Viral coefficient – Unveiling the Holy Grail of online marketing

MSc. (Econ.) Joni Salminen[1]
joni.salminen@tse.fi
BSc. (Econ.) Aarni Hytönen[1]
aaarhy@utu.fi
Turku School of Economics

Abstract

In this paper, we tackle some shortcomings of viral marketing models: (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 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 Horoscope (1.5M users in 15 days). For example, one founder describes his company's viral growth as follows (Fletcher, 2006):

> *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.*

[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).

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 model diffusion of peer-marketing messages in a finite social network?* We approach this question by developing a model measuring the growth of number of visitors that results 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 model by the Internet marketing professional Andrew Chen (2008), and finally our own model aimed at improving the other models.

a. The base model

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

$$v = x * y,$$

where v = viral growth factor
y = efficiency of the viral loop
x = number of persons a user invites

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.

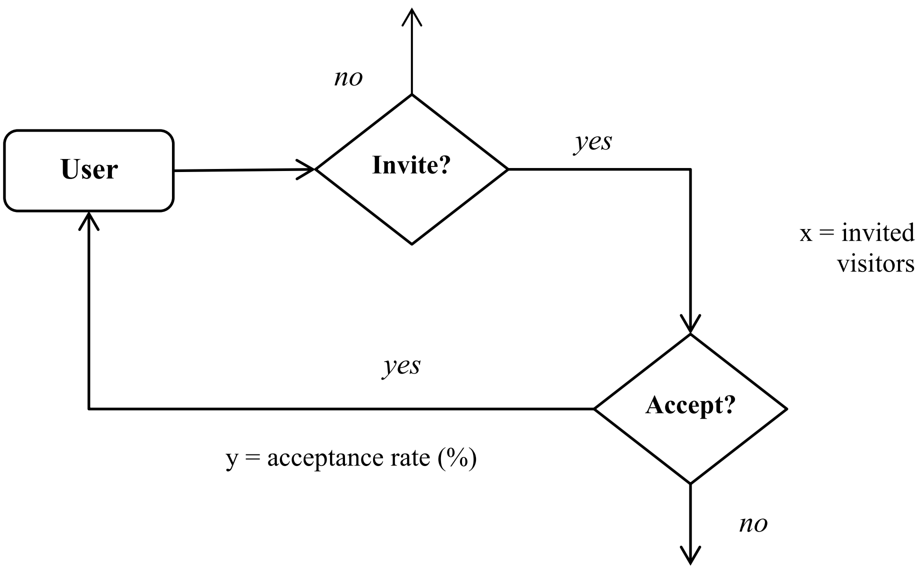

Figure 1 **Process of viral marketing (Tokuda 2008)**

According to this model[3], one achieves viral growth if

$$x * y > 1,$$

i.e. earlier invited visitors invite more than one new visitor[4] and more than one of them accepts the invitation[5]. 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, it is a circular model that does not tell much of the dynamics between the factors influencing viral growth. More specifically, its 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. 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.

As such, there are three possibilities for viral growth in terms of time (Figure 2): 1) *expiration* where viral coefficient is negative [C]; 2) *stability*, where growth factor remains constantly over one but results in linear growth [B]; and 3) *inflation* in which the factor leads to exponential growth [denoted as A in the figure].

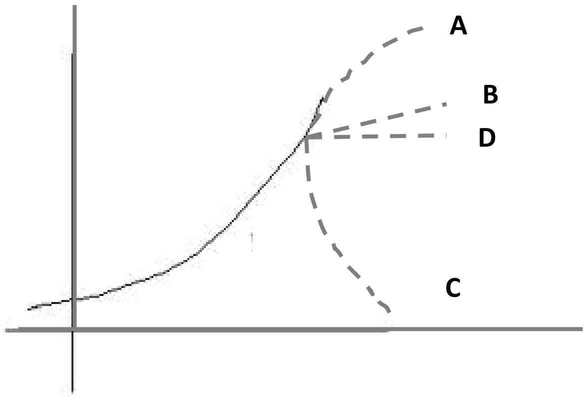

Figure 2 Possible relationships between viral coefficient and time

For viral messages whose topicality is high (e.g. news), 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 [D in graph]. Therefore, when topicality is high, the viral coefficient is high and *vice versa*.

---

[3] 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).

[4] 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).

[5] 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.

