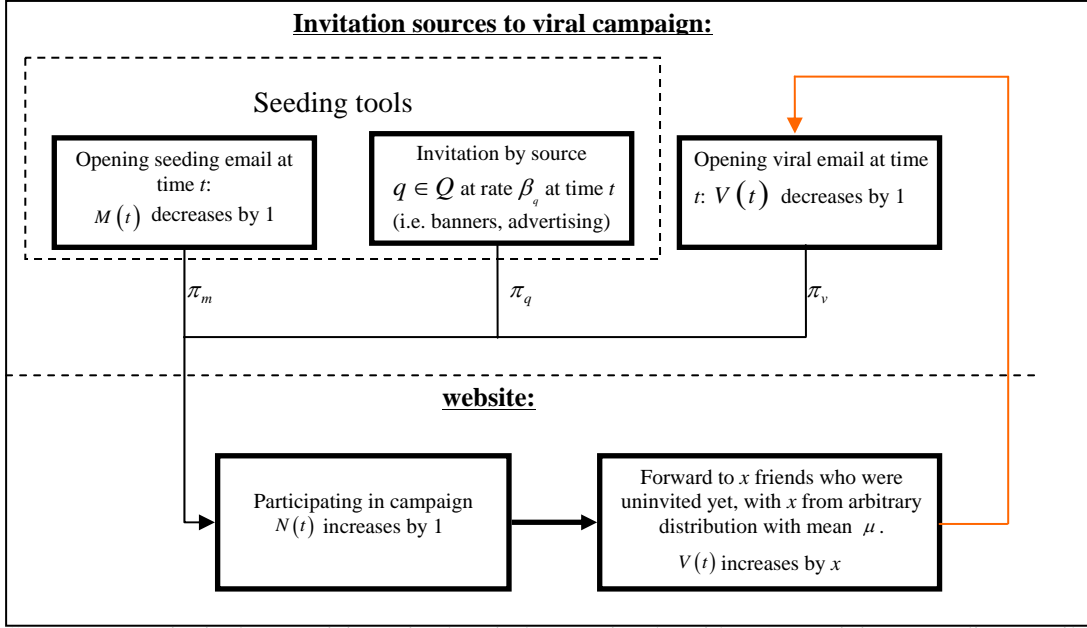
Note: Customers are invited to participate in the viral campaign by either 1) receiving a seeding email from the company, 2) via another seeding source q such as a banner or advertising, or 3) receiving a viral email from a friend. A customer participates with probability π_b, depending on the source b of the invitation. If the customer decides to participate in the viral campaign, $N(t)$ increases by 1. After participation, the customer invites x friends who did not receive an invitation or participate yet, where x is generated from an arbitrary chosen distribution with mean μ. These x invited customers become members of $V(t)$, hence $V(t)$ increases by x.

time t_0. Therefore $M(t_0)$, indicating the number of unopened seeding emails at t_0 sent by a company, equals 1. After $t_1 - t_0$ time units, which is assumed to have an exponential distribution with mean $1/\lambda_m$, customer A opens the email message and participates in the viral campaign, for example, by clicking a link directed to the campaign website. Consequently, $M(t_1) = M(t_0) - 1 = 0$, and $N(t_1)$, indicating the reach of the viral marketing campaign up to time t_1, equals $N(t_1) = N(t_0) + 1 = 1$. After participation, customer A sends two emails to friends B and C via the 'invite a friend' button. For that reason $V(t_1)$, representing the number of customers with an unopened viral email in their mailbox, equals $V(t_1) = V(t_0) + 2 = 2$. The time that customers B and C need to open this message is assumed to be independent and identically

14

Figure 4: Realization of the Stochastic Viral Branching Process When a Company Initially Seeds One Customer (t is a continuous clock time)

	t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
Number of consumers with unopened seeding email from company $M(t)$	1	0	0	0	0	0	0	0	0
Number of consumers with unopened viral email from a friend $V(t)$	0	2	4	3	2	5	4	4	3
Total cumulative number participants $N(t)$	0	1	2	2	3	4	4	5	5
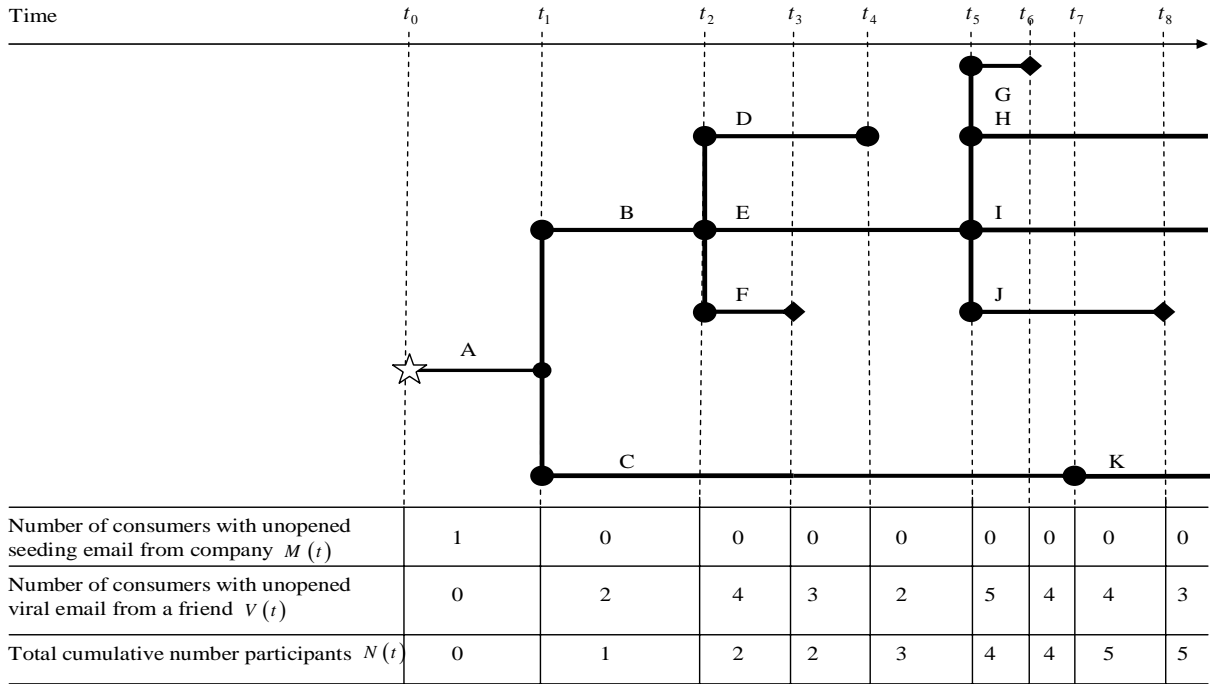
Note: At t_0 customer A is invited to the viral marketing campaign, in this case through receiving a seeding email sent by the company (☆), but this could also be due to a banner or advertising. Hence, $M(t_0)=1$. At t_1 customer A participates in the viral campaign (indicated by ●) $\left(N(t_1)=1\right)$, after opening the email $\left(M(t_1)=0\right)$, and decides to forward the message to two friends B and C $\left(V(t_1)=2\right)$. At t_2 customer B participates in the campaign $\left(N(t_2)=2\right)$, after opening the email from friend A and forwards it to three new friends: D, E, and F $\left(V(t_2)=V(t_1)-1+3=4\right)$. Subsequently at t_3, customer F opens the email and is not interested in the campaign (indicated by ◆), i.e. $V(t_3)=V(t_2)-1=3$, after which customer D opens the email $\left(V(t_4)=V(t_3)-1=2\right)$, and participates in the campaign $\left(N(t_4)=N(t_3)+1=3\right)$ but does not forward the message to friends. At t_5, customer E opens the email, starts participating in the campaign and forwards the message to four friends: G, H, I, and J, i.e. $N(t_5)=4$ and $V(t_5)=V(t_4)-1+4=5$. At t_6, customer G opens the email from friend E, but is not interested in the campaign $\left(V(t_6)=4\right)$. Then at t_7, customer C opens the email, and participates in the campaign $\left(N(t_7)=5\right)$ and forwards a message to friend K $\left(V(t_7)=V(t_7)-1+1=4\right)$. Finally, at t_8 customer J opens the email but does not participate, hence $V(t_8)=3$ and $M(t_8)$ and $X(t_8)$ do not change.

exponentially distributed with mean $1/\lambda_v$, which may be different from the time assumed for

customer A. In the example in Figure 4, customer B opens the email from friend A after t_2-t_1

time units, and customer C takes t_7-t_1 time units. Finally, in this example, at time t_8, we

observe that $M(t_8) = 0$, $V(t_8) = 3$, and $N(t_8) = 5$. In the next subsection we derive the equations of our Viral Branching Model for $M(t)$, $V(t)$, and $N(t)$.

3.3 Derivation of the Viral Branching Process Equations

Branching processes are an important class of Markov processes (Ross 1997). The memoryless property of the exponential distribution of the time between state transitions leads to a continuous time Markov process. Hence, the vector $Z(t) = (M(t), V(t), N(t))'$ follows a three-dimensional continuous time Markov process since

$$P(Z(t+t') = \mathbf{j} \mid Z(t') = \mathbf{i}, Z(r) = \mathbf{k}, 0 \leq r < t')$$ equals $P(Z(t+t') = \mathbf{j} \mid Z(t') = \mathbf{i})$. Where

$\mathbf{i} = (i_m, i_v, i_n)'$, $\mathbf{j} = (j_m, j_v, j_n)'$, and $\mathbf{k} = (k_m, k_v, k_n)'$ are nonnegative integers counting respectively the number of unopened seeding emails (indicated by subscript *m*), unopened viral emails (indicated by subscript *v*), and number of participants (indicated by subscript *n*) for different time periods: t', $t+t'$, and r respectively. In the viral marketing process without a company's interfering, the variable $M(t)$ strictly decreases and switches to state $i_m - 1$ every time a customer opens a seeding email, given that $M(t)$ was in state i_m. An important tool for a marketer to increase the value of $M(t)$ with a value *K* is by sending *K* seeding emails to a list of customers. The transitions of $V(t)$ in the viral process are more complex, as these depend on the process $M(t)$, and may both decrease as well as increase over time. When a customer opens a viral email, $V(t)$ may decrease by one if the customer does not forward the message to friends. However, $V(t)$ increases if 1) a customer opens a *seeding* email and forwards it to one or more friends, 2) a customer opens a *viral* email and forwards it to two or more friends, and 3) a

16

customer participates via another source ($q \in Q$) in the campaign and forwards it to one or more

friends. The third possibility, i.e. that customers randomly enter the viral marketing campaign

from 'outside', is an important extension of traditional branching processes and is called

immigration (Kendall 1949; Sevast'yanov 1957). We assume that the immigration rate equals

$\pi_q \beta_q$ for source $q \in \{1, 2, .., Q\}$, hence the average time between two customers that participate in

the viral campaign due to immigration is exponentially distributed with rate $1 \Big/ \sum\limits_{q=1}^{Q} \pi_q \beta_q$. Finally,

the variable $N(t)$, which depends on both processes $M(t)$ and $V(t)$, strictly increases and does

so every time a customer participates in the viral campaign. This may be due to opening an email

from a friend, or due to seeding by a company.

　　Differential equations play a crucial role in determining the values of the interrelated state

variables $Z(t)$ over time in a continuous time Markov process. Kolmogorov's backward and

forward equations are convenient to derive the differential equations that the state transition

probabilities should satisfy (Ross 1997). This research uses the forward equations to derive these

differential equations, as these are more convenient to solve compared to the backward equations

and also lead to unique solutions for all generalizations of branching processes (Harris 1963).

Because the Viral Branching Model is new to the literature, we derive and solve these

differential equations in the Web Appendix A. Next, we provide the solutions of the expectations

of $M(t)$, $V(t)$, and $N(t)$.

3.3.1 The Conditional Expected Number of Unopened Seeding Emails $M(t)$

As derived in Web Appendix A, the conditional expected number of unopened seeding emails at

time t, given that at time t', with $0 \le t' \le t$, there are i_m unopened seeding emails, equals:

$$E\left(M\left(t\right)|M\left(t'\right)=i_m\right)=i_m e^{-\lambda_m\left(t-t'\right)}. \tag{1}$$

Clearly, as λ_m is always positive, $M\left(t\right)$ decreases exponentially over time and reaches zero as

time passes. A marketer, however, may increase $M\left(t\right)$ by sending an additional set of seeding

emails to a list of customers, i.e. marketers control the value i_m directly.

3.3.2 The Conditional Expected Number of Unopened Viral Emails $V\left(t\right)$

The conditional expected number of unopened viral emails at time t, given i_v unopened viral

emails at time t', equals (see Web Appendix A):

$$E\left(V\left(t\right)|V\left(t'\right)=i_v\right)=i_v e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}+K_1\left(e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-e^{-\lambda_m\left(t-t'\right)}\right)+K_2\left(e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-1\right), \tag{2}$$

with, $K_1=\dfrac{\lambda_m\pi_m\mu i_m}{\lambda_v\left(\pi_v\mu-1\right)+\lambda_m}$, and $K_2=\dfrac{\displaystyle\sum_{q=1}^{Q}\pi_q\beta_q\mu}{\lambda_v\left(\pi_v\mu-1\right)}$. In (2), $\pi_v\mu$ represents the infection rate of

the viral marketing campaign, which is smaller than μ because not every customer who receives

an email decides to participate. Note that if $\pi_v\mu>1$, $V\left(t\right)$ grows exponentially and reaches

infinity when t becomes very large.

3.3.3 The Conditional Expected Number of Participants in the Viral Campaign $N\left(t\right)$

Web Appendix A shows that the conditional expected number of participants $N\left(t\right)$, given i_n

participants at time t', equals:

$$E\left(N\left(t\right)|N\left(t'\right)=i_n\right)=i_n+K_3\left(e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-1\right)+K_4\left(e^{-\lambda_m\left(t-t'\right)}-1\right)+K_5\left(t-t'\right), \tag{3}$$

with: $K_3=\dfrac{\pi_v}{\left(\pi_v\mu-1\right)}\left(K_1+K_2+i_v\right)$, $K_4=\dfrac{i_m\pi_m\left(\lambda_v-\lambda_m\right)}{\lambda_m+\lambda_v\left(\pi_v\mu-1\right)}$, and $K_5=-\dfrac{\displaystyle\sum_{q=1}^{Q}\pi_q\beta_q}{\pi_v\mu-1}$. Equation (3)

represents highly non-linear effects of the model parameters on the reach of the campaign $N\left(t\right)$.

Fortunately, the model parameters are estimated on the disaggregate level, and hence equation (3)

is not used in the estimation procedure. In fact, it is relatively straightforward to code this

equation in a spreadsheet program, which calculates the expected reach of the campaign based on

the individual-level parameter estimates μ, π_b, λ_m, λ_v, and β_q.

3.4 Estimating the Model Parameters

The strength of the Viral Branching Model is that its parameters can be estimated using the

individual-level data obtained from viral marketing campaigns as described in Section 2.2.

Hence, in contrast to most models in marketing, we do not estimate the model parameters using

the functional form as represented by equations (1) to (3), and data on the actual process

variables $Z(t)$. Instead, we use the dynamically generated database (see Section 2.2) containing

the individual-level data of the process from which we infer the model parameters. The estimates

based on these individual-level data are subsequently inserted into the model to predict the

number of participants over time. This approach is similar to pretest market models (Hauser and

Wisniewski 1982; Shocker and Hall 1986), including: SPRINTER (Urban 1970), PERCEPTOR

(Urban 1975), ASSESSOR (Silk and Urban 1978), TRACKER (Blattberg and Golanty 1978),

and MOVIEMOD (Eliashberg, Jonker, Sawhney, and Wierenga 2000) that predict market shares

or diffusion curves based on customers' trial and adoption processes. For these models, the

process parameters are estimated before the start of the diffusion process using data from surveys

and experiments. For our Viral Branching Model, we estimate the parameter values directly from

the individual-level data that become available from the viral process of interest and that are

stored in a dynamic database. The model parameters can be quickly estimated reliably because

this database contains many customers already in the campaign's early stages.

   We now describe how the basic parameters of the Viral Branching Model can be estimated for

a given time period. In order to do so, we first discretize the time period $[0,..,T]$ into $d = 1,..,D$

time periods, with period $d = [t_{d-1}, .., t_d]$, $t_0 = 0$ and $t_D = T$. Note that we still account for a

continuous time viral branching process, but allow the model parameters to vary across time

periods d. Hence, we estimate μ_d, π_{bd}, β_{qd} λ_{md}, and λ_{vd} for each time period d. In the

empirical application, each time period d corresponds to one day that the viral campaign is

online. For each period d, we observe $c = 1, .., n_d$ customers that participate in the viral campaign.

