Viral coefficient - Unveiling the Holy Grail of online marketing

Joni Salminen, Aarni Hytonen

We present an evolved version of Chen's (2008) viral coefficient. By doing so, we tackle some shortcomings of the simplified viral coefficient model: (1) lack of time component, (2) lack of carrying capacity, and (3) duplicate problem. Our model is based on invitations sent by individuals to one another in a finite social network and can be used to determine the viral growth of visitors in a website at a specific time. By using logistic function and basic theory of finance, the model focuses on website visits that are easily measurable, discrete events, and constitute the starting point for more advanced type of conversions, such as sales.

Keywords: viral marketing, viral coefficient, internet marketing

Introduction

Viral marketing[1], or "going viral", has long been the goal of internet marketers. Finding a formula that would make viral diffusion replicable has been described as the "Holy Grail" of online marketing (Hood, 2012). Essentially, viral marketing is a variant of word of mouth or peer marketing, in which consumers disseminate commercial messages on behalf of the company. Ideal characteristics of successful viral marketing include self-replication, voluntary dissemination with little control by the firm, and use of social networks[2] (Leskovec, Adamic, & Huberman, 2005). Although the phenomenon of word of mouth marketing is not novel (Dichter, 1966), the potential of diffusion rate is critically higher in the Internet than offline (Cheung & Thadani, 2010), as its small world characteristics make it possible to reach any individual through relatively few connections (Schnettler, 2009). Despite the hardship relating to achieving a replicable model for viral growth, there are several documented examples of such a growth — e.g. Hotmail.com, Applifier (55M users in three months) and Rock You's Horoscope application (1.5M users in 15 days). For example, the founder of One List[3] describes the company's viral growth as follows [8]:

> *Like, with ONElist, the grand total of all the advertising I ever did for that company was I spammed some guy who had posted to Usenet looking for a mailing list provider. And he was in Norway; this was on a Saturday evening in January of '98, and I just said, 'Hey, try my service.' The next day, I wake up, and not only had he created a list, ten of his friends had created lists. We had hundreds of users, just within the span of a few hours and one email. After 11 months we had a million users. Just from that.*

The need for exact and useful metrics of viral marketing has been recognized by many authors (e.g., Richardson & Domingos, 2002; Dellarocas & Narayan, 2006) The question we ask in this paper is: *how to apply the viral coefficient to model diffusion of peer-marketing messages in a finite social network?* We approach this question by improving a basic model measuring the growth of number of visitors resulting from invitations by other visitors of a website.

Three models of viral coefficient

In the following part, we will examine three models of viral growth. We will start from a discrete model used by practitioners, and then present a more advanced, continuous model by the Internet marketing professional Andrew Chen (2008), and finally our own model developed aimed at improving the other models.

a. The base model

[1]"Diffusion of information about the product and its adoption over the network" (Leskovec et al., 2005).

[2]Making viral diffusion, ultimately, a social process (Rogers, 1995).

[3]A Web application to manage email newsletters.

The base model is commonly used by practitioners of Internet marketing to understand viral marketing. The base model is (Tokuda, 2008):

$$v = x * y$$

$$\begin{aligned} where\ v &= viral\ growth\ factor \\ y &= efficiency\ of\ the\ viral\ loop \\ x &= number\ of\ persons\ a\ user\ invites \end{aligned}$$

In other words, the model measures the number of new customers the average existing customer generates through invitations. The process is depicted in the following figure.

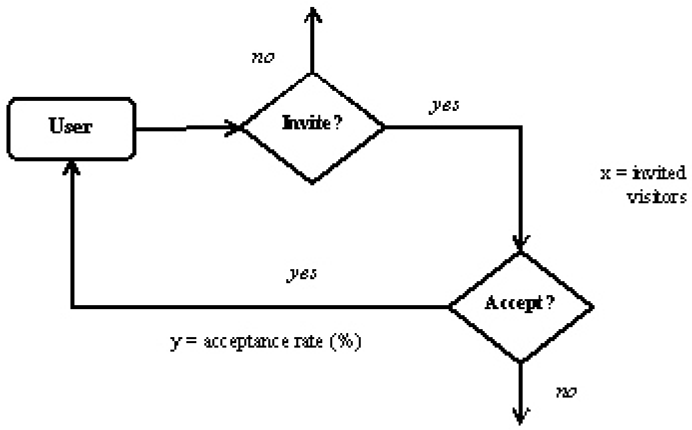

Figure 1. Process of viral marketing (Tokuda 2008).

