Communications of the Association for Information Systems

Volume 24 Article 32

6-1-2009

Impacts of Tie Characteristics on Online Viral Diffusion

Janghyuk Lee
Korea University Business School, janglee@korea.ac.kr

Jong-Ho Lee
Korea University Business School

Dongwon Lee
Korea University Business School

Communications of the Association for Information Systems

CAIS

Impacts of Tie Characteristics on Online Viral Diffusion

Janghyuk Lee

Department of Marketing, Korea University Business School

janglee@korea.ac.kr

Jong-Ho Lee

Department of Marketing, Korea University Business School

Dongwon Lee*

Department of MIS, Korea University Business School

Abstract:

To explain the viral diffusion process, most of previous studies focus on the structure of social networks and the existence of the hub. We extend the scope of analysis from a single node to a tie between the sender and the receiver to explain the impact of tie characteristics on the viral diffusion measured by its speed (i.e., how quickly) and by its volume (i.e., how much viral). Based on our analysis results using a viral marketing data of 30,035 sender-receiver ties, we find that (1) the more heterogeneous the tie is, the quicker the response occurs; and (2) heavy viral generators tend to be connected to each other. Taken together, this research broadens the study of online viral diffusion by applying tie characteristics in terms of the volume and speed of the viral.

Keywords: Assortativity, homophily, social network, viral marketing, Web, word-of-mouth

Volume 24, Article 32, pp. 545-556, June 2009

*corresponding author, mislee@korea.ac.kr
The manuscript was received 10/6/2008 and was with the authors 1 month for 1 revision.

Communications of the Association for Information Systems

I. INTRODUCTION

The success of viral marketing contrasts the rising reticence among consumers toward advertisements. According to technographic survey results conducted by Forrester Research [Nail 2005], more consumers do not pay attention to advertisements (from 27 percent in 2002 to 33 percent in 2004) and less consumers found advertisements entertaining (from 41 percent in 2002 to 21 percent in 2004). A dramatic drop was recorded for the role of advertising to collect information of new products. Only 47 percent of respondents thought that advertisements are a good way to learn about new products (drop from 78 percent in 2002).

The diffusion of information from one person to another has been considered as an effective communication channel [Godes and Mayzlin 2004; Herr et al. 1991] since Katz and Lazrasfeld [1955] explored the role of personal contacts as the most effective source of information for the brand switching decision in the case of foods, household goods, and motion pictures. Due to rapid technology developments, offline-based personal contacts have been extended into various electronic forms through e-mails, instant messengers, blogs, online communities, etc. [Kalika et al. 2008; Watson-Manheim and Bélanger 2007]. Especially, the development of recent social computing services such as social networking (e.g., MySpace, Facebook), media sharing (e.g., YouTube, Flickr), micro-blog (e.g., Tumblr, Twitter), social mapping (e.g., Dodgeball by Google, Beacon Buddy by Helio), and tagging (e.g., Dgm8, Socialight) enhances the power of personal influence; they capture the attention of managers to leverage interactive consumer viral marketing through the use of social computing across industries [Lakshmipathy 2007; Parameswaran and Whinston 2007].

Viral marketing is one of the well-known marketing techniques that leverages electronic personal contacts in preexisting social networks to achieve marketing objectives [Chevalier and Mayzlin 2006; Godes and Mayzlin 2004; Leskovec et al. 2007]. Self-replicating processes in which consumers spread voluntarily product-related messages distinguish viral marketing from other marketing activities that try to deliver the message directly to target consumers [Anderson 1998]. Compared to offline word-of-mouth, viral marketing takes actions to systematically measure consumer behavior, which provides enhanced manageability for the marketer to intervene in the marketing process at any time if necessary [Helm 2000; Subramani and Rajagopalan 2003].

Since the success of MSN's Hotmail, which opened 12 million accounts in its first year, 1996, and the 1999 hit film *The Blair Witch Project* demonstrated the power of viral marketing in the early Internet era [Neuborne 2001], the number of successful cases of viral marketing have cumulated rapidly. For example, campaigns such as Burger King's "The Subservient Chicken," Cadbury's Diary Milk "Gorilla Advert," and the film, *The Dark Knight*, superseded one another on such hit parades [Walker et al. 1994]. The recent success of the viral campaign "Mr. Hopewood" organized by BigFix, a California-based software company, expanded the application area of viral marketing into the business-to-business (B2B) environment beyond the business-to-consumer (B2C) industry [Elliot 2007]. The viral empowered by information technologies in various forms such as through e-mails, blogs, social network services, and consumer-to-consumer (C2C) e-commerce influences consumers effectively by providing useful and impartial information [Parameswaran and Whinston 2007]. The rating of TV shows was also influenced by the dispersion of online conversation generated across Usenet communities [Godes and Mayzlin 2004]. At online bookstores, such as Amazon.com and BarnsandNoble.com, their book sales ranking was significantly influenced by the quality of online reviews with a relatively more influence of negative reviews than positive ones [Chevalier and Mayzlin 2006].