3.4.1 Estimating the average number of forwarded emails ($\mu = \mu * (1 - \theta)$):

Each customer c in period d forwards y_{cd} emails to friends. We introduce variable u_{cdj}, which

equals one if email $j \in \{1, .., y_{cd}\}$ forwarded by customer c in period d reaches a customer who

already participated or already received an invitation, zero otherwise. Hence, the effective

number of forwarded emails equals $x_{cd} = y_{cd} - \sum_{j=1}^{y_{cd}} u_{cdj}$. These x_{cd} emails are automatically stored

in the dynamically updated database by adding x_{cd} rows, i.e. rows $R_{c-1,d} + 1$ to $R_{c-1,d} + x_{cd}$ (see

Section 2.2). $R_{c-1,d}$ represents the number of rows in the database up to customer c-1 in period d,

which corresponds to the cumulative number of customers who already participated or were

already invited up to customer c-1 in period d-1. Given variables y_{cd} and u_{cdj}, it is relatively

easy to estimate both parameters, $\mu *$ and θ_d, as follows:

$$\mu * = \frac{1}{n_d} \sum_{d=1}^{D} \sum_{c=1}^{n_d} y_{cd} \text{ , and} \tag{4}$$

$$\theta_d = \frac{1}{n_d} \sum_{c=1}^{n_d} \frac{\sum_{j=1}^{y_{cd}} u_{cdj}}{y_{cd}} . \tag{5}$$

As described above, for prediction we expect the probability that an email is ineffective, i.e.

$P(u_{cdj} = 1)$, to increase as a function of $R_{n_{d-1}, d-1}$. We use a binary logit specification to estimate

20

this increase:

$$P\left(u_{cdj}=1\right)=\frac{\exp\left(\alpha_1+\alpha_2 R_{c-1,d}\right)}{1+\exp\left(\alpha_1+\alpha_2 R_{c-1,d}\right)}.\tag{6}$$

For prediction of $R_{n_{d'},d'}$ in period $d'>D$ after the observation period $[1,..,D]$, we use the

following equation:

$$R_{n_{d'},d'}=n_{d'}\mu_{d'}+\sum_{q=1}^{Q}\pi_{qd'}\beta_{qd'}\cdot\left(t_{d'}-t_{d'-1}\right)+K_{d'},\tag{7}$$

where $n_{d'}\mu_{d'}=\left(N\left(t_{d'}\right)-N\left(t_{d'-1}\right)\right)\mu_{d'}$ represents the expected number of forwarded emails in

period d', $\sum_{q=1}^{Q}\pi_{qd'}\beta_{qd'}\cdot\left(t_{d'}-t_{d'-1}\right)$ the expected number of customers who join the campaign due

to seeding activities $q\in Q$, and $K_{d'}$ represents the number of seeding emails that a company

sends in period d'. Given the predicted value of $R_{n_{d'},d'}$, we use (6) to predict $\theta_{d'+1}$ as

$P\left(u_{n_{d'},d'j}=1\right)$, which in combination with (4) leads to the predicted value of

$\mu_{d'+1}=\mu*\left(1-\theta_{d'+1}\right)$. We use this procedure iteratively to forecast the viral process for all future

periods of interest.

3.4.2 Estimating the probabilities (π_m,π_v) and the distribution parameters (λ_m,λ_v) of the time

to participate:

In general, we do not observe when an invited customer opens an email and decides to delete it,

and hence, to exit the campaign (see Figure 2). Therefore, we need to infer π_{md} and λ_{md}, and

π_{vd} and λ_{vd}[3] simultaneously from the observed number of participants in the viral marketing

campaign for each period d. Because the time between receiving a seeding email and

---

[3]In the empirical application, we assume both λ_{md} and λ_{vd} to be equal across days during the week, and across

days during weekends. However, both λ_{md} and λ_{vd} are allowed to be different during weekends and weekdays.

participation is assumed to be exponentially distributed, the probability that customers open an

email in period d, given they receive a seeding email before this period, equals:

$\int_{t_{d-1}}^{t_d} \lambda_{md} e^{-\lambda_m t} \, dt = e^{-\lambda_{md} t_{d-1}} - e^{-\lambda_{md} t_d}$. Hence, the probability of participating in period d, after

receiving a seeding email equals: $\psi_d = \pi_{md} \left(e^{-\lambda_{md} t_{d-1}} - e^{-\lambda_{md} t_d} \right)$. Given that K_d customers receive a

seeding email in period d, we observe in each time period $d, d+1, .., D$ how many of these

customers h_d participate, which has a multinomial distribution[4]

$[h_d, h_{d+1}, .., h_D] \sim MN\left(K_d; \psi_d, \psi_{d+1}, .., \psi_D\right)$. Because of the many observations available after only

short time periods, the parameters π_{md} and λ_{md} can be estimated using maximum likelihood. π_{vd}

and λ_{vd} are estimated in a similar fashion.

3.4.3 Estimating the immigration rate $\pi_q \beta_q$ due to seeding tool q:

Parameters β_{qd} and π_{qd}, representing the number of customers who visit the campaign website

due to seeding tool q in time period d, and π_{qd} representing the fraction of these customers who

also start participating, are directly observed and stored in the dynamically updated database. For

specific seeding tools such as banners, a marketer frequently has the opportunity to buy a

specific amount of clicks on the banner to the website. In this case, β_{qd} does not need to be

estimated and can be directly determined (i.e. set) by the marketing manager.

4. Empirical Study: A Real Life Viral Campaign

4.1 Description of the Campaign

From Friday April 1, 2005 to Friday May 6, 2005, a large financial services provider ran a viral

marketing campaign. The goal of this campaign was to promote financial services to highly

---

[4] In the empirical application we assume that the number of emails sent in period d is uniform over time, hence the
expected probability that a customer opens a seeding email in period d, given that it was received at time τ in

period d equals $\int_{t_{d-1}}^{t_d} \int_0^{t_d - \tau} \lambda_{md} e^{-\lambda_{md} t} \, dt \, d\tau = 1 - \frac{1}{\lambda_{md}} \left(1 - e^{-\lambda_{md}(t_d - t_{d-1})} \right)$.

educated potential customers aged between 20 and 29. The structure of the campaign is as shown in Figure 1. Customers participated in the campaign while playing a game during which they answered questions which led to a career profile. Then, in return for a guaranteed prize, participants could fill out an online form requesting personal information. After filling out this information, participants were informed that they could win bigger prizes if they invited one or more of their friends to the campaign by sending emails via the 'send to a friend' button. Software connected to the campaign website checked in real-time whether the email addresses of these friends were valid (i.e. each email address was filled out only once, emails were not sent to the participants themselves, and the viral email did not bounce within a pre-specified time period).

The viral campaign was online on April 1, but the organization started seeding on April 4. However, because of the novelty of the campaign, employees of the organization already started participating and inviting their contacts before the campaign was formally seeded. This resulted in 846 participants at the end of Day 3. To seed the campaign, the organization bought 6,400 banner clicks to the campaign website between April 4 and April 14 by placing a banner on a popular website. Of the 6,400 visitors, 2,200 people decided to participate in the viral campaign. Furthermore, on April 4 and 7, the marketing agency sent 4,500 and 24,258 seeding mails, respectively, to customers who agreed to receive promotional emails. These marketing activities and the resulting viral process resulted in a total of 228,351 participants by Day 36 since the viral campaign was online. Figure 5 summarizes the marketing activities around the viral campaign and the resulting number of participants by day over time. This Figure shows that the daily number of participants grew rapidly during the first 11 days, after which it slowly decreased over time. Note that during weekends the number of participants is lower, which is due to the fact that during these days customers read their email less frequently compared to weekdays, as is also

Figure 5: Events and Number of Participants by Day during the Viral Campaign
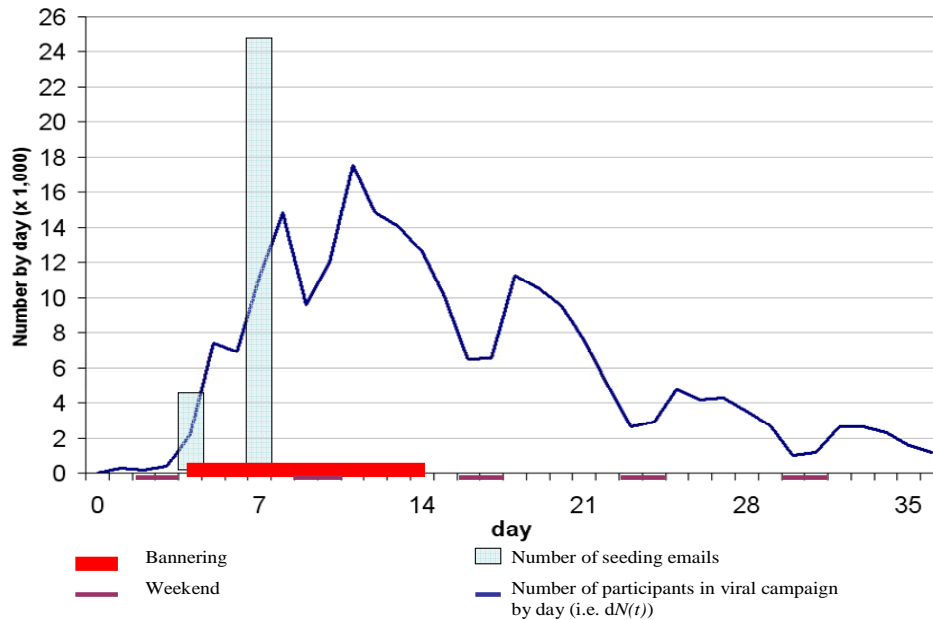
Note: The viral campaign started on a Friday and was online for 36 days. On Day 4, the number of participants grew rapidly due to marketing activities. On this day, the company sent 4,500 seeding emails and placed banners on websites that generated 200 participants by day for 11 consecutive days. On Day 7, the company sent an additional set of 24,258 seeding emails to further promote the viral campaign.

shown in the following section.

4.2 Data Description

All 228,351 participants in the viral campaign registered on the campaign website by providing their email addresses. Hence, we know the email address of each participant and the time they participated in the viral campaign. Furthermore, we also obtained the email addresses of over 1 million friends who were invited (some of which are also among the 228,351 because they actually participated), and the 28,758 seeding email addresses that the marketing agency used to seed the campaign. Given these data, we coded, for each participant, how many viral emails were sent by counting the number of viral emails that were sent to new customers who had not participated yet or had not received an invitation at the moment the emails were sent.

Next to the number of emails a participant sent, we also coded how and when a participant was invited. Unfortunately, the marketing agency did not retain the source by which a participant

was invited in their database. Therefore, we were only able to identify the source through which participants were invited by matching sent seeding and viral email addresses with the registered email addresses of participants. Using this procedure we were able to determine the source of invitation to the campaign website for 73 percent of the participants. Most of the remaining 27 percent of the customers registered under a different email address through which they were invited, most likely because of privacy concerns. This percentage closely corresponds to findings of a recent survey that showed that 42 percent of internet users have more than one email account, and that 33 percent of them provide email addresses that would not identify them personally (Wireless News 2006). From this 27 percent, we know that between April 4 and 14, 2,200 participated due to bannering. Hence, we randomly assigned 2,200 of these participants, equally distributed over the 11 days, to the banner as source of invitation. Subsequently, we computed for each day the proportions of participants for which we knew whether they were invited by a viral or seeding email. For example, on Sunday April 10, 9,245 participants (98.5 %) participated due to a viral email and 145 participants (1.5 %) participated after being invited by a seeding email. On this day, after excluding 200 participants due to banners, there were 2,406 participants for which we did not observe the source of invitation. Hence, we randomly selected 98.5% of these 2,406 participants, and we assumed that they started participating due to a viral email. For the remaining 1.5% of the participants, we assumed they were invited by a seeding email. Sensitivity analyses showed that our results are not sensitive to different choices of proportions to allocate these customers to seeding email or viral email invitation sources[5]. We repeated this procedure for all days during the campaign, so that all participants were assigned a source through which they were invited.

---

[5] In the sensitivity analyses we varied the proportions to allocate consumers to seeding emails from zero to twice as many customers as expected from the observed proportions.

In summary, after these computations, our data set consists of 228,351 lines corresponding to participants. Each line contains the identity of the participant, the date of participation, the source of invitation, the date that the participant received the invitation, the number of emails that are sent to friends, and how many of these friends already participated or were already invited.

5. Results

5.1 Performance of the Viral Branching Model

Using the procedures as described in 3.4.1 to 3.4.3 we were able to estimate the model parameters, which were subsequently plugged in equations (1) to (3) to predict the number of participants by day. To capture the effect that customers read their email less frequently during weekends, we estimated different distribution parameters of the time to participate for the weekdays and for the weekends. Using our parameter estimates, we assessed the Viral Branching Model's fit and its predictive performance. In addition to using all data during the 36 days that the campaign was online, we also estimated the parameters using only the first part of our data-set and then developed forecasts for the remaining days of the 36-day period. Because we were interested in how early in the process we would be able to accurately predict the spread of the campaign, we estimated the parameters using the data obtained in four different time periods and then developed forecasts for the remaining days of the 36-day period (i.e. hold-out periods). Because marketing activities only started on Day 4, we choose the first calibration period to be Day 1 to 7, just after the company seeded the campaign. This led to the following five scenarios:

1. Calibration Period: Day 1–7   Forecasting (Hold-out) Period: Day 8-36
2. Calibration Period: Day 1-14   Forecasting (Hold-out) Period: Day 15-36
3. Calibration Period: Day 1-21   Forecasting (Hold-out) Period: Day 22-36
4. Calibration Period: Day 1-28   Forecasting (Hold-out) Period: Day 29-36
5. Calibration Period: Day 1-36.

Furthermore, we examined whether it is worthwhile to treat viral emails separately from

seeding emails in our model. In order to test this, we also estimated a restricted version of our model by setting $\pi_m = \pi_v$ and $\lambda_m = \lambda_v$, which we call the nested Viral Branching Model. Finally, we also compared the predictive accuracy of the nested and the non-nested VBM with the simplest form of the Bass model (Bass 1969), and with an extended version of the Bass model which served as benchmarks. For the extended Bass model, we followed Kamakura and Balasubramanian (1988) and Parker (1992) and allow the market potential $\overline{N_d}$[6] to be a function of marketing activities and the innovation parameter a_d to be different for weekdays and days of the weekend, leading to the following extended Bass model:

$$N(d) - N(d-1) = \left(a_d + b\frac{N(d-1)}{\overline{N_d}} \right)\left(\overline{N_d} - N(d-1) \right). \tag{8}$$

In (8), b represents the imitation parameter, $a_d = \gamma_{a0} + \gamma_{a1} \cdot weekend(d)$, where $weekend(d)$ represents a dummy which equals one if Day d is during the weekend, zero otherwise, and

$\overline{N_d} = \gamma_{\overline{N}0} + \gamma_{\overline{N}1} \cdot \sum_{i=1}^{d} K_i + \gamma_{\overline{N}2} \cdot \sum_{i=1}^{d} \beta_i$, with K_i the number of seeding emails sent on Day i, and β_i

the number of customers who start participating due to bannering on Day i. The parameters of the Bass model are estimated so that they optimally fit the process $N(t)$, while the Viral Branching Model approach estimates parameters at the disaggregate level and, does not choose parameter values to optimize the fit of $N(t)$. The Bass model and its extended version, therefore, serve as a strong benchmark for our Viral Branching Model. This is particularly true when we compare the in-sample fit over the calibration period[7].

---

[6] To avoid confusion with the parameters of the Viral Branching Model, we slightly deviated from conventional notation of the Bass model.
[7] We tried several alternative specifications to incorporate marketing activities and weekend effects by incorporating these in functions for the innovation parameter a, imitation parameter b, and the market potential $\overline{N}$. We selected the best performing model as the extended Bass model.

In Table 1 and Table 2 we present the results of the five scenarios for the different models. Table 1 shows the in-sample fit statistics (RMSE and MAPE) and the forecasting accuracy (MAPE) for the *cumulative* number of participants (i.e. the reach $N(t)$) of the viral marketing campaign. Table 2 presents these statistics for the fit and prediction of the models for the *increase* (i.e. $dN(t)$) in the number of participants by day.