According to this model[4], one achieves viral growth if

$$x * y > 1,$$

i.e. earlier invited visitors always invite more than one new visitor[5] and always more than one of them accepts the invitation[6]. This mechanism for viral growth satisfies the requirements of self-replicating growth defined in the epidemiological studies of viral diffusion (Khelil, Becker, Tian, & Rothermel, 2002). However, as it is a circular model it does not tell much of the dynamics between factors influencing viral growth. More specifically, the major problem is the *lack of time value*. Many viral messages "jump the shark", becoming quickly outdated (Chen, 2008). For example news items expire within a day. As such, there are three possibilities for viral growth in terms of time (see Figure 2): 1) *expiration* where viral coefficient diminishes and growth stops; 2) *stability*, where growth factor remains constantly over one; and 3) *inflation* which is a consequence of network effect or increase in topicality. For viral messages whose topicality is high, an increase in time equals a drop in viral coefficient — that is, propensity to send and accept invitations. This may lead into dramatic drop in viral growth, where the cumulated number of visitors will plateau — depicted as case B in graph. Consequently, one can argue that when topicality is high, the viral coefficient is high and *vice versa*. Consider also the cumulative growth of visitors: if expiration is rapid, a long term strategy with a low viral coefficient may bring better results than short-term

[4]Note that the model is essentially compatible with models of communication where central entities are sender, receiver, channel and message (e.g., Shannon, Weaver, & Shannon, 1998).

[5]The decision of inviting others is a function of the so-called viral loop, i.e. "the steps a user goes through between entering the site to inviting the next set of new users" (Chen 2008).

[6]Acceptance is influenced by perception of interestingness, which is sometimes referred to as viral hook. The invitation itself can relate to many things, such as visiting a website, downloading or installing an application, or other incremental action.

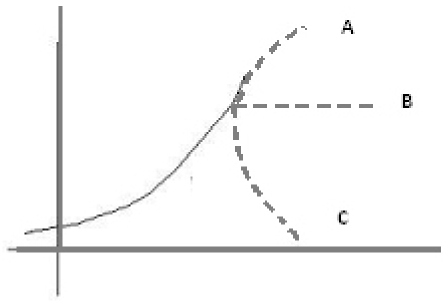

Figure 2. Post-viral: What happens after growth?

campaigns[7]. Hence, examining viral growth factor at a specific point in time is always susceptible to error of predictive accuracy. Marketing researchers have examined the phenomenon of fatigue, or maturity in product life cycle models (see Rink & Swan, 1979 for a dated but comprehensive review). However, the major difference is that product lifecycle changes take place during longer periods of time, typically years, whereas major fluctuation for viral diffusion may take place within intra-day periods.

An example of case A, which we define as viral growth through network effects, is a Web service called Dropbox that allows visitors to store their files in the "cloud" where they are accessible from anywhere. The firm offers what we refer to as *dual-sided referral incentive* which promises both the referral visitor and the invitee additional benefit — that is, additional storage space — upon accepting an invitation. Hence, the more people invited, the more one can accumulate additional storage space, up to a certain threshold. This becomes a powerful incentive for to invite as many prospects as possible, and subsequently for them to accept the invitation. In spite of potential enhancing influence of network effects to viral growth, their existence typically sets a requirement of other visitors[8] — consider for example, Google and Facebook. Having additional users of Google search does not increase one's benefits of use; in contrast, Facebook is more valuable in use the more one has friends there[9], and in fact is quite pointless to use alone.

The cumulative growth curve can never turn to negative slope when examining visitors or number of visit — that is, case C is impossible in our model. This is because a visit is a discrete, irreversible event that cannot be undone. In contrast, case C is possible when replacing 'visitor' with 'customer', as customers can become non-customers at any point of their lifetime. Therefore, C demonstrates the fundamental difference of measuring visitors and customers, or our model and lifetime models. It would make sense to expect that visitors and customers behave differently in disseminating viral messages since becoming a customer can be seen as a sign of higher engagement — therefore, it is invaluable to make the distinction. However, it is not wise to assume that only customers spread viral messages, but non-paying visitors do so as well. Further, the concepts user and customer are often used interchangeably, so that the question "who is the customer?" seems at times elusive in online business[10]. Yet, in our model the distinction is important, if nothing more than for the sake of mathematics.

b. Chen's model

Acknowledging the shortcomings of the base model, the Internet marketer Andrew Chen has

[7]This can be seen e.g. in the lifetime value of blog articles; even when they are not shared frequently, because the content remains searchable and linkable the website receives visitors, in theory, ad infinitum. In the case of low viral coefficient the sum of visitors grows slowly; looks of high viral coefficient, on the other hand, may be deceptive as it is more prone to deflation, in particular if it correlates strongly with topicality factor (time). Finally, topicality may fluctuate according to trends, e.g. re-vitalizing expired content.