Previous studies on viral marketing and social networks mostly focus on the network structure and how it affects the diffusion process of information. For example, Watts and Strogatz [1998] explained how a random link can substantially improve the connectivity of networks by reducing the average length among network participants. Using mobile telecommunications data, Onnela et al. [2007] showed the role of weak ties to connect remote communities as well as that of strong ties to maintain local communities, and simulation results demonstrated that information diffusion can be slowed down in a network which has unevenly weighted links. Because marketing messages have to be timely and interesting to consumers, the total volume and speed of the viral generated within a given period time is a critical indicator to gauge the performance of a viral marketing campaign [Leskovec et al. 2007]. To complement the findings of viral marketing from a structural perspective, therefore, we raise two key questions to better understand the dynamics of viral marketing in this study: (1) how much the viral will be generated (i.e., the volume of the viral); and (2) how quickly the viral is diffused (i.e., the speed of the viral) across different tie characteristics. More specifically, we examine the diffusion process of the viral, based on a data set of viral marketing, to explain how the tie characteristics affect the total size of the viral composed of the volume and the speed of diffusion at the individual level.

The remainder of the paper is organized as follows. We first present, in Section II, the theoretical background of viral marketing, social networking, and the formal basis for a set of hypotheses to examine the viral diffusion patterns affected by the tie characteristics. The third section provides the research methodology along with the description of data and applied empirical models. The fourth section describes our empirical findings: (1) the more heterogeneous the tie is, the quicker the response occurs, and (2) heavy viral generators tend to be connected to each other. The fifth section concludes with a discussion of a number of implications for research and practice, as well as the limitations and possible extensions for future research.

II. RESEARCH BACKGROUND AND HYPOTHESES

In this section, we review previous studies on the tie characteristics in social networks and viral marketing, and provide a set of hypotheses to examine the viral diffusion patterns in terms of the speed and volume affected by the tie characteristics.

Tie Characteristics in Social Networks

A number of previous studies on social networks focus on the tie strength or characteristics and its impact on the diffusion of information. A tie consists of a sender and a receiver depending on the direction of information flow [Watts and Strogatz 1998]. Its relationship can be defined by how strongly the tie is connected and how the sender and the receiver share the same characteristics. The stronger the social tie connection between two individuals, the more similar they tend to be [Granovetter 1973; McPherson and Smith-Lovin 1987]. The concept of the tie strength and homophily has been considered as synonymous or as related constructs [Gatignon and Robertson 1985; Rogers 1995].

The *tie strength* includes closeness, intimacy, support, and association [Frenzen and Davis 1990]. Strong ties are characterized by the degree of intimacy and special meaning through a voluntary investment, which have frequent interactions in multiple contexts under a sense of mutuality of the relationship [Walker et al. 1994]. Brown and Reingen [1987] measure the tie strength in a social network of piano teachers with 130 sender-receiver dyads based on the following constructs: (1) relationship category (e.g., friend, neighbor, relative, acquaintance, and others); (2) frequency of communications (e.g., daily, weekly, biweekly, and monthly); and (3) relationship importance (e.g., “merely an acquaintance” and “feels so close to it is hard to imagine life without him/her”). Frenzen and Davis [1990] measure the tie strength by checking the level of confidence, the likelihood to spend free time to give and receive a large favor, and the closeness of the ties. In a transaction-based measurement, Onnela et al. [2007] compute the total conversation time between two users to measure tie strength in a mobile telecommunication network with 4.6 million users. Leskovec et al. [2007] count the frequency of recommendation contacts between individuals as a proxy for the tie strength in a network with about 4 million users.