Overall, when analyzing the fit of the models, the results in Table 1 and Table 2 (see also Figure 6) indicate that our Viral Branching Model (VBM) does very well in fitting the spread of the viral marketing campaign. The fit of the nested VBM, where the effectiveness of seeding emails is assumed to be equal to that of viral emails, is extremely low. This confirms the importance of incorporating different parameters for viral and seeding emails. Furthermore, although the standard Bass model does not seem to fit the process well, the extended Bass model fits the process $N(t)$ better than our Viral Branching Model based on RMSE (1.83 vs. 6.98 for the total estimation period). Interestingly however, compared to the extended Bass model, the Viral Branching Model fits the cumulative process better based on MAPE (.05 vs. .22), and the differenced process, $dN(t)$ based on both measures (RMSE: 1.23 vs. 1.30; MAPE: .18 vs. .31).

This result is due to the fact that the parameters of the extended Bass model are chosen so that they optimize RMSE of the cumulative number of participants, and suggests that the Viral Branching Model better captures the actual process, which becomes apparent in the forecasting performance. As indicated by the results in Tables 1 and 2, and in contrast to all three competing models, the Viral Branching Model is able to accurately predict the spread of the campaign already on Day 7, when the campaign was still not fully seeded. The nested version of the model is not able to predict the number of participants accurately in the early stages of the campaign, and only starts doing better at the end of the campaign when the viral process has almost died out

28

Table 1: Model Performance – Cumulative Number of Participants in a Time Period

Estimation Period	Model	In sample fit		Out of sample forecast (MAPE) for days			
		RMSE[1]	MAPE[2]	8-14	15-21	22-28	29-36
Day 1-7	VBM	1.79	.07	.09	.03	.07	.14
	Nested VBM	4.02	.23	.39	.60	.37	.25
	Standard Bass Model	8.73	2.58	.51	.77	.82	.84
	Extended Bass Model	0.48	0.24	.08	.19	.33	.39
Day 1-14	VBM	4.47	.05	-	.02	.03	.03
	Nested VBM	44.41	.48	-	.21	.38	.46
	Standard Bass Model	15.85	2.66	-	.09	.25	.32
	Extended Bass Model	1.12	.40	-	.15	.33	.39
Day 1-21	VBM	6.06	.06	-	-	.01	.02
	Nested VBM	83.60	.58	-	-	.06	.14
	Standard Bass Model	14.79	2.51	-	-	.03	.10
	Extended Bass Model	2.35	.43	-	-	.02	.02
Day 1-28	VBM	3.48	.04	-	-	-	.01
	Nested VBM	116.54	.66	-	-	-	.01
	Standard Bass Model	12.85	2.07	-	-	-	.04
	Extended Bass Model	2.04	.28	-	-	-	.00
Day 1-36	VBM	6.98	.05	-	-	-	-
	Nested VBM	119.70	.61	-	-	-	-
	Standard Bass Model	9.90	1.65	-	-	-	-
	Extended Bass Model	1.83	.22	-	-	-	-

1. RMSE: Root Mean Squared Errors are multiplied by 1,000.
2. MAPE: Mean Absolute Percentage Error.

and does not attract many new customers. A similar phenomenon is true for the standard Bass

model. Although the extended Bass model does slightly better, it is not able to predict the

number of customers in the campaign after Day 7 or Day 14. As a matter of fact, after Day 14,

the extended Bass model hugely under predicts at 134,682 whereas the prediction of the Viral

Branching Model is at 221,429, which is very close to the true ultimate level of 228,351.The

extended Bass model starts to predict the process relatively well only after Day 21, while the

nested model and standard Bass only start to predict well after Day 28. The fact that the extended

Bass model is not able to predict the process at Day 7 or 14 confirms previous research findings

that forecasts can only be made after the inflection point (Lenk and Rao 1990), which seems to

Table 2: Model Performance – Participants by Day

Estimation Period	Model	In sample fit		Out of sample forecast (MAPE) for days			
		RMSE[1]	MAPE[2]	8-14	15-21	22-28	29-36
Day 1-7	VBM	1.12	.11	.15	.50	.61	.93
	Nested VBM	1.91	.25	.79	.88	.83	.99
	Standard Bass Model	3.73	3.26	1.00	1.00	1.00	1.00
	Extended Bass Model	0.84	0.32	.30	.59	.93	.99
Day 1-14	VBM	1.16	.08	-	.22	.24	.35
	Nested VBM	8.40	.57	-	.92	1.51	1.43
	Standard Bass Model	3.18	2.80	-	.75	.98	1.00
	Extended Bass Model	0.84	0.36	-	.82	1.00	1.00
Day 1-21	VBM	0.96	.07	-	-	.15	.31
	Nested VBM	7.65	.68	-	-	.80	1.35
	Standard Bass Model	3.18	2.62	-	-	.63	.91
	Extended Bass Model	1.62	.46	-	-	.18	.29
Day 1-28	VBM	1.01	.11	-	-	-	.33
	Nested VBM	8.49	.64	-	-	-	.34
	Standard Bass Model	2.85	2.23	-	-	-	.75
	Extended Bass Model	1.45	.35	-	-	-	.24
Day 1-36	VBM	1.23	.18	-	-	-	-
	Nested VBM	6.57	.62	-	-	-	-
	Standard Bass Model	2.57	1.88	-	-	-	-
	Extended Bass Model	1.30	.31	-	-	-	-

1. RMSE: Root Mean Squared Errors are multiplied by 1,000.
2. MAPE: Mean Absolute Percentage Error.

occur after Day 14 (see Figure 6).

5.2 Parameter Estimates of the Viral Branching Model

In addition to using the Viral Branching Model for forecasting the spread of the viral marketing campaign, we also used its parameter estimates to gain insight into the spread of information in the viral campaign. Table 3 presents the parameter estimates for our Viral Branching Model[8]. When we examine the parameter estimates, a number of observations can be made. First, on average participants sent out over four ($\mu* = 4.15$) viral emails to friends. Second, the probability that these friends start participating after receiving such an email is, on average, .26.

---

[8] We did not estimate β_q for the banners, because the company bought a fixed amount of 6,400 clicks.

Figure 6: Model Performance for Different Estimation Periods

7 days estimation period

14 days estimation period

21 days estimation period

28 days estimation period

— observed ■ Viral Branching Model ▲ Extended Bass Model
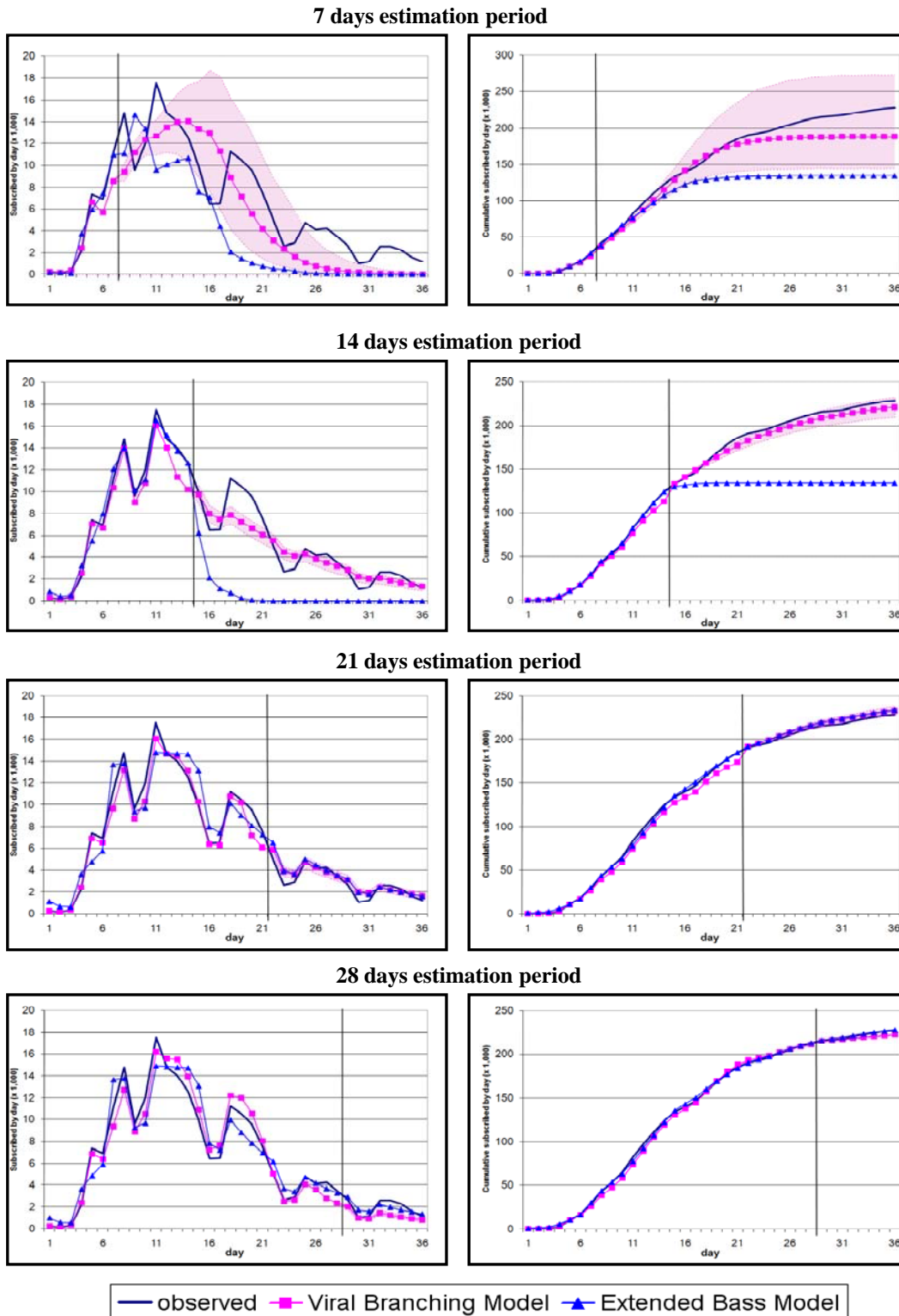
Note: Left (right) graphs reflect the (cumulative) number of participants by day for the four different calibration periods for the Viral Branching Model (■), and the Bass Model (▲). The actual values are indicated by the line (—). The shaded areas represent 95 percent prediction intervals of the Viral Branching Model (See Web Appendix B for its derivation).

Interestingly, this leads to an average infection rate of 1.08 (i.e., $\pi_v \mu^*$) at the start of the campaign, which shows that this particular viral campaign is extremely successful as the infection rate is larger than one. Hence, the number of participants grows exponentially. Note that as expected, the proportion of emails sent to customers who already received an invitation or already participated θ gradually increases over time as a function of the number of participants and people who already received an invitation, R. As explained in Section 3.4.1, equation (6), this increase is captured by a binary logit regression. The results of this analysis confirmed our expectations with $\alpha_1 = 2.99$ ($p < .01$), and $\alpha_2 = 7.24 \cdot 10^{-7}$ ($p<.01$). Consequently, at the end of the campaign the average infection rate is smaller than one and equals .87, which means that the number of additional participants does decrease over time as shown in Figure 5. This infection rate is still substantially larger than those reported by Watts and Peretti (2007), who find infection rates between .041 and .769. This emphasizes the success of the specific campaign we studied.

As expected, the probability of participation after receiving an email from a friend ($\pi_v = .26$) is substantially higher than the probability of participation after receiving a seeding email sent by a company ($\pi_m = .12$). The source of the email strongly influences its effectiveness, which is also apparent in the forecasts of the nested VBM. Interestingly, the probability of participation after a

Table 3: Parameter Estimates

	μ^*	θ	π_m	π_v	π_q	$1/\lambda_m$		$1/\lambda_v$	
						week	weekend	week	weekend
Day 1-7	4.59	4.06%	.06	.25	.34	0.69	-[1]	1.12	1.06
Day 1-14	4.29	6.47%	.10	.26	.34	1.75	2.77	1.33	1.51
Day 1-21	4.23	7.13%	.11	.26	.34	2.80	3.85	1.53	2.15
Day 1-28	4.19	7.38%	.12	.26	.34	3.31	4.39	1.59	2.80
Day 1-36	4.15	7.64%	.12	.26	.34	3.88	5.03	1.64	3.24

1. The response time to the seeding emails at the weekend could not be estimated because there were no responses, as the first seeding emails were sent just after the first weekend the campaign was online.

banner click is relatively high (i.e. $\pi_q = .34$), and even higher than that of customers who received a viral email of a friend. This is probably due to the fact that customers who click on a banner are already interested in the campaign. Still, 66 percent of these customers decide not to participate and quickly leave the campaign's landing page. The source of the email also affects the amount of time people participate in the viral campaign ($1/\lambda.$). This is more than two times shorter when the email is received from a friend rather than from a company (1.64 days vs. 3.88 days during weekdays). Note that we allowed for different estimates for λ_m, and λ_v for emails sent during weekdays and those sent during the weekend. At weekends, people probably read their emails less often leading to longer times to participate, which results in fewer participants at weekends as shown in Figure 3.

In the next Section, we explore further implications of the parameter estimates of our Viral Branching Model by examining the effects of two alternative what-if scenarios.

5.3 What-if Analyses

The Viral Branching Model does not only allow us to predict the spread of the viral marketing campaign over time, it also enables us to forecast the spread if different marketing activities are pursued. This possibility to perform what-if analyses allows marketers to use the model to support decisions about modifying the campaign in order to reach their objectives. To illustrate this possibility, we explore the effects of two alternative marketing activities. Using the model parameters of the VBM based on the estimation period of 14 days, we predict how the spread of the viral marketing campaign is different if 1) an additional 10,000 seeding emails are sent on Day 15; and 2) an additional 10,000 clicks are bought through banners that are set online for one week from Day 15 to Day 22.

Table 4 summarizes the effects of these two alternative marketing campaigns. The additional

Table 4: Predicted Effects of What-if Scenarios

Marketing activity on Day 15	Predicted cumulative number of participants on Day 36	Predicted number of additional participants	Predicted number of additional participants per click/seed
Actual marketing strategy	221,429	-	-
Extra bannering for one week: 10,000 clicks	242,595	21,166	2.17 participants/click
Extra seeding: 10,000 emails	227,640	6,211	0.62 participants/seed

10,000 seeding emails results in an additional reach of 6,211 participants at the end of the campaign on Day 36. This means that on average .62 additional participants will be reached for every seeding email. This is the number of people that directly participate by responding to the seeding email and indirectly through receiving a viral email with an invitation from a friend. It is remarkable that the effect of buying 10,000 additional banner clicks is substantially higher. This leads to an additional reach of 21,166 participants at the end of the campaign and means that the additional reach for every click is 2.17. Again, this is the sum of people who start participating directly after they have clicked the banner and the subsequently invited contacts through viral emails. Apparently, the bannering approach benefits from a self-selection mechanism. People who click on a banner may have an interest in the campaign and are then also more likely to participate and send viral emails to their friends. These effects are reflected in the model by the different probabilities of participating after receiving a seeding email ($\pi_m = .10$ for Day 1 to 14, see Table 3), and after clicking on a banner ($\pi_q = .34$, see Section 5.2). Of course, the difference between the effectiveness of these approaches will also depend on the quality of the mailing database, the characteristics of the website where the banners are placed, and the costs of these seeding tools. Figure 7 graphically shows the difference in the spread of the campaign if the two alternative scenarios are executed. It is interesting to see that effects of the additional marketing expenditures on Day 15 or shortly after do not only have an immediate effect but also a more long term effect. This is due to the indirect or viral effect following the direct effect of these

Figure 7: Results of What-if Analyses
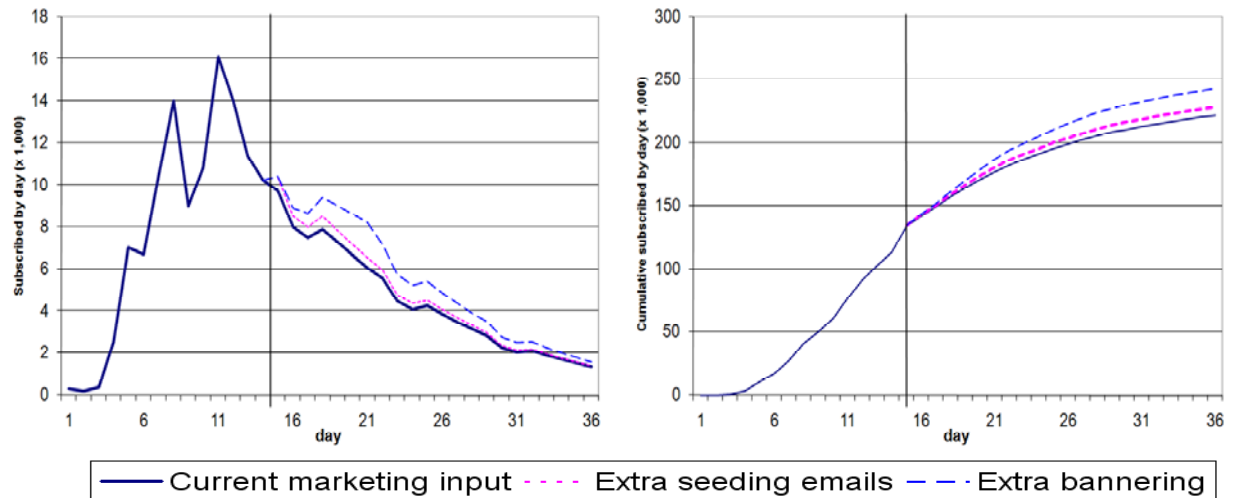
Note: Left (right) panel reflects the predictions on day 14 for the (cumulative) number of participants by day for the current marketing activities (▬) and for 2 different scenarios. In the first scenario, an additional set of 10,000 seeding emails is sent (····), in the second scenario, an additional 10,000 clicks to the campaign website are generated via bannering (‒ ‒ ·)

marketing activities. Hogan, Lemon and Libai (2004) label this the 'ripple' effect and they find

that ignoring this effect may underestimate the effectiveness of advertising campaigns. The same

is true for viral marketing campaigns and the ripple effect is likely to be even stronger for these

types of campaigns because participants are actively encouraged to further spread the campaign

among their friends. Once the rates of banner clicks and seeding emails are known, a company

can determine which seeding method is most cost-effective. Once the company can also put a

dollar value on a customer that participates (e.g. customer lifetime value) it is possible to

determine if it is profitable to carry out a particular additional seeding.