[8]Termed critical mass in most contexts.

[9]Up to a certain point, of course — cognitive and biological constraints apply also in the Web (see Dunbar & Dunbar, 2010; Goncalves, Perra, & Vespignani, 2011).

[10]Consider a person reading free news articles at a news portal: is he a customer or not? Clearly he consumes the content - on the other hand, he is not paying for it. However, one can argue he is "paying" for it by giving attention which is monetized by the service provider by showing advertisements (Pujol, 2010). Hence, one has a saying about online business models: "If you are not the customer, you are the product."

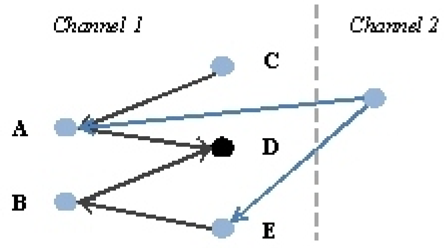

Figure 3. Sharing pattern and the duplicate problem

developed a model of viral growth. The parameters of his model include (Chen, 2008):

$$\begin{aligned} a &= invite\ conversion\ rate(\%) \\ b &= average\ invites\ per\ person \\ c &= initial\ target\ group \\ d &= carrying\ capacity \\ t &= time \end{aligned}$$

The underlying idea is that we start with an initial target group (c) — this is equivalent to the concept of "patient zero", because the initial target group is the basis for viral growth. At each invite-accept iteration loop, the size of the group increases[11], as members send invitations which have a certain conversion rate[12] (a). Each subsequent batch of members has to exceed the previous batch in order to "go viral" and the ratio is the viral coefficient. Based on these variables, Chen's (2008) initial logic is

$$V(t) = c * (1 + b * a)^t$$

The major problem in this model is the lack of carrying capacity — that is, the model assumes viral growth *ad infinitum*. To solve this unrealistic assumption, Chen (2008) adds the concept of carrying capacity, which we define here as:

$$d(a, b, c) = carrying\ capacity,$$

describing the maximum number of people exposed to the viral message — or, in marketing terms, reach[13]. The maximum theoretical carrying capacity is of course limited to the size of the network, defined as the number of unique nodes linking to one another. Because each node (visitor) is connected to other nodes of his social network, by acknowledging small world characteristics, it is theoretically possible to reach all users. However, it is not meaningful to use this figure as the carrying capacity since in practice it will be only a small fraction of users who see even a successful viral campaign (Leskovec et al., 2005). Whereas Chen (2008) uses a constant here, we argue for a factor that scales according to invites, thereby rapidly eliminating "dead ends". Because the number of invitations is unknown prior to *de facto* viral growth, we argue d is the sum of total connections by the users inviting other users minus duplicates. The problem of duplicates is illustrated below.

The figure depicts the diffusion of a message among a set of five users. The duplicate problem occurs with D, because both A and B share to him. If this is not taken into account, the

[11]The visitor base increases when at least one person from the new iteration accepts invitation.

[12]The conversion rate is calculated simply by dividing the number of accepted invitations by the total number of invites sent; it indicates the "viral power" of a message.

[13]It is important to acknowledge the diffusion has a specific boundary which is e.g. the finite size of a particular social network. In Facebook's case, for example, this would be 800 million people.

model incurs duplicates, thereby skewing the number of people shared to[14] as well as the carrying capacity which is a derivative of this. Consider for example a network with strong membership coherence; for example, if A and B would share 50% of their contacts — which is not impossible in a coherent social network - the invitations could experience considerable overlap[15]. Increased frequency of messages can also be interpreted as spam[16], leading to negative attitudes towards a brand[17]. Further, invitations may take place across channels, e.g. different social networks, which complicates tracking. Overlap among exposed members also creates a problem when inserting real numbers into the formula: if we measure all invitations without discounting duplicates, we get an artificially lower conversion rate[18], because we should only count recommendations based on number of unique individuals, not numbers of invitations sent[19]. The higher coherence inside a recommendation network, the more likely overlap in recommendations because people partially share the same contacts.