In a number of previous studies, demographic-related criteria have been used to measure the tie similarity called *homophily*. In offline environments, Brown and Reingen [1987] measure it dichotomously as to whether a tie is homophilous or heterophilos. Ties are classified as homophilous if they have identical information on at least three of four criteria: age, gender, education, and occupation. To measure homophily in online environments, De Bruyn and Lilien [2008] use categorical measures for gender and education, and seven-point scales for age (i.e., 1= I am much younger, 4= same age, 7= I am much older) and occupation (i.e., 1= not at all similar, 7= extremely similar). In association with structural measures of social networks, Catanzaro et al. [2004] present an assortative model that explains the tendency of nodes (i.e., individuals) to have similar degrees (i.e., the number of links) to be connected each other in a network of scientists.

Viral Diffusion Speed

In marketing’s diffusion models of new products, Bemmaor and Lee [2002] suggest an individual adoption model based on Gamma shifted Gompertz distribution which integrates the heterogeneity of the propensity to buy with the shape parameter γ. The more heterogeneous the diffusion of products becomes (i.e., the smaller value of γ), the flatter the unit adoption rate becomes over the period of the diffusion like the exponential distribution with the constant rate of the adoption. If it gets homogeneous, the adoption rate accelerates once it takes off. In empirical research, Gatignon and Robertson [1985] find the positive correlation between the homogeneity and the speed of diffusion: the more homogeneous the social system is, the faster the diffusion of new products gets. However, in their recent meta analysis on the diffusion speed with the tie characteristics, Van den Bulte and Stremersch [2004] offer a counter example; system homogeneity has a negative impact on the speed of diffusion. At the aggregate level, they show that the relative weight of imitation (q/p ratio), representing the speed of the diffusion, becomes smaller with higher levels of country’s homogeneity (explained by “Gini” index) gets larger. They argue that the greater a society's heterogeneity the more individuals wish to be seen as different with a subsequent higher uptake of innovations.

Overall, the speed in the diffusion models actually consists of three elements: response speed, generated volume of referral, and response rate. As the performance of at least one of these elements improves, the diffusion of products speeds up. Compared to previous diffusion studies that explore the heterogeneity of overall market, we assess its impact at the individual level.

When an individual is contacted by the viral, the response speed is one of the key elements that determine the total volume of the viral. The faster the speed, the larger the total volume becomes in a fixed time period. If the individual-level diffusion is analogous to that of aggregate level, as Van den Bulte and Stremersch [2004] found in their meta analysis, the recipient may respond faster to the sender of heterogeneous characteristics than to one of similar characteristics . Therefore, we present our first hypothesis about the effects of the tie characteristics on the speed of the viral diffusion.

> ***H1 (The Speed of Viral Diffusion Hypothesis):*** *The more heterogeneous the tie characteristics are, the quicker the receiver responds to the sender's contact.*

Viral Diffusion Volume

To ascertain the volume of the viral, previous studies offer us a range of viral volumes captured in the form of word-of-mouth across various business contexts. According to a marketing research company, TARP (now e-Satisfy), dissatisfied customers generate larger volume of word-of-mouth (i.e., nine to ten people) than satisfied customers (i.e., four to five people) in the case of Coca Cola complaint handling [Rosen 2000]. In customer satisfaction index surveys, on average, respondents talked to 9.5 people annually in Sweden and eight people in United States [Anderson 1998]. In the case of Dell USA in 2002, promoters who responded 9 or 10 on a question of willingness for recommendation on a 10-point scale talked to eight people on average but detractors who responded from 1 to 6 talked to five people on average [Reichheld 2006]. The limitation of the findings is that the volume of the viral depends on customer experience, not on the tie characteristics. Therefore, we explore another angle in relation to the tie characteristics. In a study of social networks, the tendency of nodes (i.e., consumers in our case) to be connected with similar ones is called as *assortativity* [Newman and Park 2003]. So, we suggest the following hypothesis about the effects of the tie characteristics on the volume of the viral diffusion.

> ***H2 (The Volume of Viral Diffusion Hypothesis)****: The more heterogeneous tie characteristics are, the larger the viral the receiver generates.*

III. RESEARCH METHODOLOGY

Today, the Internet provides unprecedented opportunities to collect data involving more observations (e.g., of human decisions, prices, products, etc.) with lower costs, fewer strict assumptions, and more realism for empirical research involving technology and marketing-related phenomena. So, quasi-experimental designs in IS and e-commerce can take advantage of natural experiments in the real world and enable a researcher to distinguish between outcomes associated with different levels of impacts, access or use of IT. A *quasi-experiment*, also called a *natural experiment*, is a type of quantitative research design conducted to explain relationships and clarify why certain events happen [Campbell 1988]. The quasi-experimental method puts an emphasis on the understanding of the source of variation used to estimate key parameters [Meyer 1995]. In this section, therefore, we provide the description of our viral marketing data in the context of quasi-experiments and applied empirical models—a Cox regression model for the viral speed and a regression model based on the negative binomial distribution for the viral volume.