6. Discussion

Viral marketing is a relatively new way of approaching markets and communicating with

customers and can potentially achieve a large reach and a fast spread among target audiences.

Often these campaigns are relatively inexpensive since customer networks take care of spreading

the messages and no expensive media exposure needs to be purchased. The dependency on these

networks requires new modeling techniques to predict how a campaign will evolve over time and

how many customers will receive the message and participate. Using insights from epidemiology to describe the spread of viruses as a branching process, we have derived and applied a new model to predict the reach of a viral marketing campaign. In addition to predicting the spread of information, our Viral Branching Model also incorporates the effects of marketing activities such as seeding emails, bannering, and traditional advertising on this process, which standard branching models do not allow for. This enables marketers to accurately forecast the effects of their marketing activities and to analyze a variety of what-if scenarios. The application of our model on a real life viral marketing campaign shows that it is able to accurately forecast the reach of a viral marketing campaign after only a few days that the campaign is online and the company just started seeding the campaign.

Deriving the functional form of the Viral Branching Model requires solving complex differential equations. This results in closed-form solutions for the expected reach of viral marketing campaigns. Interestingly, this complex functional form of the reach is not needed to estimate the model parameters. Instead, they can be estimated relatively easily using the individual-level data that become available in large numbers early in the campaign. In fact, the functional form of the Viral Branching Model can be implemented in a spreadsheet program such as Excel, and the values of the parameter estimates can be plugged into the model to derive the reach of the viral marketing campaign over time. This makes our Viral Branching Model useful and implementable as a marketing decision support system (Lilien and Rangaswamy 2004). In addition, the model parameters provide valuable insights for managers to improve their viral marketing campaigns, because they are easily interpretable. For instance, it is insightful to monitor the switching probabilities as presented in Figure 1. A low probability means a bottleneck in the viral process, and marketers can then be advised to take appropriate measures to increase these probabilities. De Bruyn and Lilien (2008) show how these switching probabilities

depend on characteristics of the sender and the receiver of the viral email and their relationships. It would also be interesting to investigate how marketers could influence this process by changing, for example, the subject line of an email which in turn influences the probability of opening an email (Bonfrer and Drèze 2009). The number of emails sent by a participant is another important parameter that positively influences the reach of the campaign. Marketers can influence this parameter by changing the incentives to forward viral emails. Finally, in our empirical example, customers seem to read their emails less frequently during weekends compared to weekdays. This implies that it is more effective to send seeding emails on a weekday. Next to accurately forecasting and investigating alternative scenarios, managers can also use our model to compute the additional number of customers that a participant will generate in the viral marketing campaign. As shown by Hogan et al. (2004), the effectiveness of advertising is underestimated if word-of-mouth or the 'ripple' effect is not taken into account. Our model incorporates this ripple effect directly.

In our research we only focused on the number of participants in a viral marketing campaign. However, an interesting feature of online marketing is the possibility to track the behavior of visitors on websites (Manchanda, Dubé, Goh, and Chintagunta 2006). This allows marketers not only to investigate the number of customers who visited the campaign website, but also to inspect the quality of these visits. An interesting opportunity for future research would be to study the impact of viral marketing campaigns by integrating the reach of the campaign with behavioral data, such as the time customers spend on the website, which pages they visit, whether they subscribe for a service or buy specific products.

We applied the Viral Branching Model to one specific viral marketing campaign. Future research should investigate the performance of our model on other viral marketing campaigns. More interestingly, using a large set of viral marketing campaigns, it would be useful to

determine the relationships between viral marketing campaign characteristics and the value of the model parameter estimates. This will provide interesting insights into what makes a campaign successful and under which circumstances. Furthermore, such insights could be useful to predict the reach of viral marketing campaigns even before their launch. In addition to relating model parameters to campaign characteristics, it would also be valuable to investigate how model parameters evolve over time during the course of a viral marketing campaign. For instance, in our research we found that response times are slower during weekends and that the number of effectively forwarded emails decreases as more customers are invited. It is possible that in other campaigns other parameters evolve as well. For instance, the effectiveness of seeding activities may change if more customers joined the campaign. How to design these seeding tools effectively is another fruitful area for future research. For example, in a field experiment one could study the effect of timing and different formats of seeding emails and banners on traffic to the campaign website. Moreover, the effect of other media, such as blogs, and search engines would be valuable to study.

To conclude, this paper is the first to describe and predict the spread of electronic word of mouth in viral marketing campaigns. Our approach captures the interactions between customers as they are directly observed in viral marketing campaigns. Furthermore, it shows how offline and online marketing activities affect these interactions. We believe that our Viral Branching Model is a valuable tool to develop and optimize viral marketing campaigns.

References

Athreya, K. B. and P. E. Ney (1972), *Branching Processes*. Berlin: Springer-Verlag.

Bartlett, M. S. (1960), *Stochastic Population Models in Ecology and Epidemiology*. London: Methuen.

Bass, F. M. (1969), "A New Product Growth for Model Consumer Durables," *Management Science*, 15 (5), 215-227.

Biyalogorsky, E., E. Gerstner, and B. Libai (2001), "Customer Referral Management: Optimal Reward Programs," *Marketing Science*, 20 (1), 82-95.

Blattberg, R. and J. Golanty (1978), "Tracker: An Early Test Market Forecasting and Diagnostic Model for New Product Planning," *Journal of Marketing Research*, 15 (May), 192-202.

Bonfrer, A. and X. Drèze (2009), "Real-Time Evaluation of E-Mail Campaign Performance," *Marketing Science*, 28 (2), 251-263.

Chiu, H.-C., Y.-C. Hsieh, Y.-H. Kao, and M. Lee (2007), "The Determinants of Email Receivers' Disseminating Behaviors on the Internet," *Journal of Advertising Research*(December), 524-534.

De Bruyn, A. and G. L. Lilien (2008), "A Multi-Stage Model of Word of Mouth Influence through Viral Marketing," *International Journal of Research in Marketing*, 25 (3), 151-163.

Dorman, K. S., J. S. Sinsheimer, and K. Lange (2004), "In the Garden of Branching Processes," *SIAM Review*, 46 (2), 202-229.

Eliashberg, J., J.-J. Jonker, M. S. Sawhney, and B. Wierenga (2000), "Moviemod: An Implementable Decision Support System for Pre-Release Market Evaluation of Motion Pictures," *Marketing Science*, 19 (3), 226-243.

Godes, D., D. Mayzlin, Y. Chen, S. Das, C. Dellarocas, B. Pfeiffer, B. Libai, S. Sen, M. Shi, and P. Verlegh (2005), "The Firm's Management of Social Interactions," *Marketing Letters*, 16 (3/4), 415-428.

Harris, T. E. (1963), *The Theory of Branching Processes*. Berlin: Springer-Verlag.

Hauser, J. R. and K. J. Wisniewski (1982), "Application, Predictive Test, and Strategy Implications for a Dynamic Model of Consumer Response," *Marketing Science*, 1 (2), 143-179.

Hogan, J. E., K. N. Lemon, and B. Libai (2004), "Quantifying the Ripple: Word-of-Mouth and Advertising Effectiveness," *Journal of Advertising Research*, September, 271-280.

Kalyanam, K., S. McIntyre, and J. T. Masonis (2007), "Adaptive Experimentation in Interactive Marketing: The Case of Viral Marketing at Plaxo," *Journal of Interactive Marketing*, 21 (3), 72-85.

Kamakura, W. A. and S. K. Balasubramanian (1988), "Long-Term View of the Diffusion of Durables: A Study of the Role of Price and Adoption Influence Processes Via Tests of Nested Models," *International Journal of Research in Marketing*, 5, 1-13.

Kendall, D. G. (1949), "Stochastic Processes and Population Growth," *Journal of the Royal Statistical Society: Series B*, 11 (2), 230-264.

Lenk, P. J. and A. G. Rao (1990), "New Models from Old: Forecasting Product Adoption by Hierarchical Bayes Procedures," *Marketing Science*, 9 (1), 42-53.

Lilien, G. L. and A. Rangaswamy (2004), *Marketing Engineering: Computer-Assisted Marketing Analysis and Planning*, (Revised Second Edition ed.). Victoria, BC, Canada: Trafford Publishing.

Manchanda, P., J.-P. Dubé, K. Y. Goh, and P. K. Chintagunta (2006), "The Effect of Banner Advertising on Internet Purchasing," *Journal of Marketing Research*, 43 (February), 98-108.

Moe, W. (2003), "Buying, Searching, or Browsing: Differentiating between Online Shoppers Using in-Store Navigational Clickstream," *Journal of Consumer Psychology*, 13 (1&2), 29-39.

Morrissey, B. (2007), "Clients Try to Manipulate 'Unpredictable' Viral Buzz," *Adweek*, 48 (March 19), 12.

New Media Age (2007), "Red Nose Day Viral Game Played 1.16m Times," (March 29), 3.

Parker, P. M. (1992), "Price Elasticity Dynamics over the Adoption Life Cycle," *Journal of Marketing Research*, 29 (3), 358-367.

Phelps, J. E., R. Lewis, L. Mobilo, D. Perry, and N. Raman (2004), "Viral Marketing or Electronic Word-of-Mouth Advertising: Examining Consumer Responses and Motivations to Pass Along Email," *Journal of Advertising Research*(December), 333-348.

Ross, S. M. (1997), *Introduction to Probability Models*. San Diego, CA: Academic Press.

Sevast'yanov, B. A. (1957), "Limit Theorems for Branching Stochastic Processes of Special Form," *Theory of Probability and its Applications*, 2 (3), 321-331.

Shocker, A. D. and W. G. Hall (1986), "Pretest Market Models: A Critical Evaluation," *Journal of Product Innovation Management*, 3, 86-107.

Silk, A. J. and G. L. Urban (1978), "Pre-Test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," *Journal of Marketing Research*, 15 (May), 171-191.

Urban, G. L. (1970), "Sprinter Mod III: A Model for the Analysis of New Frequently Purchased Consumer Products," *Operations Research*, 18 (5), 805-854.

Urban, G. L. (1975), "Perceptor: A Model for Product Postioning," *Management Science*, 21 (8), 858-871.

Watts, D. J. and J. Peretti (2007), "Viral Marketing for the Real World," *Harvard Business Review*, May, 22-23.

Wireless News (2006), "Truste/TNS Survey: Most Internet Users Are Not Taking Action to Protect Online Privacy," Dec. 8, 1.

Wyck, S. v. (2007), "Viral Is Worth the Investment," *B&T Weekly*, 57 (February 23), 14.

Derivation of the Viral Branching Process Variables: $M(t)$, $V(t)$, **and** $N(t)$

Web Appendix A derives the expectations of the three stochastic processes $M(t)$, $V(t)$, and

$N(t)$ of the viral branching model. The process denoted by $M(t)$ captures the number of

unopened seeding emails. The process $V(t)$ captures the number of unopened viral emails and it

depends on $M(t)$, and includes immigration that is the number of viral emails may also increase

due to consumers that participate because of other sources $q \in Q$ than seeding or viral emails,

such as banners and traditional advertising. Finally, the process $N(t)$ denotes the number of

participants in the viral campaign and depends on both processes $M(t)$ and $V(t)$. Since the

viral branching model, represented by the processes $M(t)$, $V(t)$, and $N(t)$, is a continuous

time Markov process, we can derive the Kolmogorov forward equations. This is done in the first

Section of Web Appendix A. These differential equations represent the probability distributions

that the three stochastic processes should satisfy. Since these differential equations do not have a

closed form solution, we use them in the second section to derive the differential equations of the

probability generating functions. In the final section we use these probability generating

functions to derive closed-form solutions for the first moments of $M(t)$, $V(t)$, and $N(t)$.

1. Derivation of the Kolmogorov Forward Equations

Let $P_{\mathbf{ik}}(t)$ denote the transition probability of switching from state $\mathbf{i} = \left(i_m, i_v, i_n\right)'$ to

$\mathbf{k} = \left(k_m, k_v, k_n\right)'$ in time t (i.e., $P_{\mathbf{ik}}(t) = P\left(Z(t+s) = \mathbf{k} \mid Z(s) = \mathbf{i}\right)$, with $s > 0$ and

$Z(t) = \left\{M(t), V(t), N(t)\right\}$, (see Ross 1997)), where $\mathbf{i} = \left(i_m, i_v, i_n\right)'$ and $\mathbf{k} = \left(k_m, k_v, k_n\right)'$ are

nonnegative integers counting respectively the number of unopened seeding emails (indicated by

subscript m), unopened viral emails (indicated by subscript v), and number of participants

(indicated by subscript n). The Kolmogorov forward equations are defined as follows (Ross

1997):

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{\mathbf{ik}}(t) = \sum_{\mathbf{j} \neq \mathbf{k}} h_{\mathbf{jk}} P_{\mathbf{ij}}(t) - w_{\mathbf{k}} P_{\mathbf{ik}}(t), \tag{A1}$$

for all $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, with $\mathbf{j} = \left(j_m, j_v, j_n \right)'$. In (A1), $w_{\mathbf{k}}$ indicates the rate at which the process makes a transition given it is in state $\mathbf{k}$. This transition occurs due to the three types of sources $b \in \{m, v, Q\}$, i.e. when 1) a customer opens a seeding email (m), 2) a customer opens a viral email (v), and 3) a customer participates in the viral campaign by accepting an invitation from another source $q \in Q$. Because of the assumptions that the time between receiving a seeding or viral email and participating in the campaign is exponentially distributed with parameters λ_m and λ_v respectively, a transition from state k due to a seeding email occurs at rate $k_m \lambda_m$ and a transition due to a viral email happens at rate $k_v \lambda_v$ (i.e., the number of unopened seeding and viral emails multiplied by the speed in which seeding and viral emails are opened respectively[9]). We model the third possibility, i.e. the process making a transition due to other sources Q given it is in state $\mathbf{k}$, using an immigration process (Harris 1963). This allows consumers to participate in the viral campaign at a given exponentially distributed rate, without being invited by seeding or viral emails. Since a customer participates in the viral campaign due to source $q \in Q$ at rate $\pi_q \beta_q$, where β_q is the exponentially distributed rate at which customers are invited by seeding tool q and π_q is the probability that such a customer subsequently participates in the campaign, given that it is invited by source q. Hence, given seeding sources Q, transitions from state $\mathbf{k}$ due to these sources occur at rate $\sum_{q=1}^{Q} \pi_q \beta_q$. Because all rates are independent and exponentially distributed, we add these three possibilities of making a transition from state $\mathbf{k}$, to arrive at the overall rate $w_{\mathbf{k}}$ at which the process makes a transition equals from state $\mathbf{k}$:

$$w_{\mathbf{k}} = k_m \lambda_m + k_v \lambda_v + \sum_{q=1}^{Q} \pi_q \beta_q. \tag{A2}$$

In (A1), $h_{\mathbf{jk}}$ represents the instantaneous transition rates that equal $h_{\mathbf{jk}} = w_{\mathbf{j}} r_{\mathbf{jk}}$ (Ross 1997), where $r_{\mathbf{jk}}$ denotes the probability that a transition will occur into state $\mathbf{k}$ given that the process is currently in state $\mathbf{j}$. To derive $r_{\mathbf{jk}}$, note that transitions may occur due to three types of sources of

---

[9] Note that if X_1, X_2, .., X_k are independent exponentially distributed random variables with parameter λ, than the minimum of these random variables, i.e. $\min \{ X_1, X_2, .., X_k \}$, is exponentially distributed with parameter $k\lambda$.

invitation $b \in \{m, v, Q\}$. Therefore, we define $p_{j_z k_z}^{z,b}$ to denote the transition probability of process $z \in \{m, v, n\}$, representing respectively the number of unopened seeding emails (m), number of unopened viral emails (v), and number of participants (n), due to invitation source type $b \in \{m, v, Q\}$. Using these definitions, the probability that the process switches from state $\mathbf{j}$ to state $\mathbf{k}$ due to invitation source b equals: $r_{\mathbf{jk}} = r_{(j_m, j_v, j_n),(k_m, k_v, k_n)} = \prod_{z \in \{m,v,n\}} p_{j_z k_z}^{z,b} = p_{j_m k_m}^{m,b} p_{j_v k_v}^{v,b} p_{j_n k_n}^{n,b}$.