Returning to Chen's (2008) model, the "discount" on the conversion rate should be related to the total percentage of invited visitors. The assumption here is that the only people who will not accept the invitation are the people who have already done it. This can be described mathematically so that

$$V(t) = c * \left(1 + \frac{b * a}{d(a,b,c)}\right)^t$$

We see that the carrying capacity depends on the parameters a,b,c. However; this model is also too simple since it does not consider saturation which is a crucial assumption of exponential growth models. The marketer is interested in saturation, because it describes the point where the firm dominates the market and no more customers can be found. We will tackle saturation in the advanced model.

c. Advanced model of viral growth

In this third model, we are going to use the basic theories of finance combined with logistic function. We claim not a perfect model but a good starting point for some major contribution in theory of viral marketing. To begin, in finance the grow factor of money saved in the bank account is derived from the following equation.

$$V(t) = (1+i)^t * c.$$

When applying the more convenient equation, we get the continuous factor

$$V(t) = e^{pt} * c.$$

From these two equations it is possible to solve p

$$1 + i = e^p$$

[14]Note that the act of inviting can remain two or more in case of duplicates, but we are interested in the number of people invited, not in the act of inviting itself. Also, as Leskovec et al. (2005) discovered, the propensity for desired action does not necessarily increase with additional word-of-mouth communication, but may actually decrease as a result of fatigue. This is a clear distinction from epidemic models where the probability of infection increases with repeated interaction.

[15]Therefore, the more coherent the network is, the more likely the overlap of invitations. For further implications see Leskovec et al. (2005); and Centola (2010) for neighboring effects.

[16]I.e., unsolicited and intrusive messages - to counter control for such effects, marketers are able to use online tools for analyzing the sentiment of customers (Kalyanam, McIntyre, & Masonis, 2007).

[17]Note that from the recipient's perspective it is irrelevant if the repetition comes from one source or several sources; it is nevertheless excessive unsolicited information.

[18]Consider two cases: 1) four invitations are sent to two persons, one converts to visit; and 2) two invitations are sent to the same persons, one converts. The conversion rate in former, counting by number of invitations, is 1/4=25%, and in the latter by number of people invited, 1/2=50%.

[19]Similar problem is associated with measuring visits vs. unique visitors; the latter gives less accurate information about the customers, in which the marketer is interested.

and

$$p = \ln(1+i).$$

In theory of finance, the factor i refers to interest but in viral marketing we have the factor $b * a$. By using the previous structure we get

$$V(t) = e^{t*\ln(1+i)} * c = e^{t*\ln(1+ba)} * c$$

with the adjusted conversion rate the form is

$$V(t) = c * e^{t*\ln\left(1+\frac{ba}{d(a,b,c)}\right)}.$$

The formula depicts increase in time. How is saturation then taken into consideration in our model? In this case, we rely on the logistic function by applying the basic form

$$P(t) = \frac{1}{1+e^{-t}}.$$

In our case we get

$$P(t) = \frac{d(a,b,c)}{d(a,b,c) + V(t)^{-t}} = \frac{d(a,b,c)}{d(a,b,c) + c * e^{t*\ln\left(1+\frac{ba}{d(a,b,c)}\right)^{-t}}}$$

Knowing the rules of logarithmic calculation, the final form of this function and, thus, our model for viral marketing coefficient, is

$$P(t) = \frac{d(a,b,c)}{d(a,b,c) + c * e^{-t^2*\ln\left(1+\frac{ba}{d(a,b,c)}\right)}}.$$

Limitations

There are several limitations to this paper, some of which are shared with other models of viral marketing while others touch specifically our model.

a. Specific limitations

The major shortcoming in the model is the function d whose nature is still unknown to us, and therefore not well defined in the model. Yet, we assume structure of recommendation networks is *ad hoc* as oppose to static — although the social network of a set of individuals may remain relatively stable, the diffusion process is not identical due to differences in acceptance which leads to unpredictable paths of diffusion[20]. Since we cannot predict diffusion *a priori*, we cannot accurately predict the structure of the network either (although we can define limitations to its size). This is what causes trouble in determining carrying capacity. For example, in Facebook the crucial action is *liking* interesting content - not necessarily inviting other visitors[21]. The action of liking results in the content appearing on the activity feed of a person's friends[22]. However, the carrying capacity is different for invitations versus liking. In our invitation based model, d is the sum of all invitations sent in each round of iteration that the viral growth entails. In liking, reach

[20]Highly clustered networks perform better when measuring cooperation (Suri & Watts, 2011) - this can explain low rates of viral diffusion - as clustering density decreases, so does the willingness to cooperate. Further, during the diffusion process the network is subject to both churn and adding users.

[21]'Liking' is non-object-intentional and inviting is object-intentional action.

[22]I.e., the viral growth, modeled as the outcome of diffusion in an inter-connected node based network is skewed by non-predictable diffusion path - in practice this means Internet marketers are trying to find ways of "gaming" the algorithm and thus improving the prominence of their message as oppose to competing messages. However, the algorithm is most likely self-learning, meaning that the performance of the message among the base of users is used as a criterion to determine its prominence.

of the message is determined by a mediator[23]. Because the mediator hides the exact formula for diffusion (for introduction, see TechCrunch, 2010), the factor can only be described at a general level, and appears random[24]. How d differs then is that in the like-based model it would be more unpredictable — due to seeming randomness of selection algorithms — and there is much more freedom in the viral growth pattern, as the degree of friction in liking is less than that of invitations; essentially, users are disseminating the message with a minimal effort. In theory this results in d scaling relatively easily.