The cumulative growth curve can never turn to negative slope when examining number of visits. This is because a visit is an irreversible event that cannot be undone. In contrast, [C] is possible when measuring customer relationship, as customers can become non-customers (*churn*) at any point of their lifetime. Therefore, [C] demonstrates the fundamental difference of measuring visitors and customers, our model and lifetime models. One would expect that visitors and customers behave differently in disseminating viral messages since becoming a customer (*converting*) is an act of higher engagement than a website visit. However, not only customers but also visitors spread viral messages.

b. Chen's model

Acknowledging the shortcomings of the base model, the Internet marketer Andrew Chen has developed a model of viral growth. The parameters of his model include (Chen, 2008):

a = invite conversion rate (%)
b = average invites per person
c = initial target group
d = carrying capacity
t = time

The initial target group (c) is equivalent to the concept of "patient zero", the basis for viral growth. At each invite–accept iteration loop, the size of the group increases[6], as members send invitations which have a certain conversion rate[7] (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:

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

describing the maximum number of people exposed to the viral message – or, in marketing terms, reach[8]. 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 – however, in practice, only a small fraction of users see even a successful viral campaign (Leskovec et al., 2005). Whereas Chen (2008) uses a constant here, we argue for a factor

[6] The visitor base increases when at least one person from the new iteration accepts invitation.
[7] 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.
[8] 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.

that scales according to invites, thereby rapidly eliminating “dead ends”. In 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 can find no more visitors. 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. To begin, in finance the grow factor of money saved in a 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

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

From these two equations it is possible to solve ρ

$$1 + i = e^{\rho}$$

and

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

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 this case, we rely on the logistic function by applying the basic form

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

Therefore 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 and discussion

The major shortcoming in the model is the function *d* whose nature is still unknown to us, and therefore not well defined. Yet, we assume recommendation networks to be dynamic by nature – although the connections of individuals remain relatively stable, the diffusion itself is idiosyncratic due to differences in acceptance rates which lead to unpredictable paths of diffusion[9]. 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 explicitly inviting other visitors[10]. The action of liking items is primarily targeted towards the object of interest and only indirectly leads to diffusion. Therefore, the carrying capacity is different for invitations versus actions such as 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 of the message is determined by the mediator[11]. Because the mediator hides the exact formula for diffusion (see TechCrunch, 2010), the factor can only be described at a general level[12]. How *d* differs then is that in a like-based model it would be more

[9] 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.

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

[11] 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.

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

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[13]. 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[14].

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[15]. The notable difference is that when modeling the growth of user base, one has to consider additional factors, such as *churn* and *loyalty*. For instance, if the churn rate is higher than the viral coefficient, the business loses customers[16].

Managerial implications

By understanding the parameters, marketers can measure their product's viral growth at a given point in time. Contrasting this information to marketing efforts will give insight into the performance of those actions. Although viral growth is ideally a self-sustaining process, following advice can be given:

---

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

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

[15] 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.

[16] 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.

1. *Viral diffusion can be assisted by increasing the 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. *Marketers can find and persuade key influencers* – when the audience reached through advertising remains a fraction of the carrying capacity, the marketer is better off focusing his efforts on the subset of members that are more likely to spread the message. Identifying influencers does not necessarily require highly sophisticated methods or tools, but it requires creating an interesting proposal.
3. *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.

Marketers may consider 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 campaigns[17]. The main strategies for maintaining a high viral coefficient seem to be either increasing topicality of the message or network externalities of the product.

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[18]. 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). Qualitative research is needed to understand why certain content is being shared more than other. This work has been pioneered by e.g. Berger and Milkman (2009) and Jihua (2011).

---

[17] 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.

[18] 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.

References

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

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

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.

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

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

Dichter, E. (1966). {How word-of-mouth advertising works}. *Harvard Business Review*, *44*(6), 147–160.

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

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

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

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

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

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.

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

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

Kempe, D., Kleinberg, J., & Tardos, É. (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

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

Klein, J. P. (1992). *Survival Analysis: State of the Art*. Springer.

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

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.
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
Pujol, N. (2010). Freemium: Attributes of an Emerging Business Model. *SSRN eLibrary*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1718663
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
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
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
Rogers, E. M. (1995). *Diffusion of Innovations*. Simon and Schuster.
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
Shannon, C. E., Weaver, W., & Shannon. (1998). *The Mathematical Theory of Communication* (1ST ed.). University of Illinois Press.
Spitzer, F. (2001). *Principles of Random Walk*. Springer.
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
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
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/
Terho, H. (2008) Customer portfolio management – The construct and performance. Turku School of Economics, Series A-4: 2008. Doctoral dissertation.
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.
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
Yang, S., & Allenby, G. (2003). Modeling Interdependent Consumer Preferences. *Journal of Marketing Research*, *40*(3), 282–294.