Hence, given the three types of seeding sources $b \in \{m, v, Q\}$, and the fact that $h_{\mathbf{jk}} = w_{\mathbf{j}} r_{\mathbf{jk}}$ and using (A2), we get:

$$h_{\mathbf{jk}} = j_m \lambda_m p_{j_m k_m}^{mm} p_{j_v k_v}^{vm} p_{j_n k_n}^{nm} + j_v \lambda_v p_{j_m k_m}^{mv} p_{j_v k_v}^{vv} p_{j_n k_n}^{nv} + \sum_{q=1}^{Q} \pi_q \beta_q p_{j_m k_m}^{mq} p_{j_v k_v}^{vq} p_{j_n k_n}^{nq}. \qquad (A3)$$

Note that the process $M(t)$ only decreases when a customer opens a seeding email of the company, i.e. $p_{j_m k_m}^{mm} = 1$ when $j_m = k_m + 1$, zero otherwise, and does not change due to other sources $b = \{v, Q\}$, i.e. $p_{j_m k_m}^{mv} = p_{j_m k_m}^{mq} = 1$ for all $q \in Q$ when $j_m = k_m$, zero otherwise. On the other hand, $V(t)$ may change due to all three types of sources b. First, due to opening a seeding email (m), a customer decides to send one or more viral emails after participating in the viral campaign due to opening a seeding email. Second, due to opening a viral email (v) a customer decides to forward viral emails to two or more friends, i.e. $V(t)$ increases, or a customer decides not to invite any friend and $V(t)$ decreases by one. Third, due to source $q \in Q$, a customer participates in the campaign and decides to invite one ore more friends by sending a viral email. When the change is due to company activities, i.e. seeding (m) or other sources $q \in Q$, $V(t)$ cannot decrease. Hence, given that a consumer participates in the campaign with probability π_m due to opening a seeding email, $p_{j_v k_v}^{vm} = \begin{cases} 0 & \text{if } k_v < j_v \\ \pi_m \phi_{k_v - j_v} & \text{if } k_v \geq j_v \end{cases}$, where $\phi_{k_v - j_v}$ indicates the probability that a consumer sends $k_v - j_v$ viral emails to friends that have not been invited or did not participate yet. Similarly $p_{j_v k_v}^{vq} = \begin{cases} 0 & \text{if } k_v < j_v \\ \pi_q \phi_{k_v - j_v} & \text{if } k_v \geq j_v \end{cases}$ when the change is due to source $q \in Q$ with probability π_q. However, as described above, when a customer participates with probability

π_v after receiving a viral email, $V(t)$ may also decrease which gives the following:

$$p_{j_v k_v}^{vv} = \begin{cases} 0 & \text{if } k_v < j_v \\ \pi_v \phi_{k_v - j_v + 1} & \text{if } k_v \geq j_v - 1 \end{cases}$$. Next, since $N(t)$ counts the number of participants that

participated in the viral campaign, and at most one participant can start participating in the viral

campaign, $p_{j_n k_n}^{nb} = 1$ if $j_n = k_n - 1$, and zero otherwise for all sources $b = \{m, v, Q\}$.

Using these derivations of the transition probabilities $p_{j_z k_z}^{z,b}$ in combination with (A3), the

Kolmogorov forward equations (A1) of a viral marketing campaign become:

$$
\begin{aligned}
\frac{d}{dt} P_{ik}(t) &= (k_m + 1)\lambda_m \left(\pi_m \sum_{j_v=0}^{k_v} \phi_{k_v - j_v} P_{(i_m, i_v, i_n)(k_m+1, j_v, k_n-1)}(t) + (1 - \pi_m) P_{(i_m, i_v, i_n)(k_m+1, k_v, k_n)}(t) \right) \\
&+ \lambda_v \left(\pi_v \sum_{j_v=1}^{k_v+1} j_v \cdot \phi_{k_v - j_v + 1} P_{(i_m, i_v, i_n)(k_m, j_v, k_n-1)}(t) + (1 - \pi_v)(k_v + 1) P_{(i_m, i_v, i_n)(k_m, k_v+1, k_n)}(t) \right) \\
&+ \sum_{q=1}^{Q} \pi_q \beta_q \sum_{j_v=0}^{k_v} \phi_{k_v - j_v} P_{(i_m, i_v, i_n)(k_m, j_v, k_n-1)}(t) \\
&- \left(k_m \lambda_m + k_v \lambda_v + \sum_{q=1}^{Q} \pi_q \beta_q \right) P_{(i_m, i_v, i_n)(k_m, k_v, k_n)}(t)
\end{aligned}
\qquad \text{(A4)}
$$

Equation (A4) consists of four parts (corresponding to the four lines at the right-hand-side of the

equation). Recalling that the first part of (A4) denotes:

$$(k_m + 1)\lambda_m \left(\pi_m \overbrace{\sum_{j_v=0}^{k_v} \phi_{k_v - j_v} P_{(i_m, i_v, i_n)(k_m+1, j_v, k_n-1)}(t)}^{\text{Customer accepts seeding invitation}} + (1 - \pi_m) \overbrace{P_{(i_m, i_v, i_n)(k_m+1, k_v, k_n)}(t)}^{\text{Customer rejects seeding invitation}} \right) \quad \text{(A4.1)}$$

accounts for two situations. In the first situation, the customer opens the seeding invitation and

participates in the campaign with probability π_m, and the process $V(t)$ changes from j_v to k_v if

this customer forwards $k_v - j_v$ viral emails which happens with probability $\phi_{k_v - j_v}$. Furthermore,

$N(t)$ increases by 1 and hence j_n should equal k_{n-1} in order to switch to k_n. In the second

situation when the customer opens a seeding email but decides not to participate in the viral

campaign, which happens with probability $1 - \pi_m$, only the process $M(t)$ changes and

decreases by one, $V(t)$ and $N(t)$ are left unchanged. Similarly, recalling that the second part of

(A4) also denotes two situations:

$$\lambda_v \left(\pi_v \overbrace{\sum_{j_v=1}^{k_v+1} j_v \cdot \phi_{k_v-j_v+1} P_{(i_m,i_v,i_n)(k_m,j_v,k_n-1)}}^{\text{Customer accepts viral invitation}} (t) + (1-\pi_v) \overbrace{(k_v+1) P_{(i_m,i_v,i_n)(k_m,k_v+1,k_n)}}^{\text{Customer rejects viral invitation}} \right). \qquad \text{(A4.2)}$$

In the first situation, the customer opens a viral email and participates the viral campaign with probability π_v. If the process switches from state **j** to state **k**, this customer needs to send $k_v - j_v + 1$ viral emails which happens with probability $\pi_v \phi_{k_v-j_v+1}$, and the arrival rate of such a customer equals $\lambda_v j_v$. In this situation, $N(t)$ increases by 1 and hence j_n should equal k_{n-1} in order to switch to k_n. In the second situation of (A4.2), the customer decides to reject the viral invitation and $V(t)$ decreases by 1 (so $j_v = k_v + 1$), leaving the other two process $M(t)$ and $N(t)$ unchanged. This situation occurs with probability $(1-\pi_v)$ and at speed $\lambda_v j_v = \lambda_v (k_v + 1)$.

The third part of (A4):

$$\sum_{q=1}^{Q} \pi_q \beta_q \sum_{j_v=0}^{k_v} \phi_{k_v-j_v} P_{(i_m,i_v,i_n)(k_m,j_v,k_n-1)}(t), \qquad \text{(A4.3)}$$

represents participation due to seeding sources $q \in Q$ at rate $\sum_{q=1}^{Q} \beta_q$ with probabilities π_q. In this case $M(t)$ remains the same, $N(t)$ increases by one, hence $j_n = k_n - 1$. Furthermore, the process $V(t)$ may increase from state j_v to k_v if the customer forwards $k_v - j_v$ viral emails which occurs with probability $\phi_{k_v-j_v}$. Finally, recalling that part four of (A4):

$$-\left(k_m \lambda_m + k_v \lambda_v + \sum_{q=1}^{Q} \pi_q \beta_q \right) P_{(i_m,i_v,i_n)(k_m,k_v,k_n)}(t), \qquad \text{(A4.4)}$$

incorporates the rate $w_{\mathbf{k}}$ at which the process makes a transition (see also A2).

Solving equation (A4) for arbitrary combinations of **i**, **k**, and t results in the complete probability distribution of the viral marketing campaign over time. However, the computations are highly cumbersome, as there is generally no analytical solution that expresses its probability distribution, except for very special cases such as the birth and death process (Athreya and Ney 1972). However, it is possible to derive the differential equation of the probability generating function of the process using equation (A4) (Athreya and Ney 1972; Harris 1963), which we describe in the following Section.

2. Derivation of the Probability Generating Function

Each probability distribution has a unique probability generating function from which we are able to derive its moments. Therefore, probability generating functions are popular representations of distributions especially when analytical representations are unknown. The probability generating function $F_{\mathbf{i}}(\mathbf{s},t)$ of the viral branching process is defined as follows:

$$F_{\mathbf{i}}(\mathbf{s},t)=F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)=\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}P_{(i_m,i_v,i_n),(k_m,k_v,k_n)}(t)\,s_m^{k_m}s_v^{k_v}s_n^{k_n},\quad\text{with }|\mathbf{s}|\le\mathbf{1}.\quad(A5)$$

To derive the conditional moments of the corresponding distribution, we only need to differentiate to $\mathbf{s}$ and evaluate the resulting equation in $\mathbf{s}=\mathbf{1}$. For example

$$E\big(N(t)\,|\,N(t')=i_n\big)=\frac{\mathrm{d}}{\mathrm{d}s_n}F_{\mathbf{i}}(\mathbf{s}=\mathbf{1},t),\text{ and }E\big(M(t)\,|\,M(t')=i_m\big)=\frac{\mathrm{d}}{\mathrm{d}s_m}F_{\mathbf{i}}(\mathbf{s}=\mathbf{1},t).\text{ To obtain the}$$

differential equation that $F_{\mathbf{i}}(\mathbf{s},t)$ must satisfy, we multiply (A4) by $s_m^{k_m}s_v^{k_v}s_n^{k_n}$, and sum the resulting equation over k_m, k_v and k_n. For (A4.1) this leads to:

$$\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{k_m}s_v^{k_v}s_n^{k_n}(k_m+1)\lambda_m\left(\pi_m\sum_{j_v=0}^{k_v}p_{k_v-j_v}P_{(i_m,i_v,i_n)(k_m+1,j_v,k_n-1)}(t)+(1-\pi_m)P_{(i_m,i_v,i_n)(k_m+1,k_v,k_n)}(t)\right).\;(A6)$$

Letting k_m run from 1 to infinity, and recognizing that $P_{(i_m,i_v,i_n)(k_m,j_v,k_n-1)}=0$ for $k_n=0$, leads to the following result for the first part of (A6):

$$\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{k_m}s_v^{k_v}s_n^{k_n}(k_m+1)\lambda_m\pi_m\sum_{j_v=0}^{k_v}p_{k_v-j_v}P_{(i_m,i_v,i_n)(k_m+1,j_v,k_n-1)}(t)=$$
$$\sum_{k_m=1}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{k_m-1}s_v^{k_v}s_n^{k_n+1}k_m\lambda_m\pi_m\sum_{j_v=0}^{k_v}p_{k_v-j_v}P_{(i_m,i_v,i_n)(k_m,j_v,k_n)}(t)\qquad.\;(A7.1)$$

Noting that $\displaystyle\sum_{k_v=0}^{\infty}s_v^{k_v}\sum_{j_v=0}^{k_v}\phi_{k_v-j_v}P_{(i_m,i_v,i_n)(k_m,j_v,k_n)}(t)=\sum_{k_v=0}^{\infty}s_v^{k_v}\sum_{k=0}^{\infty}\phi_k s_v^k P_{(i_m,i_v,i_n)(k_m,k_v,k_n)}(t)$ in (A7.1), leads to:

$$\sum_{k_m=1}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{k_m-1}s_v^{k_v}s_n^{k_n+1}k_m\lambda_m\pi_m\sum_{k=0}^{\infty}\phi_k s_v^k P_{(i_m,i_v,i_n)(k_m,k_v,k_n)}(t).\qquad(A7.2)$$

Note that $\displaystyle\sum_{k_m=1}^{\infty}k_m s_m^{k_m-1}=\sum_{k_m=0}^{\infty}k_m s_m^{k_m-1}=\frac{\mathrm{d}}{\mathrm{d}s_m}\sum_{k_m=0}^{\infty}s_m^{k_m}$, and taking into account (A5), this leads to the following result for the first part of (A6):

$$s_n \lambda_m \pi_m \sum_{k=0}^{\infty} \phi_k s_v^{\ k} \frac{\mathrm{d} F_{(i_m, i_v, i_n)}\left(s_m, s_v, s_n, t\right)}{\mathrm{d} s_m}. \tag{A7.3}$$

Similarly, by letting k_m run from 1 to infinity, the second part of (A6) becomes:

$$\sum_{k_m=0}^{\infty} \sum_{k_v=0}^{\infty} \sum_{k_n=0}^{\infty} s_m^{\ k_m} s_v^{\ k_v} s_n^{\ k_n} \left(k_m + 1\right) \lambda_m \left(1 - \pi_m\right) P_{(i_m, i_v, i_n)(k_m+1, k_v, k_n)}(t) =$$

$$\sum_{k_m=1}^{\infty} \sum_{k_v=0}^{\infty} \sum_{k_n=0}^{\infty} s_m^{\ k_m-1} s_v^{\ k_v} s_n^{\ k_n} k_m \lambda_m \left(1 - \pi_m\right) P_{(i_m, i_v, i_n)(k_m, k_v, k_n)}(t) \tag{A8.1}$$

Similar to step from (A7.2) to (A7.3), we observe that $\sum\limits_{k_m=1}^{\infty} k_m s_m^{\ k_m-1} = \dfrac{\mathrm{d}}{\mathrm{d} s_m} \sum\limits_{k_m=0}^{\infty} s_m^{\ k_m}$. Combining

this with definition (A5), (A8.1) equals:

$$\lambda_m \left(1 - \pi_m\right) \frac{\mathrm{d} F_{(i_m, i_v, i_n)}\left(s_m, s_v, s_n, t\right)}{\mathrm{d} s_m} \tag{A8.2}$$

Multiplying (A4.2) by $s_m^{\ k_m} s_v^{\ k_v} s_n^{\ k_n}$, and summing the resulting equation over k_m, k_v and k_n,

leads to:

$$\sum_{k_m=0}^{\infty} \sum_{k_v=0}^{\infty} \sum_{k_n=0}^{\infty} s_m^{\ k_m} s_v^{\ k_v} s_n^{\ k_n} \lambda_v \left(\pi_v \sum_{j_v=1}^{k_v+1} j_v \cdot \phi_{k_v-j_v+1} P_{(i_m, i_v, i_n)(k_m, j_v, k_n-1)}(t) + \left(1 - \pi_v\right)\left(k_v+1\right) P_{(i_m, i_v, i_n)(k_m, k_v+1, k_n)} \right). \tag{A9}$$