There are several alternative approaches to model viral growth that have not been explored in this paper. These include e.g. survival analysis (for overview, see Klein, 1992); random walk (Spitzer, 2001); and probabilistic models (Dellarocas & Narayan, 2006) measuring individuals' propensity to invite others. For example, survival analysis could be used to analyze large data sets to identify patterns of viral expiration. By combining this with qualitative classification one can create typologies for viral growth of different product categories — this information can be applied to predict the viral growth of a product within a specific category, and understanding the difference of viral factor between distinct products[25]. The random walk approach is hindered by "social laws" — for example proximity, preference influence, social identification, and network externalities (Yang & Allenby, 2003). Hence, if the data seems random the analyst has most probably failed in finding the patterns and relationships between individuals[26].

Finally, what happens to other variables once an independent variable changes? Such an analysis to our model would be possible through empirical data. Empirical data might also bring substantial modifications to the underlying theoretical assumptions of our model, if it would not provide accurate results. Second, the question of profit optimization: which point of time is optimal for profit? To answer this, one has to take into account what happens *after* the visitor visits the website — that is, in the sales funnel. This paper bases its theoretical assumptions on visitors, not users or customers[27] . The notable difference is that when modeling the growth of user base, one has to consider additional factors, such as *churn* and loyalty rate. To demonstrate, if the churn rate is higher than the viral coefficient, the business loses customers[28].

b. General limitations

The general limitations to measuring viral growth that we have identified include:

1. **Number of patient zeros.** The number of patient zeros can be influenced by the marketer through advertising, but what is the optimal ratio between advertising and viral growth? This question relates to whether or not assume a critical mass which, in turn, relates to network effects which are not measurable straightforward but can be proxied through the viral growth factor itself! In such a circular argument, viral growth may in fact model the network effects of the viral entity in question, not the appeal of the viral message. This is a very important distinction for marketers, as discussed in implications[29].

[23]For example, in Facebook the Edgerank algorithm determines which posts are shown to which infividuals. There are two relaxations to secrecy of the algorithm: 1) a general model of EdgeRank has been made public by Facebook, and 3) general statistics of reach are revealed to firms about their posts.

[24]In contrast, object-intentional messages such as invitations by email are always directed to specific recipients whose number and quality is known.

[25]However, large data sets are a requisite because the exponential nature of viral growth risks skewing the results.

[26]Yang and Allenby (2003) offered to solve this with their "Bayesian spatial autoregressive discrete-choice model."

[27]The chain of roles is as follows: visitor →user→ customer, so that one has to visit the site to become a user, and use it to become a customer. This does not, however, claim that this transformation could not happen instantly, only that visits may hold a different utility for the firm and should not be weighed equally in marketing decisions.

[28]Because new customers are not able to compensate for the loss of old ones. This has been contrasted to a leaking bucket: no matter how much one adds water; that is, acquires new customers, the amount of water remains the same.

[29]Consider two products: Product A that has a strong inherent virality but low marketing efforts, and Product B that has low inherent virality but high marketing virality - e.g., a viral campaign. Which one would perform better? We'd argue for Product A because the focus of interest is genuinely on the product when disseminating,

2. **Quality of patient zeros.** In addition to number, the selection of proper subset of patient zeros is an issue — as asked by Kempe, Kleinberg, and Tardos (2003), which segment should the marketer choose to optimize the cascade of viral diffusion? The traditional answer is "early adopters" that are most willing to adopt novelties (Rogers, 1995). However, beyond the scope of innovations the problem endures - further, who are early adopters and how can marketers find them? The answer has been sought by many (e.g. Wang, Cong, Song, and Xie, 2010) but no definitive solution exists yet.
3. **Quality of invites.** Visitors' ability to create effective invitations affects the acceptance rate which, in turn, affects the viral coefficient — this process is not controlled by the firm at all, and may be handled poorly[30]. The viral message can also have negative valence towards the firm (Richins, 1983)[31], so not all diffusion is in fact desirable.
4. **Value of visitors.** We know from marketing theory that the value of different customers varies (Gupta et al., 2006) — the same applies to website visitors. Marketers may want to segment customers into different groups based on their profitability and manage customer portfolios accordingly (Terho, 2008). Although the value of a customer is a function of *lifetime*, viral coefficient does not consider the variance among customers and, consequently, the profit earned by the firm[32] . This is a major shortcoming and present in most models of viral growth; most likely due to their short-term, temporal focus. Some authors have succeeded in dealing with this issue by calculating the optimal marketing spend per customer (Richardson & Domingos, 2002).
5. **Aggregate problem.** Like all aggregates, viral coefficient lacks the insight into the unique individual; therefore, the depth of quality of infection, i.e. attitude towards the viral message, cannot be determined by this means. Yet, marketers *are* interested in individuals' attitudes towards a brand or content. Through aggregates we can estimate interest through invitations - when viral coefficient is higher than zero, the general attitude can be judged to be positive[33]. Further, one can apply agent-based modeling to introduce purposefulness (Goldenberg, Libai, & Muller, 2010; Stonedahl, Rand, & Wilensky, 2010).
6. **Isolation of channels.** The marketer is interested in optimizing his marketing spending, which means choosing the most efficient mix of channels. Channels tend to differ by performance[34], so aggregating them under one measure makes the marketer unaware of differences. Although one can measure channels separately, this does not sustain in measuring cross-channel diffusion[35]. However, one could experiment by normalizing each channel and comparing the viral diffusion *sourcing* between them. This means one would account for cross-channel spread but only so that length of viral diffusion is examined separately by

whereas in Product B the marketer "tricks" the attention through peripheral route.

[30]In fact, when the format of a viral message is decided by the invitee, large quality variance follows. Yet, the marketer may be able to standardize the message by "building in virality" (e.g. standard message in Hotmail signatures).

[31]I.e. negative word of mouth: "The service in that restaurant was so bad" could be a classical example heard by almost everyone by some point from their peers.

[32]It can easily be calculated that if a campaign A brings ten new customers with average lifetime value of 100 , and campaign B (as a result of higher viral coefficient) 100 new customers with average lifetime value of five euro, we'll get $A : 10 * 100$€= 1.000€versus $100 * 5$€= 500€; hence, higher viral growth does not contribute to profit, ceteris paribus.

[33]Yet, this does not indicate the valence of attitude; consider, for example, a negative story of a brand that spreads like wildfire. By measuring virality the firm sees it's doing a "good job", although the attitudes towards the brand are likely to become worse and worse by further diffusion of the message (note that this is an inherent risk of any messages disseminated by groups or individuals other than the firm). Examining the variance, we may distinguish between recommendation (invitation), referral (leading to action), and disproval (refusal - either due to disinterest or negative valence of message).

[34]The viral effect across channels varies depending on factors such as the match between the content and the audiences. Consider a case of two channels: the first channel has a very low viral coefficient; e.g., marketing dog food to cat owners, whereas in another channel it would be high; e.g., marketing in Facebook group for dog owners. The average viral growth of the visitors to the site would be the average of these two channels; whereas the manager would benefit more from channel-specific data as he already knows the efforts made in different channels.

[35]E.g., person A sharing from channel A to person B in channel B who converts; where to attribute conversion?

comparison of patient zeros between channels.

7. **Instability of network.** The predictive ability of a model is determined by the nature of the network and the nature of the product. Because all data is derived from an idiosyncratic setting, the stability of the network topology may become an issue. In other words, if connections frequently change and the network loses coherence, then the predictive ability of the model may suffer. However, it is unclear how stable social networks are; one hand, there are constant dynamics involved; on the other hand, despite the change of consistency, the behavior of viral diffusion perhaps will not dramatically change, as social behavior remains similar despite of changing invitation pairs[36].

Managerial implications

Despite its limitations, our model offers grounds for improving accuracy and utility of viral marketing theory and practice. First, the model resolves some of the shortcomings of existing models, thus aiming for improved theoretical accuracy. Second, by knowing the parameters, marketers are able to measure their product's viral growth at a certain point in time with improved precision. Contrasting this information to marketing efforts in a given period will give insight into the performance of specific marketing actions, which is helpful for guiding decision making. Although viral growth is ideally a self-sustaining process, following advice can be given:

1. *a. Viral diffusion can be assisted by increasing number of patient zeros through advertising* — this is particularly beneficial for websites with network effects, because the utility of the service to additional users grows in proportion to user base - consider the incentives of a visitor in an empty discussion board.
2. *a. Marketers can find and persuade key influencers* — they play a key role because the limits of audience reached through advertising remains a fraction of the entire carrying capacity, so the marketer is better of focusing his efforts on the subset of members that are more likely to respond positively to its message. Identifying influencers can be based on combination of qualitative and quantitative research through search engines, and does not necessarily require highly sophisticated methods or tools.
3. *c. Marketers can apply experiments to address problems of modeling.* Kalyanam et al. (2007) argued that adaptive experimentation improves efficiency of viral marketing efforts because marketers are able to fine-tune their decisions based on data fitted on simple metrics. Lans, Bruggen, Eliashberg, and Wierenga (2009) developed a branching model to extrapolate a large-scale diffusion from relatively small datasets that can be used to predict reach in various what-if scenarios. We suggest that in particular the first invitation iteration is useful in extrapolating the conversion rate of further iterations; thus, the marketer may be able to quickly test and adjust the viral appeal until a good match between message and audience is found.