Noting that $P_{(i_m, i_v, i_n)(k_m, j_v, k_n-1)} = 0$ for $k_n = 0$ leads to the following for the first part of (A9):

$$\sum_{k_m=0}^{\infty} \sum_{k_v=0}^{\infty} \sum_{k_n=0}^{\infty} s_m^{\ k_m} s_v^{\ k_v} s_n^{\ k_n} \lambda_v \pi_v \sum_{j_v=1}^{k_v+1} j_v \cdot \phi_{k_v-j_v+1} P_{(i_m, i_v, i_n)(k_m, j_v, k_n-1)}(t) =$$

$$\sum_{k_m=0}^{\infty} \sum_{k_v=0}^{\infty} \sum_{k_n=0}^{\infty} s_m^{\ k_m} s_v^{\ k_v} s_n^{\ k_n+1} \lambda_v \pi_v \sum_{j_v=1}^{k_v+1} j_v \cdot \phi_{k_v-j_v+1} P_{(i_m, i_v, i_n)(k_m, j_v, k_n)}(t) \tag{A10.1}$$

Noting that $\sum\limits_{k_v=0}^{\infty} s_v^{\ k_v} \sum\limits_{j_v=1}^{k_v+1} j_v \phi_{k_v-j_v+1} P_{(i_m, i_v, i_n)(k_m, j_v, k_n)} = \sum\limits_{k_v=0}^{\infty}\left(k_v+1\right) s_v^{\ k_v} \sum\limits_{k=0}^{\infty} \phi_k s_v^{\ k} P_{(i_m, i_v, i_n)(k_m, k_v+1, k_n)}$ in (A10.1),

leads to:

$$\sum_{k_m=0}^{\infty} \sum_{k_v=0}^{\infty} \sum_{k_n=0}^{\infty} s_m^{\ k_m} \left(k_v+1\right) s_v^{\ k_v} s_n^{\ k_n+1} \lambda_v \pi_v \sum_{k=0}^{\infty} \phi_k s_v^{\ k} P_{(i_m, i_v, i_n)(k_m, k_v+1, k_n)}(t). \tag{A10.2}$$

Letting k_v run from 1 to infinity and observing that (A10.3) is equal to zero if $k_v = 0$, (A10.2)

can be written as:

$$\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{\ k_m}k_v s_v^{\ k_v-1}s_n^{\ k_n+1}\lambda_v\pi_v\sum_{k=0}^{\infty}\phi_k s_v^k P_{(i_m,i_v,i_n)(k_m,k_v,k_n)}(t).\tag{A10.3}$$

Note again that $\displaystyle\sum_{k_v=0}^{\infty}k_v s_v^{\ k_v-1}=\frac{\mathrm{d}}{\mathrm{d}\,s_v}\sum_{k_v=0}^{\infty}s_v^{\ k_v}$, which leads to the following expression for (A10.3):

$$\lambda_v s_n \pi_v \sum_{k=0}^{\infty}\phi_k s_v^k \frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}\,s_v}.\tag{A10.4}$$

By letting k_v run from 1 to infinity, and observing that the rhs of (A11.1) equals zero if $k_v=0$,

the second part of (A9) becomes:

$$\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{\ k_m}s_v^{\ k_v}s_n^{\ k_n}\lambda_v\left(1-\pi_v\right)\left(k_v+1\right)P_{(i_m,i_v,i_n)(k_m,k_v+1,k_n)}=$$
$$\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{\ k_m}s_v^{\ k_v-1}s_n^{\ k_n}\lambda_v\left(1-\pi_v\right)k_v P_{(i_m,i_v,i_n)(k_m,k_v,k_n)}\tag{A11.1}$$

Note again that $\displaystyle\sum_{k_v=0}^{\infty}k_v s_v^{\ k_v-1}=\frac{\mathrm{d}}{\mathrm{d}\,s_v}\sum_{k_v=0}^{\infty}s_v^{\ k_v}$, which leads to the following expression for (A11.1):

$$\lambda_v\left(1-\pi_v\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}\,s_v}.\tag{A11.2}$$

The multiplication of (A4.3) by $s_m^{\ k_m}s_v^{\ k_v}s_n^{\ k_n}$, and summing the resulting equation over k_m , k_v

and k_n , leads to:

$$\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{\ k_m}s_v^{\ k_v}s_n^{\ k_n}\sum_{q=1}^{Q}\pi_q\beta_q\sum_{j_v=0}^{k_v}\phi_{k_v-j_v}P_{(i_m,i_v,i_n)(k_m,j_v,k_n-1)}(t).\tag{A12}$$

Taking into account that $\displaystyle\sum_{k_v=0}^{\infty}s_v^{\ k_v}\sum_{j_v=0}^{k_v}\phi_{k_v-j_v}P_{(i_m,i_v,i_n)(k_m,j_v,k_n-1)}(t)=\sum_{k_v=0}^{\infty}s_v^{\ k_v}\sum_{k=0}^{\infty}\phi_k s_v^k P_{(i_m,i_v,i_n)(k_m,k_v,k_n-1)}(t)$ as

noted above, and recognizing that $P_{(i_m,i_v,i_n)(k_m,k_v,k_n-1)}=0$ for $k_n=0$, (A12) can be written as:

$$\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty}s_m^{\ k_m}s_v^{\ k_v}s_n^{\ k_n+1}\sum_{q=1}^{Q}\pi_q\beta_q\sum_{k=0}^{\infty}\phi_k s_v^k P_{(i_m,i_v,i_n)(k_m,k_v,k_n)}(t).\tag{A13.1}$$

Given the definition in (A5), (A13.1) can be written as:

$$s_n\sum_{q=1}^{Q}\pi_q\beta_q\sum_{k=0}^{\infty}\phi_k s_v^k \cdot F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right).\tag{A13.2}$$

Finally, the multiplication of (A4.4) by $s_m{}^{k_m} s_v{}^{k_v} s_n{}^{k_n}$, and summing the resulting equation over k_m, k_v and k_n, leads to:

$$-\sum_{k_m=0}^{\infty}\sum_{k_v=0}^{\infty}\sum_{k_n=0}^{\infty} s_m{}^{k_m} s_v{}^{k_v} s_n{}^{k_n}\left(k_m\lambda_m + k_v\lambda_v + \sum_{q=1}^{Q}\pi_q\beta_q \right) P_{(i_m,i_v,i_n)(k_m,k_v,k_n)}(t) =$$

$$-\left(k_m\lambda_m + k_v\lambda_v + \sum_{q=1}^{Q}\pi_q\beta_q \right)\cdot F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t) \tag{A14}$$

Given these derivations, the differential equation of the probability generating function of the viral branching model is equal to the sum of equations (A7.3), (A8.2), (A10.4), (A11.2), (A13.2), and (A14), which equals:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t) &= \lambda_m\left((1-\pi_m) + s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v{}^k - s_m \right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{\mathrm{d}s_m} \\
&+ \lambda_v\left((1-\pi_v) + s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v{}^k - s_v \right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{\mathrm{d}s_v} \\
&+ \sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v{}^k - 1 \right)F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)
\end{aligned}
\tag{A15}
$$

Using (A15), we are now able to derive the moments $E\big(M(t)\,|\,M(t')\big)$, $E\big(V(t)\,|\,V(t')\big)$, and $E\big(N(t)\,|\,N(t')\big)$, with $0 \le t' \le t$, of the viral marketing processes in the next Section.

3. Derivation of the moments of the Viral Branching Model

Derivation of $E\big(M(t)\,|\,M(t')=i_m\big)$:

Let, $M(i_m,t) = E\big(M(t)\,|\,M(t')=i_m\big) = \dfrac{\mathrm{d}}{\mathrm{d}s_m}F_{(i_m,i_v,i_n)}(s_m=1, s_v=1, s_n=1, t)$. Differentiating (A15) to s_m leads to the following equation:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s_m}F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t) &= -\lambda_m\frac{\mathrm{d}F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{\mathrm{d}s_m} \\
&+ \lambda_m\left((1-\pi_m) + s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v{}^k - s_m \right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{\mathrm{d}s_m\,\mathrm{d}s_m} \\
&+ \lambda_v\left((1-\pi_v) + s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v{}^k - s_v \right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{\mathrm{d}s_v\,\mathrm{d}s_m} \\
&+ \sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v{}^k - 1 \right)\frac{F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{\mathrm{d}s_m}
\end{aligned}
\tag{A16}
$$

Setting $s_m = s_v = s_n = 1$ in (A16), and by observing that $\sum_{k=0}^{\infty} \phi_k s_v^{\ k} = 1$ if $s_v = 1$, we get $M\left(i_m, t\right)$ by solving the following differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} M\left(i_m, t\right) = -\lambda_m \cdot M\left(i_m, t\right). \tag{A17}$$

Using the fact that $M\left(i_m, t'\right) = i_m$, we get:

$$M\left(i_m, t\right) = i_m e^{-\lambda_m\left(t - t'\right)}. \tag{A18}$$

Clearly, as λ_m is always positive, $M\left(t\right)$ decreases exponentially over time and reaches zero as time passes by. A marketer, however, may increase $M\left(t\right)$ by sending an additional set of seeding emails to a list of customers, i.e., a marketer controls the value i_m directly.

Derivation of $E\left(V\left(t\right) | V\left(t'\right) = i_v\right)$:

Let $V\left(i_v, t\right) = E\left(V\left(t\right) | V\left(t'\right) = i_v\right) = \dfrac{\mathrm{d}}{\mathrm{d}s_v} F_{\left(i_m, i_v, i_n\right)}\left(s_m = 1, s_v = 1, s_n = 1, t\right)$. Differentiating (A15) to s_v leads to the following equation:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s_v} F_{\left(i_m, i_v, i_n\right)}\left(s_m, s_v, s_n, t\right) \;=\; & \lambda_m s_n \pi_m \sum_{k=0}^{\infty} k\phi_k s_v^{\ k-1} \frac{\mathrm{d} F_{\left(i_m, i_v, i_n\right)}\left(s_m, s_v, s_n, t\right)}{\mathrm{d}s_m} \\
& + \lambda_m \left(\left(1 - \pi_m\right) + s_n \pi_m \sum_{k=0}^{\infty} \phi_k s_v^{\ k} - s_m\right) \frac{\mathrm{d} F_{\left(i_m, i_v, i_n\right)}\left(s_m, s_v, s_n, t\right)}{\mathrm{d}s_m\, \mathrm{d}s_v} \\
& + \lambda_v \left(s_n \pi_v \sum_{k=0}^{\infty} k\phi_k s_v^{\ k-1} - 1\right) \frac{\mathrm{d} F_{\left(i_m, i_v, i_n\right)}\left(s_m, s_v, s_n, t\right)}{\mathrm{d}s_v} \\
& + \lambda_v \left(\left(1 - \pi_v\right) + s_n \pi_v \sum_{k=0}^{\infty} \phi_k s_v^{\ k} - s_v\right) \frac{\mathrm{d} F_{\left(i_m, i_v, i_n\right)}\left(s_m, s_v, s_n, t\right)}{\mathrm{d}s_v\, \mathrm{d}s_v} \\
& + \sum_{q=1}^{Q} \pi_q \beta_q s_n \sum_{k=0}^{\infty} k\phi_k s_v^{\ k-1} F_{\left(i_m, i_v, i_n\right)}\left(s_m, s_v, s_n, t\right) \\
& + \sum_{q=1}^{Q} \pi_q \beta_q \left(s_n \sum_{k=0}^{\infty} \phi_k s_v^{\ k} - 1\right) \frac{\mathrm{d} F_{\left(i_m, i_v, i_n\right)}\left(s_m, s_v, s_n, t\right)}{\mathrm{d}s_v}
\end{aligned}
\tag{A19}
$$

Setting $s_m = s_v = s_n = 1$ in (A19), and by observing that $\sum_{k=0}^{\infty} \phi_k s_v^{\ k} = 1$, and $\sum_{k=0}^{\infty} k\phi_k s_v^{\ k} = \mu$, i.e. the expected number of forwarded viral emails to friends that did not participate or have not been invited yet. Note that in the paper, $\mu = \mu^*\left(1 - \theta\right)$, where μ^* denotes the average number of

forwarded viral emails, and θ denotes the probability of sending a viral email to a friend that has already received an invitation or that already participated. Furthermore, note that if $s_v = 1$, and ,

$$\frac{\mathrm{d}\, F_{(i_m, i_v, i_n)}\left(s_m = 1, s_v = 1, s_n = 1, t\right)}{\mathrm{d}\, s_m} = M\left(i_m, t\right)$$ we get $V\left(i_v, t\right)$ by solving the following differential

equation:

$$\frac{\mathrm{d}}{\mathrm{d}\,t}V\left(i_v, t\right) = \lambda_m \pi_m \mu \cdot M\left(i_m, t\right) + \lambda_v\left(\pi_v \mu - 1\right)\cdot V\left(i_v, t\right) + \sum_{q=1}^{Q}\pi_q \beta_q \mu, \tag{A20}$$

Using the fact that $V\left(i_v, t'\right) = i_v$, and $M\left(i_m, t\right) = i_m e^{-\lambda_m\left(t-t'\right)}$, we get:

$$V\left(i_v, t\right) = i_v e^{\lambda_v\left(\pi_v \mu - 1\right)\left(t-t'\right)} + K_1\left(e^{\lambda_v\left(\pi_v \mu - 1\right)\left(t-t'\right)} - e^{-\lambda_m\left(t-t'\right)}\right) + K_2\left(e^{\lambda_v\left(\pi_v \mu - 1\right)\left(t-t'\right)} - 1\right), \tag{A21}$$

after solving (A20), with $K_1 = \dfrac{\lambda_m \pi_m \mu i_m}{\lambda_v\left(\pi_v \mu - 1\right) + \lambda_m}$, and $K_2 = \dfrac{\sum_{q=1}^{Q}\pi_q \beta_q \mu}{\lambda_v\left(\pi_v \mu - 1\right)}$. (A21) consists of three

components. The first component, not directly under the marketer's control, depends on the

number of unopened viral emails at $t = t'$, i.e. i_v. These customers may invite new customers by

opening their emails and forwarding it to their friends. When $\pi_v \mu < 1$, this process dies out as

time passes by. The second component depends on the number of unopened seeding emails i_m at

time t' and the subsequent viral process, and is therefore under marketers control. Because

$\left(e^{\lambda_v\left(\pi_v \mu - 1\right)\left(t-t'\right)} - e^{-\lambda_m\left(t-t'\right)}\right)$ goes to zero when t gets very large, the second component goes to zero

as well. The third component is also under marketers control and depends on seeding activities Q

and the subsequent viral process. Interestingly, this component reaches an equilibrium larger

than zero that equals $-K_2$, which is nonnegative when $\pi_v \mu - 1 < 0$. However, when marketers

quit their seeding activities (i.e. $\beta_q = 0$ for all $q \in Q$), K_2 becomes zero, and the process dies

out.

Derivation of $E\left(N(t)\,|\,N(t') = i_n\right)$:

Let $N\left(i_n,t\right)=E\left(N\left(t\right)\mid N\left(t'\right)=i_n\right)=\dfrac{d}{ds_n}F_{\left(i_m,i_v,i_n\right)}\left(s_m=1,s_v=1,s_n=1,t\right)$. Differentiating (A15) to

s_n leads to the following equation:

$$
\begin{aligned}
\frac{d}{dt\,ds_n}F_{\left(i_m,i_v,i_n\right)}\left(s_m,s_v,s_n,t\right) &= \lambda_m\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\,k}\frac{d\,F_{\left(i_m,i_v,i_n\right)}\left(s_m,s_v,s_n,t\right)}{d\,s_m}\\
&+\lambda_m\left(\left(1-\pi_m\right)+s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-s_m\right)\frac{d\,F_{\left(i_m,i_v,i_n\right)}\left(s_m,s_v,s_n,t\right)}{d\,s_m\,d\,s_n}\\
&+\lambda_v\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\,k}\frac{d\,F_{\left(i_m,i_v,i_n\right)}\left(s_m,s_v,s_n,t\right)}{d\,s_v}\\
&+\lambda_v\left(\left(1-\pi_v\right)+s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-s_v\right)\frac{d\,F_{\left(i_m,i_v,i_n\right)}\left(s_m,s_v,s_n,t\right)}{d\,s_v\,d\,s_n}\\
&+\sum_{q=1}^{Q}\pi_q\beta_q\sum_{k=0}^{\infty}\phi_k s_v^{\,k}F_{\left(i_m,i_v,i_n\right)}\left(s_m,s_v,s_n,t\right)\\
&+\sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-1\right)\frac{d\,F_{\left(i_m,i_v,i_n\right)}\left(s_m,s_v,s_n,t\right)}{d\,s_n}
\end{aligned}
\qquad (A22)
$$

Setting $s_m=s_v=s_n=1$ in (A22), and by observing that $\sum_{k=0}^{\infty}\phi_k s_v^{\,k}=1$ if $s_v=1$,

$\dfrac{d\,F_{\left(i_m,i_v,i_n\right)}\left(s_m=1,s_v=1,s_n=1,t\right)}{d\,s_m}=M\left(i_m,t\right)$, and $\dfrac{d\,F_{\left(i_m,i_v,i_n\right)}\left(s_m=1,s_v=1,s_n=1,t\right)}{d\,s_v}=V\left(i_m,t\right)$ we get

$N\left(i_n,t\right)$ by solving the following differential equation:

$$
\frac{d}{dt}N\left(i_n,t\right)=\lambda_m\pi_m\cdot M\left(i_m,t\right)+\lambda_v\pi_v\cdot V\left(i_v,t\right)+\sum_{q=1}^{Q}\pi_q\beta_q\,. \qquad (A23)
$$

Using the fact that $N\left(i_n,t'\right)=i_n$, and the solutions for $M\left(i_m,t\right)$ and $V\left(i_v,t\right)$, we get:

$$
N\left(i_n,t\right)=i_n+K_3\left(e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-1\right)+K_4\left(e^{-\lambda_m\left(t-t'\right)}-1\right)+K_5\left(t-t'\right), \qquad (A24)
$$

with: $K_3=\dfrac{\pi_v}{\left(\pi_v\mu-1\right)}\left(K_1+K_2+i_v\right)$, $K_4=\dfrac{i_m\pi_m\left(\lambda_v-\lambda_m\right)}{\lambda_m+\lambda_v\left(\pi_v\mu-1\right)}$, and $K_5=-\dfrac{\sum_{q=1}^{Q}\pi_q\beta_q}{\pi_v\mu-1}$. Equation (A24)

consists of 4 components. Because the cumulative number of participants in the viral campaign is

strictly increasing, the first component represents the number of participants at time t', i.e.