Future research

In terms of guidelines for future research, we agree with Cheung and Thadani (2010) in that a coherent theory, or a set of theories, of viral marketing is needed — so far, researchers have focused on applying previous theories into viral phenomenon. In our view, the lack of *absolute* analytical accuracy can be compensated through insight on factors influencing the propensity of inviting — i.e., the visitors' motives of sending, accepting and refusing invitations[37]. For example, game theory can be applied to analyze incentives in a viral loop (Kempe et al., 2003). Others have made attempts to better understand *why, how* and *to whom* messages are sent (e.g. Phelps, Lewis, Mobilio, Perry, & Raman, 2004). In addition, qualitative research is needed to understand

[36]For example, Hill (2006) found that "network neighbors" convert considerably higher than more distant nodes - this behavior is likely to occur regardless of which individual nodes communicate.

[37]These are myriad and seem at first introduce an additional layer of complexity — however, a good theory has the power of reducing and simplifying complex phenomena into a coherent set of logical assumptions.

why certain content is being shared more than other — or crack the secrets of viral content. This work has been pioneered by e.g. Berger and Milkman (2009) and Jihua (2011).

References

[1] Berger J.A., & Milkman K.L. (2009). What Makes Online Content Viral? SSRN eLibrary. Retrieved from
$http://papers.ssrn.com/sol3/papers.cfm?abstract_id = 1528077$

[2] Centola D. (2010). The Spread of Behavior in an Online Social Network Experiment. Science, 329(5996), 1194-1197. doi:10.1126/science.1185231

[3] Chen A. (2008) Facebook viral marketing: When and why do apps "jump the shark?" $http://andrewchen.co/2008/03/05/facebook - viral - marketing - when - and - why - do - apps - jump - the - shark$, retrieved June 21, 2012.

[4] Cheung C.M.K., & Thadani D.R. (2010). The State of Electronic Word-Of-Mouth Research: A Literature Analysis. PACIS 2010 Proceedings. Retrieved from http://aisel.aisnet.org/pacis2010/151

[5] Dellarocas C., & Narayan R. (2006). A Statistical Measure of a Population's Propensity to Engage in Post-Purchase Online Word-of-Mouth. Statistical Science, 21(2), 277-285. doi:10.1214/088342306000000169

[6] Dichter E. (1966). How word-of-mouth advertising works. Harvard Business Review, 44(6), 147-160.

[7] Dunbar R., & Dunbar R.I.M. (2010). How Many Friends Does One Person Need?: Dunbar's Number and Other Evolutionary Quirks. Harvard University Press.

[8] Fletcher M. (2006) Presentation at Startup SIG.
$http://www.niallkennedy.com/blog/2006/05/mark - fletcher - bloglines - onelist.html$, retrieved October 10, 2012.

[9] Goldenberg J., Libai B., & Muller E. (2010). The chilling effects of network externalities. International Journal of Research in Marketing, 27(1), 4-15. doi:10.1016/j.ijresmar.2009.06.006

[10] Goncalves B., Perra N., & Vespignani A. (2011). Validation of Dunbar's number in Twitter conversations. arXiv:1105.5170. doi:10.1371/journal.pone.0022656

[11] Gupta S., Hanssens D., Hardie B., Kahn W., Kumar V., Lin N., Ravishanker N., et al. (2006). Modeling Customer Lifetime Value. Journal of Service Research, 9(2), 139-155. doi:10.1177/1094670506293810

[12] Hill S. (2006). Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. Statistical Science, 21(2), 256-276. doi:10.1214/088342306000000222

[13] Hood, D. (2012) Going viral: Researchers in quest of marketing's 'holy grail'.
$http://www.washingtontimes.com/news/2012/feb/23/going - viral - researchers - in - quest - of - marketings - hol/?page = all$, retrieved September 27.

[14] Jihua Z. (2011). Research on propagation mechanism of viral marketing. Business Management and Electronic Information (BMEI), 2011 International Conference on (Vol. 1, pp. 325 -327). doi:10.1109/ICBMEI.2011.5916939

[15] Kalyanam K., McIntyre S., & Masonis J.T. (2007). Adaptive experimentation in interactive marketing: The case of viral marketing at Plaxo. Journal of Interactive Marketing, 21(3), 72-85. doi:10.1002/dir.20086

[16] Kempe D., Kleinberg J., & Tardos E. (2003). Maximizing the spread of influence through a social network. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '03 (pp. 137-146). New York, NY, USA: ACM. doi:10.1145/956750.956769

[17] Khelil A., Becker C., Tian J., & Rothermel K. (2002). An epidemic model for information diffusion in MANETs (p. 54). ACM Press. doi:10.1145/570758.570768

[18] Klein J.P. (1992). Survival Analysis: State of the Art. Springer.

[19] Lans R. van der Bruggen G. van Eliashberg J., & Wierenga B. (2009). A Viral Branching Model for Predicting the Spread of Electronic Word of Mouth. Marketing Science. doi:10.1287/mksc.1090.0520

[20] Leskovec J., Adamic L., & Huberman B. (2005). The Dynamics of Viral Marketing. In EC '06: Proceedings of the 7th ACM conference on Electronic commerce (pp. 228-237). ACM Press.

[21] Phelps J.E., Lewis R., Mobilio L., Perry D., & Raman N. (2004). Viral Marketing or Electronic Word-of-Mouth Advertising: Examining Consumer Responses and Motivations to Pass Along Email. Journal of Advertising Research, 44(04), 333-348. doi:10.1017/S0021849904040371

[22] Pujol N. (2010). Freemium: Attributes of an Emerging Business Model. SSRN eLibrary. Retrieved from
$http://papers.ssrn.com/sol3/papers.cfm?abstract_id = 1718663$

[23] Richardson M., & Domingos P. (2002). Mining knowledge-sharing sites for viral marketing. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '02 (pp. 61-70). New York, NY, USA: ACM. doi:10.1145/775047.775057

[24] Richins M. (1983). Negative word-of-mouth by dissatisfied consumers: A pilot study. Journal of Marketing, 47(1). Retrieved from
$http://search.proquest.com/docview/227803126/abstract/embedded/IDFPJ9BPOK3O8PK3?source = fedsrch$

[25] Rink D.R., & Swan J.E. (1979). Product life cycle research: A literature review. Journal of Business Research, 7(3), 219-242. doi:10.1016/0148-2963(79)90030-4

[26] Rogers E.M. (1995). Diffusion of Innovations. Simon and Schuster.

[27] Schnettler S. (2009). A structured overview of 50 years of small-world research. Social Networks, 31(3), 165-178. doi:10.1016/j.socnet.2008.12.004

[28] Shannon C.E., Weaver W., & Shannon. (1998). The Mathematical Theory of Communication (1ST ed.). University of Illinois Press.

[29] Spitzer F. (2001). Principles of Random Walk. Springer.

[30] Stonedahl F., Rand W., & Wilensky U. (2010). Evolving viral marketing strategies. Proceedings of the 12th annual conference on Genetic and evolutionary computation, GECCO '10 (pp. 1195-1202). New York, NY, USA: ACM. doi:10.1145/1830483.1830701

[31] Suri S., & Watts D.J. (2011). Cooperation and Contagion in Web-Based, Networked Public Goods Experiments. PLoS ONE, 6(3), e16836. doi:10.1371/journal.pone.0016836

[32] TechCrunch — EdgeRank: The Secret Sauce That Makes Facebook's News Feed Tick. (n.d.).TechCrunch. Retrieved June 21, 2012, from
$http://techcrunch.com/2010/04/22/facebook-edgerank/$

[33] Terho H. (2008) Customer portfolio management - The construct and performance. Turku School of Economics, Series A-4: 2008. Doctoral dissertation.

[34] Tokuda L. (2008) Arch Viral: Creating Social Apps for Social Platforms (Lance Tokuda, Startonomics SF 2008).
$http://www.slideshare.net/Startonomics/arch-viral-creating-social-apps-for-social-platforms-presentation?src = embed$, retrieved June 21, 2012.

[35] Wang Y., Cong G., Song G., & Xie K. (2010). Community-based greedy algorithm for mining top-K influential nodes in mobile social networks. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '10 (pp. 1039-1048). New York, NY, USA: ACM. doi:10.1145/1835804.1835935

[36] Yang S., & Allenby G. (2003). Modeling Interdependent Consumer Preferences. Journal of Marketing Research, 40(3), 282-294.

Authors

Joni Salminen — 3rd year Ph.D, Faculty of Marketing, Turku School of Economics, Turku, Finland; E-mail: *joni.salminen@tse.fi*

Aarni Hytonen — 4th year master, Faculty of Quantitative Methods in Management, Turku School of Economics, Turku, Finland; E-mail: *aaarhy@utu.fi*