$N\left(t'\right)=i_n$. The second and third components are a mix of both participants opening seeding and

viral emails, because $V\left(t\right)$ depends on $M\left(t\right)$. When time passes by, these two components do

not generate additional participants and the total number of participants generated by these two processes equals $K_3 + K_4$. As discussed previously, a marketer may directly influence this sum by sending out additional seeding emails to a list of customers. The fourth component increases linearly in time with coefficient K_5, which depends on seeding sources $q \in Q$ and the subsequent viral process. Again, when marketers quit their seeding activities Q, K_5 gets equal to zero and $N(t)$ does not increase further.

WEB APPENDIX B

Derivation of Confidence Intervals using Second-Order Moments of the Viral Branching Process Variables

Web Appendix B describes how to obtain confidence intervals as presented in Figure 6 of the paper. For the derivation of these confidence intervals, we take into account two sources of stochasticity: 1) stochasticity due to parameter uncertainty, and 2) stochasticity due to uncertainty of the viral branching process itself. We solve the first source of stochasticity by simulating repeatedly from the distribution of the estimated parameters, and subsequently computing the expected number of participants $N(t)$ over time. However, the resulting distribution of $N(t)$ underestimates the true variation in the process, because $N(t)$ is stochastic as well. In order to take this stochasticity into account, we derive the second-order moments $M(t)(M(t)-1)$, $V(t)(V(t)-1)$, and $N(t)(N(t)-1)$ of the viral branching process variables which we denote respectively by $M_2(t)$, $V_2(t)$, and $N_2(t)$. Using these second order moments, we are able to derive the variance of the number of participants in the viral campaign which equals:

$$\text{var}(N(t)) = N_2(t) + N(t) - N(t)^2. \tag{B1}$$

Because of the large number of participants in viral marketing campaigns, we apply the Central Limit Theorem, which states that the distribution of the number of participants at time t is approximately normal with mean $N(t)$ and variance $N_2(t)$.

Using the above procedure, we simulate the distribution of $N(t)$ by repeatedly executing the following steps[10]:

Step 1) Randomly draw each of the parameters from their estimated distributions.

Step 2) Using the parameter draws from Step 1), compute the expected mean and variance of the process variable $N(t)$.

Step 3) Draw $N(t)$ from a normal distribution with mean and variance as computed in Step 2).

---

[10] In the empirical application we used 20,000 draws to simulate the 95 percent prediction intervals in Figure 6.

Using the draws generated in Step 3, it is straightforward to compute confidence intervals as we presented in Figure 6 of our paper. However, to execute these three steps repeatedly, we need a closed-form expression of the second-order moment of $N_2(t)$, which we derive next.

Second-order moments

Similar to the first-order moments in Web Appendix A, we derive the second-order moments using the differential equation of the probability generating function (A15). Using the notation in Web Appendix A, the second-order moment of the process $N(t)$ can be computed as follows:

$$N_2(i_n,t) = E\big(N(t)\big(N(t)-1\big)\,|\,N(t')=i_n\big) = \frac{\mathrm{d}}{\mathrm{d}\,s_n\,\mathrm{d}\,s_n} F_{(i_m,i_v,i_n)}\big(s_m=1, s_v=1, s_n=1, t\big). \text{ Using (A15)}$$

and (A22), we get

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t\,\mathrm{d}\,s_n\,\mathrm{d}\,s_n} F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big) \;=\;\; & 2\lambda_m\pi_m\sum_{k=0}^{\infty}\phi_k s_v{}^k \frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_m\,\mathrm{d}\,s_n} \\[4pt]
& + \lambda_m\left(\big(1-\pi_m\big)+s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v{}^k - s_m\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_m\,\mathrm{d}\,s_n\,\mathrm{d}\,s_n} \\[4pt]
& + 2\lambda_v\pi_v\sum_{k=0}^{\infty}\phi_k s_v{}^k \frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_v\,\mathrm{d}\,s_n} \\[4pt]
& + \lambda_v\left(\big(1-\pi_v\big)+s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v{}^k - s_v\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_v\,\mathrm{d}\,s_n\,\mathrm{d}\,s_n} \\[4pt]
& + 2\sum_{q=1}^{Q}\pi_q\beta_q\sum_{k=0}^{\infty}\phi_k s_v{}^k \frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_n} \\[4pt]
& + \sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v{}^k - 1\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_n\,\mathrm{d}\,s_n}
\end{aligned}
\tag{B1}
$$

Note that (B1) depends, among others, on $\dfrac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_m\,\mathrm{d}\,s_n}$, which equals

$E\big(M(t)N(t)\,|\,M(t')=i_m, N(t')=i_n\big)$ and represents the interaction between process variables $M(t)$ and $N(t)$. Hence, to derive second-order moment of the process $N(t)$, we also need to derive the second-order moments of the other processes, $M(t)$ and $V(t)$, and its interactions. We first derive $M_2(t)$ next.

Derivation of $E\big(M(t)\big(M(t)-1\big)|M(t')=i_m\big)$:

Let, $M_2(i_m,t)=E\big(M(t)\big(M(t)-1\big)|M(t')=i_m\big)=\dfrac{d}{ds_m ds_m}F_{(i_m,i_v,i_n)}\big(s_m=1,s_v=1,s_n=1,t\big).$

Differentiating (A16) to s_m leads to the following equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s_m\,\mathrm{d}s_m}F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)\;=\;-2\lambda_m\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_m\,\mathrm{d}\,s_m}$$

$$+\lambda_m\left((1-\pi_m)+s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-s_m\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_m\,\mathrm{d}\,s_m\,\mathrm{d}\,s_m}$$

$$+\lambda_v\left((1-\pi_v)+s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-s_v\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_v\,\mathrm{d}\,s_m\,\mathrm{d}\,s_m}\qquad(\mathrm{B}2)$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-1\right)\frac{F_{(i_m,i_v,i_n)}\big(s_m,s_v,s_n,t\big)}{\mathrm{d}\,s_m\,\mathrm{d}\,s_m}$$

Setting $s_m=s_v=s_n=1$ in (B2), and by observing that $\displaystyle\sum_{k=0}^{\infty}\phi_k s_v^{\,k}=1$ if $s_v=1$, we get $M_2(i_m,t)$ by

solving the following differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}M_2(i_m,t)=-\lambda_m\cdot M_2(i_m,t).\qquad(\mathrm{B}3)$$

Using the fact that $M(i_m,t')=i_m$, so that $M_2(i_m,t')=i_m(i_m-1)$ we get:

$$M_2(i_m,t)=i_m(i_m-1)e^{-2\lambda_m(t-t')}.\qquad(\mathrm{B}4)$$

Derivation of $E\big(V(t)\big(V(t)-1\big)|V(t')=i_v\big)$:

To derive the second-order moment of $V(t)$, we first need to have an expression for

$E\big(M(t)\cdot V(t)|M(t')=i_m,V(t')=i_v\big)$, as $E\big(V(t)\big(V(t)-1\big)|V(t')=i_v\big)$ depends on this. Let

$$MV(i_m,i_v,t)=E\big(M(t)\cdot V(t)|M(t')=i_m,V(t')=i_v\big)=\frac{d}{ds_m ds_v}F_{(i_m,i_v,i_n)}\big(s_m=1,s_v=1,s_n=1,t\big).$$

Differentiating (A16) to s_v, we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s_m\,\mathrm{d}s_v}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right) = -\lambda_m\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_v}$$

$$+\lambda_m s_n\pi_m\sum_{k=0}^{\infty}\phi_k ks_v^{\,k-1}\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_m}$$

$$+\lambda_m\left(\left(1-\pi_m\right)+s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-s_m\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_m\,\mathrm{d}s_v}$$

$$+\lambda_v\left(s_n\pi_v\sum_{k=0}^{\infty}\phi_k ks_v^{\,k-1}-1\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_m} \qquad (\text{B5})$$

$$+\lambda_v\left(\left(1-\pi_v\right)+s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-s_v\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_m\,\mathrm{d}s_v}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q s_n\sum_{k=0}^{\infty}\phi_k ks_v^{\,k-1}\frac{F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-1\right)\frac{F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_v}$$

Setting $s_m=s_v=s_n=1$ in (B5), and by observing that $\sum_{k=0}^{\infty}\phi_k s_v^{\,k}=1$, and $\sum_{k=0}^{\infty}k\phi_k s_v^{\,k}=\mu$, we get

the following differential equation for $MV\left(i_m,i_v,t\right)$:

$$\frac{d}{dt}MV\left(i_m,i_v,t\right)=\left(\lambda_v\left(\pi_v\mu-1\right)-\lambda_m\right)\cdot MV\left(i_m,i_v,t\right)+\lambda_m\pi_m\mu M_2\left(i_m,t\right)+\sum_{q=1}^{Q}\pi_q\beta_q\mu M\left(i_m,t\right).\quad(\text{B6})$$

Using (A18), (B4) and the fact that $MV=\left(i_m,i_v,t'\right)=i_m i_v$ to solve (B6), we get:

$$MV\left(i_m,i_v,t\right) = K_6\left(e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}e^{-\lambda_m\left(t-t'\right)}-e^{-2\lambda_m\left(t-t'\right)}\right)+K_7\left(e^{\lambda_v\left(\pi_v\mu-1\right)t}e^{-\lambda_m\left(t-t'\right)}-e^{-\lambda_m\left(t-t'\right)}\right),\quad(\text{B7})$$

$$+i_m i_v e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}e^{-\lambda_m\left(t-t'\right)}$$

with $K_6=\dfrac{\lambda_m\pi_m\mu i_m\left(i_m-1\right)}{\lambda_v\left(\pi_v\mu-1\right)+\lambda_m}$, and $K_7=\dfrac{\sum_{q=1}^{Q}\pi_q\beta_q\mu i_m}{\lambda_v\left(\pi_v\mu-1\right)}$.

Using (B7), we are able to derive the second-order moment

$V_2\left(i_v,t\right)=E\left(V\left(t\right)\left(V\left(t\right)-1\right)\mid V\left(0\right)=i_v\right)=\dfrac{d}{ds_v ds_v}F_{(i_m,i_v,i_n)}\left(s_m=1,s_v=1,s_n=1,t\right)$. Differentiating

(A19) to s_v, we get the following differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s_v\,\mathrm{d}s_v}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right) = \lambda_m s_n \pi_m \sum_{k=0}^{\infty} k(k-1)\phi_k {s_v}^{k-2}\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m}$$

$$+2\lambda_m s_n \pi_m \sum_{k=0}^{\infty} k\phi_k {s_v}^{k-1}\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_v}$$

$$+\lambda_m\left((1-\pi_m)+s_n\pi_m\sum_{k=0}^{\infty}\phi_k {s_v}^{k}-s_m\right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_v\,\mathrm{d}s_v}$$

$$+\lambda_v s_n \pi_v \sum_{k=0}^{\infty} k(k-1)\phi_k {s_v}^{k-2}\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v}$$

$$+\lambda_v\left(s_n\pi_v\sum_{k=0}^{\infty} k\phi_k {s_v}^{k-1}-1\right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_v}$$

$$+\lambda_v\left(s_n\pi_v\sum_{k=0}^{\infty}\phi_k k{s_v}^{k-1}-1\right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_v} \qquad . \quad (B8)$$

$$+\lambda_v\left((1-\pi_v)+s_n\pi_v\sum_{k=0}^{\infty}\phi_k {s_v}^{k}-s_v\right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_v\,\mathrm{d}s_v}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q s_n \sum_{k=0}^{\infty} k(k-1)\phi_k {s_v}^{k-2}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q s_n \sum_{k=0}^{\infty} k\phi_k {s_v}^{k-1}\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q s_n \sum_{k=0}^{\infty} \phi_k k{s_v}^{k-1}\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k {s_v}^{k}-1\right)\frac{\mathrm{d}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_v}$$

Setting $s_m = s_v = s_n = 1$ in (B8), and by observing that $\sum_{k=0}^{\infty}\phi_k {s_v}^{k}=1$, $\sum_{k=0}^{\infty} k\phi_k {s_v}^{k}=\mu$, and

$\sum_{k=0}^{\infty} k(k-1)\phi_k {s_v}^{k-2}=\mu_2$, where μ_2 is the second-order moment of forwarded viral emails to

friends that did not participate or have not been invited yet[11], we get $V_2\left(i_v,t\right)$ by solving the

following differential equation:

---

[11] Similar to the other parameters of the viral branching process, we estimate μ_2 directly from the individual-level

data that readily comes available during a viral marketing campaign, i.e. $\mu_2 = \frac{1}{n_d}\sum_{d=1}^{D}\sum_{c=1}^{n_d} x_{cd}\left(x_{cd}-1\right)$.

$$\frac{d}{dt}V_2\left(i_v,t\right) \;=\; K_8 \cdot V\left(i_v,t\right)+\lambda_m\pi_m\mu_2 \cdot M\left(i_m,t\right)+2\lambda_m\pi_m\mu \cdot MV\left(i_m,i_v,t\right)$$

$$+2\lambda_v\left(\pi_v\mu-1\right)V_2\left(i_v,t\right)+\sum_{q=1}^{Q}\pi_q\beta_q\mu_2 \;,\qquad (B9)$$

with $K_8 = \left(\lambda_v\pi_v\mu_2+2\sum_{q=1}^{Q}\pi_q\beta_q\mu\right)$. Using (A18), (A21), (B7) and $V_2\left(t'\right)=i_v\left(i_v-1\right)$, we get:

$$V_2\left(i_v,t\right) \;=\; K_9\left(e^{2\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}\right)+K_{10}\left(e^{2\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-e^{-\lambda_m\left(t-t'\right)}\right)$$

$$+K_{11}\left(e^{2\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-e^{-2\lambda_m\left(t-t'\right)}\right)+K_{12}\left(e^{2\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}e^{-\lambda_m\left(t-t'\right)}\right),\;\;(B10)$$

$$+K_{13}\left(e^{2\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-1\right)+i_v\left(i_v-1\right)e^{2\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}$$

with $K_9 = \dfrac{K_8\left(K_1+K_2+i_v\right)}{\lambda_v\left(\pi_v\mu-1\right)}$, $K_{10} = \dfrac{\lambda_m\pi_m\left(\mu_2 i_m-2\mu K_7\right)-K_1 K_8}{2\lambda_v\left(\pi_v\mu-1\right)+\lambda_m}$, $K_{11} = -\dfrac{2\lambda_m\pi_m\mu K_6}{2\lambda_v\left(\pi_v\mu-1\right)+2\lambda_m}$,

$$K_{12} = \frac{2\lambda_m\pi_m\mu\left(K_6+K_7+i_m i_v\right)}{\lambda_v\left(\pi_v\mu-1\right)+\lambda_m}, \text{ and } K_{13} = \frac{\sum_{q=1}^{Q}\pi_q\beta_q\mu_2-K_2 K_8}{2\lambda_v\left(\pi_v\mu-1\right)}.$$

Derivation of $E\left(N\left(t\right)\left(N\left(t\right)-1\right)\mid N\left(t'\right)=i_n\right)$:

To derive the second-order moment of $N\left(t\right)$, we need, next to (B4), (B7) and (B10), expressions

for $E\left(M\left(t\right)N\left(t\right)\mid M\left(t'\right)=i_m,N\left(t'\right)=i_n\right)$, and $E\left(V\left(t\right)N\left(t\right)\mid V\left(t'\right)=i_v,N\left(t'\right)=i_n\right)$, as

$E\left(N\left(t\right)\left(N\left(t\right)-1\right)\mid N\left(t'\right)=i_n\right)$ depends on it. Let

$$MN\left(i_m,i_n,t\right)=E\left(M\left(t\right)N\left(t\right)\mid M\left(t'\right)=i_m,N\left(t'\right)=i_n\right)=\frac{d}{ds_m ds_n}F_{\left(i_m,i_v,i_n\right)}\left(s_m=1,s_v=1,s_n=1,t\right).$$

Differentiating (A16) to s_n, we get:

$$\frac{d}{dt}\frac{d}{ds_m\,ds_n}F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t) = -\lambda_m\frac{d\,F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{ds_m\,ds_n}$$

$$+\lambda_m\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\ k}\frac{d\,F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{ds_m\,ds_m}$$

$$+\lambda_m\left((1-\pi_m)+s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\ k}-s_m\right)\frac{d\,F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{ds_m\,ds_m\,ds_n}$$

$$+\lambda_v\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\ k}\frac{d\,F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{ds_v\,ds_m} \qquad . \qquad \text{B}(11)$$

$$+\lambda_v\left((1-\pi_v)+s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\ k}-s_v\right)\frac{d\,F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{ds_v\,ds_m\,ds_n}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\sum_{k=0}^{\infty}\phi_k s_v^{\ k}\frac{F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{ds_m}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v^{\ k}-1\right)\frac{F_{(i_m,i_v,i_n)}(s_m,s_v,s_n,t)}{ds_m\,ds_n}$$

Setting $s_m=s_v=s_n=1$ in (B11), and by observing that $\sum_{k=0}^{\infty}\phi_k s_v^{\ k}=1$ and $\sum_{k=0}^{\infty}k\phi_k s_v^{\ k}=\mu$, we get

$MN(i_m,i_n,t)$ by solving the following differential equation:

$$\frac{d}{dt}MN(i_m,i_n,t) = \lambda_v\pi_v\cdot MV(i_m,i_v,t)+\lambda_m\pi_m\cdot M_2(i_m,t)+\sum_{q=1}^{Q}\pi_q\beta_q\cdot M(i_m,t)$$
$$-\lambda_m\cdot MN(i_m,i_n,t) \qquad (B10)$$

Using (A18), (B4), (B7) and the fact that $MN(i_m,i_n,t')=i_m i_n$, we get:

$$MN(i_m,i_n,t) = K_{14}(t-t')e^{-\lambda_m(t-t')}+K_{15}\left(e^{-\lambda_m(t-t')}-e^{-2\lambda_m(t-t')}\right)$$
$$+K_{16}\left(e^{\lambda_v(\pi_v\mu-1)(t-t')}e^{-\lambda_m(t-t')}-e^{-\lambda_m(t-t')}\right)+i_m i_n e^{-\lambda_m(t-t')} \qquad (B12)$$

with $K_{14}=\sum_{q=1}^{Q}\pi_q\beta_q i_m-\lambda_v\pi_v K_7$, $K_{15}=\dfrac{\lambda_m\pi_m i_m(i_m-1)-\lambda_v\pi_v K_6}{\lambda_m}$, and $K_{16}=\dfrac{\lambda_v\pi_v(K_6+K_7+i_m i_v)}{\lambda_v(\pi_v\mu-1)}$.

Let $VN(i_v,i_n,t)=E\left(V(t)N(t)\,|\,V(t')=i_v,N(t')=i_n\right)=\dfrac{d}{ds_v ds_n}F_{(i_m,i_v,i_n)}(s_m=1,s_v=1,s_n=1,t)$.

Differentiating (A19) to s_n, we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}}{\mathrm{d}s_v\,\mathrm{d}s_n}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right) = \lambda_m\pi_m\sum_{k=0}^{\infty}k\phi_k s_v^{\,k-1}\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m}$$

$$+\lambda_m s_n\pi_m\sum_{k=0}^{\infty}k\phi_k s_v^{\,k-1}\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_n}$$

$$+\lambda_m\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\,k}\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_v}$$

$$+\lambda_m\left((1-\pi_m)+s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-s_m\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_m\,\mathrm{d}s_v\,\mathrm{d}s_n}$$

$$+\lambda_v\pi_v\sum_{k=0}^{\infty}k\phi_k s_v^{\,k-1}\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v}$$

$$+\lambda_v\left(s_n\pi_v\sum_{k=0}^{\infty}k\phi_k s_v^{\,k-1}-1\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_n}\qquad.\quad\text{(B13)}$$

$$+\lambda_v\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\,k}\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_v}$$

$$+\lambda_v\left((1-\pi_v)+s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-s_v\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_v\,\mathrm{d}s_n}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\sum_{k=0}^{\infty}k\phi_k s_v^{\,k-1}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q s_n\sum_{k=0}^{\infty}k\phi_k s_v^{\,k-1}\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_n}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\sum_{k=0}^{\infty}\phi_k s_v^{\,k}\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v^{\,k}-1\right)\frac{\mathrm{d}\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{\mathrm{d}s_v\,\mathrm{d}s_n}$$

Setting $s_m=s_v=s_n=1$ in (B13), and by observing that $\sum_{k=0}^{\infty}\phi_k s_v^{\,k}=1$ and $\sum_{k=0}^{\infty}k\phi_k s_v^{\,k}=\mu$, we get

$VN\left(i_v,i_n,t\right)$ by solving the following differential equation:

$$\frac{d}{dt}VN\left(i_v,i_n,t\right) = \left(\lambda_v\pi_v\mu+\sum_{q=1}^{Q}\pi_q\beta_q\right)V\left(i_v,t\right)+\lambda_m\pi_m\mu\cdot M\left(i_m,t\right)+\sum_{q=1}^{Q}\pi_q\beta_q\mu\cdot N\left(i_n,t\right)+$$

$$\lambda_m\pi_m\cdot MV\left(i_m,i_v,t\right)+\lambda_m\pi_m\mu MN\left(i_m,i_n,t\right)+\lambda_v\pi_v\cdot V_2\left(i_v,t\right)+\sum_{q=1}^{Q}\pi_q\beta_q\mu+.$$

$$\lambda_v\left(\pi_v\mu-1\right)\cdot VN\left(i_v,i_n,t\right)$$

(B14)

Using (A18), (A21), (A24), (B7), (B10), (B12) and the fact that $VN(i_v, i_n, t') = i_v i_n$, we get:

$$
\begin{aligned}
VN(i_v, i_n, t) &= K_{17}(t-t') e^{\lambda_v(\pi_v \mu - 1)(t-t')} + K_{18}(t-t') e^{-\lambda_m t} + K_{19}\left(e^{\lambda_v(\pi_v \mu - 1)(t-t')} - e^{-\lambda_m(t-t')}\right) \\
&\quad + K_{20}\left(e^{\lambda_v(\pi_v \mu - 1)(t-t')} - e^{-2\lambda_m(t-t')}\right) + K_{21}\left(e^{\lambda_v(\pi_v \mu - 1)(t-t')} - e^{\lambda_v(\pi_v \mu - 1)(t-t')} e^{-\lambda_m t}\right) \\
&\quad + K_{22}\left(e^{2\lambda_v(\pi_v \mu - 1)(t-t')} - e^{\lambda_v(\pi_v \mu - 1)(t-t')}\right) + K_{23}\left(e^{\lambda_v(\pi_v \mu - 1)(t-t')} - 1\right) + K_{24}(t-t') \\
&\quad + i_v i_n e^{\lambda_v(\pi_v \mu - 1)(t-t')}
\end{aligned} \quad , \quad (B15)
$$

with $K_{17} = \left(\lambda_v \pi_v \mu + \sum_{q=1}^{Q} \pi_q \beta_q\right)(K_1 + K_2 + i_v) + \sum_{q=1}^{Q} \pi_q \beta_q \mu K_3 - \lambda_v \pi_v K_9$, $K_{18} = -\dfrac{\lambda_m \pi_m \mu K_{14}}{\lambda_v(\pi_v \mu - 1) + \lambda_m}$,

$$
K_{19} = \frac{\lambda_m \pi_m \mu\left(K_{15} - K_{16} + i_m i_n + i_m\right) - \left(\lambda_v \pi_v \mu + \sum_{q=1}^{Q} \pi_q \beta_q\right) K_1 + \sum_{q=1}^{Q} \pi_q \beta_q \mu K_4 - \lambda_m \pi_m K_7 - \lambda_v \pi_v K_{10} - K_{18}}{\lambda_v(\pi_v \mu - 1) + \lambda_m}
$$

$K_{20} = -\dfrac{\lambda_m \pi_m\left(K_6 + \mu K_{15}\right) + \lambda_v \pi_v K_{11}}{\lambda_v(\pi_v \mu - 1) + 2\lambda_m}$, $\quad K_{21} = \dfrac{\lambda_m \pi_m\left(K_6 + K_7 + i_m i_v + \mu K_{16}\right) - \lambda_v \pi_v K_{12}}{\lambda_m}$,

$K_{22} = \dfrac{\lambda_v \pi_v\left(K_9 + K_{10} + K_{11} + K_{12} + K_{13} + i_v(i_v - 1)\right)}{\lambda_v(\pi_v \mu - 1)}$,

$K_{23} = \dfrac{\sum_{q=1}^{Q} \pi_q \beta_q \mu - \left(\lambda_v \pi_v \mu + \sum_{q=1}^{Q} \pi_q \beta_q\right) K_2 + \sum_{q=1}^{Q} \pi_q \beta_q \mu\left(i_x - K_3 - K_4\right) - \lambda_v \pi_v K_{13} - K_{24}}{\lambda_v(\pi_v \mu - 1)}$, and

$K_{24} = -\dfrac{\sum_{q=1}^{Q} \pi_q \beta_q \mu K_5}{\lambda_v(\pi_v \mu - 1)}$.

Let $N_2(i_n, t) = E\left(N(t)(N(t) - 1) \mid N(t') = i_n\right) = \dfrac{d}{ds_n ds_n} F_{(i_m, i_v, i_n)}(s_m = 1, s_v = 1, s_n = 1, t)$.

Differentiating (A22) to s_n, we get:

$$\frac{d}{dt\,ds_n\,ds_n}F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right) = 2\lambda_m\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\ k}\frac{d\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{d\,s_m\,d\,s_n}$$

$$+\lambda_m\left(\left(1-\pi_m\right)+s_n\pi_m\sum_{k=0}^{\infty}\phi_k s_v^{\ k}-s_m\right)\frac{d\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{d\,s_m\,d\,s_n\,d\,s_n}$$

$$+2\lambda_v\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\ k}\frac{d\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{d\,s_v\,d\,s_n}$$

$$+\lambda_v\left(\left(1-\pi_v\right)+s_n\pi_v\sum_{k=0}^{\infty}\phi_k s_v^{\ k}-s_v\right)\frac{d\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{d\,s_v\,d\,s_n\,d\,s_n} \qquad (B16)$$

$$+2\sum_{q=1}^{Q}\pi_q\beta_q\sum_{k=0}^{\infty}\phi_k s_v^{\ k}\frac{d\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{d\,s_n}$$

$$+\sum_{q=1}^{Q}\pi_q\beta_q\left(s_n\sum_{k=0}^{\infty}\phi_k s_v^{\ k}-1\right)\frac{d\,F_{(i_m,i_v,i_n)}\left(s_m,s_v,s_n,t\right)}{d\,s_n\,d\,s_n}$$

Setting $s_m=s_v=s_n=1$ in (B16), and by observing that $\sum_{k=0}^{\infty}\phi_k s_v^{\ k}=1$ we get $N_2\left(i_n,t\right)$ by solving the following differential equation:

$$\frac{d}{dt}N_2\left(i_n,t\right)=2\lambda_v\pi_v\cdot VN\left(i_v,i_n,t\right)+2\lambda_m\pi_m\cdot MN\left(i_v,i_n,t\right)+2\sum_{q=1}^{Q}\pi_q\beta_q\cdot N\left(i_n,t\right). \qquad (B17)$$

Using (B12), (B15) and the fact that $N_2\left(i_n,t'\right)=i_n\left(i_n-1\right)$, we get:

$$N_2\left(i_n,t\right) = i_n\left(i_n-1\right)+K_{25}\left(t-t'\right)e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}+K_{26}\left(t-t'\right)e^{-\lambda_m\left(t-t'\right)}+K_{27}\left(1-e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}\right)$$

$$+K_{28}\left(1-e^{-\lambda_m\left(t-t'\right)}\right)+K_{29}\left(1-e^{-2\lambda_m\left(t-t'\right)}\right)+K_{30}\left(e^{\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}e^{-\lambda_m\left(t-t'\right)}-1\right) \qquad , (B18)$$

$$+K_{31}\left(e^{2\lambda_v\left(\pi_v\mu-1\right)\left(t-t'\right)}-1\right)+K_{32}\left(t-t'\right)^2+K_{33}\left(t-t'\right)$$

with $K_{25}=\dfrac{2\lambda_v\pi_v K_{17}}{\left(\lambda_v\pi_v\mu-1\right)}$, $K_{26}=-\dfrac{2\lambda_v\pi_v K_{18}+2\lambda_m\pi_m K_{14}}{\lambda_m}$,

$$K_{27}=-\frac{2\lambda_v\pi_v\left(K_{19}+K_{20}+K_{21}-K_{22}+K_{23}+i_v i_n\right)+2\sum_{q=1}^{Q}\pi_q\beta_q K_3-K_{25}}{\lambda_v\left(\pi_v\mu-1\right)},$$

$$K_{28}=\frac{2\lambda_m\pi_m\left(K_{15}-K_{16}+i_m i_n\right)+2\sum_{q=1}^{Q}\pi_q\beta_q K_4-2\lambda_v\pi_v K_{19}-K_{26}}{\lambda_m}, \quad K_{29}=-\frac{\lambda_v\pi_v K_{20}+\lambda_m\pi_m K_{15}}{\lambda_m},$$

$$K_{30} = \frac{2\lambda_m \pi_m K_{16} - 2\lambda_v \pi_v K_{21}}{\lambda_v (\pi_v \mu - 1) - \lambda_m}, \quad K_{31} = \frac{\pi_v K_{22}}{\pi_v \mu - 1}, \quad K_{32} = \lambda_v \pi_v K_{24} + \sum_{q=1}^{Q} \pi_q \beta_q K_5, \text{ and}$$

$$K_{33} = 2\sum_{q=1}^{Q} \pi_q \beta_q (i_n - K_3 - K_4) - 2\lambda_v \pi_v K_{23}.$$

References

Athreya, K. B. and P. E. Ney (1972), *Branching Processes*. Berlin: Springer-Verlag.

Harris, T. E. (1963), *The Theory of Branching Processes*. Berlin: Springer-Verlag.

Ross, S. M. (1997), *Introduction to Probability Models*. San Diego, CA: Academic Press.

Publications in the Report Series Research* in Management

ERIM Research Program: "Marketing"

2009

Map Based Visualization of Product Catalogs
Martijn Kagie, Michiel van Wezel, and Patrick J.F. Groenen
ERS-2009-010-MKT
http://hdl.handle.net/1765/15142

Embedding the Organizational Culture Profile into Schwartz's Universal Value Theory using Multidimensional Scaling with Regional Restrictions
Ingwer Borg, Patrick J.F. Groenen, Karen A. Jehn, Wolfgang Bilsky, and Shalom H. Schwartz
ERS-2009-017-MKT
http://hdl.handle.net/1765/15404

Determination of Attribute Weights for Recommender Systems Based on Product Popularity
Martijn Kagie, Michiel van Wezel, and Patrick J.F. Groenen
ERS-2009-022-MKT
http://hdl.handle.net/1765/15910

An Empirical Comparison of Dissimilarity Measures for Recommender Systems
Martijn Kagie, Michiel van Wezel, and Patrick J.F. Groenen
ERS-2009-023-MKT
http://hdl.handle.net/1765/15911

A Viral Branching Model for Predicting the Spread of Electronic Word-of-Mouth
Ralf van der Lans, Gerrit van Bruggen, Jehoshua Eliashberg, Berend Wierenga
ERS-2009-029-MKT
http://hdl.handle.net/1765/16015