Quasi-Experimental Data

We conducted our study based on an online referral campaign organized by an online travel agency located in France. The campaign went on from May 10 to June 8 in 2004 (i.e., 30 days). In this campaign, the online travel agency intended to collect e-mail addresses of participants for future direct marketing activities. The whole process of the campaign went on a specifically designed Web site where participants visited and registered themselves by providing their personal information such as name, birth date, gender, and mailing address including the zip code. Upon registration, the participants obtained a chance to play an online game to win a free holiday package for two people. In case of failure, another chance to play was given if they provided either an acquaintance's name and e-mail address or additional personal information. The e-mail address question was asked in rotation with personal information questions. Once a new e-mail address was collected, an invitation e-mail to participate in the campaign was sent automatically to the person with the referrer's name. At the beginning of the campaign, various promotional sources such as banner advertising on paid and affiliated Web sites and direct e-mails using rental lists were mobilized to attract Web surfers to the site. This agency collected 231,106 registered e-mail addresses with profile information and 218,713 of them were valid after a rigorous data cleansing process. 30.3 percent of them (i.e., 66,331) were collected purely through the viral upon invitation. Participants came from 185 different countries with 83 percent of them were from France. Figure 1 illustrates the frequency of registered participants during the campaign period. The main participants came from affiliated Web sites. This company sent e-mails from its own customer list in the beginning of the campaign and sent rental e-mails later on day 17 to increase the number of participants.

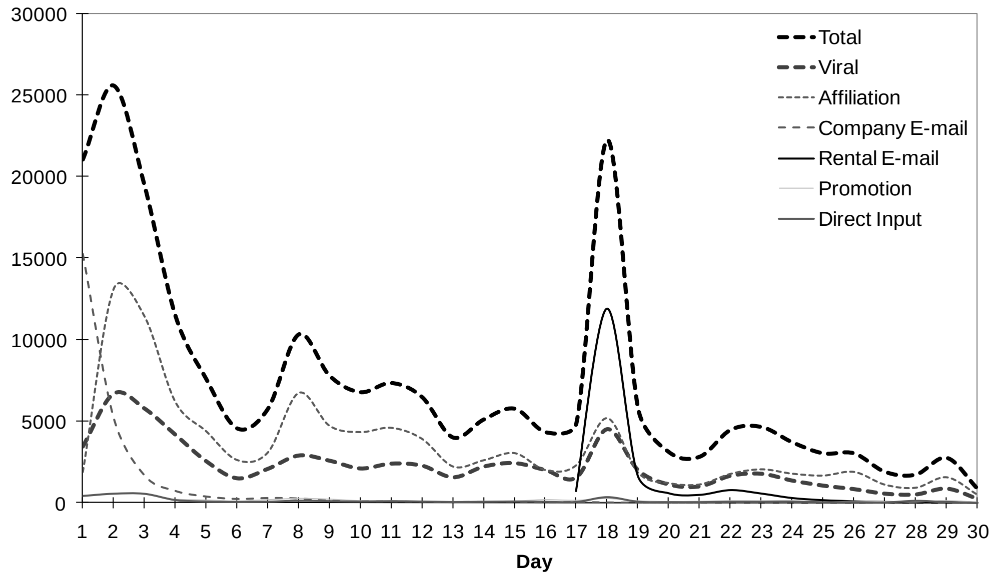

Figure 1. The Number of Registered E-Mails by Promotion Channel

Figure 2 illustrates the frequency of registered participants on a logarithm scale grouped by the wave of registration. Participants in wave 0 are those who got registered without a viral invitation. Participants in wave 1 are those who got registered upon the invitation from participants in wave 0. Therefore participants in wave 1 through 3 are viral-generated ones. The time-series pattern of each wave is almost identical, which means that the timing of registration did not influence the participant's behavior of e-mail sending and response. Tagging technology was applied to the invitation card for the visitors generated by word-of-mouth to be counted systematically. A total of 472,754 invitation e-mails were generated by registered participants.

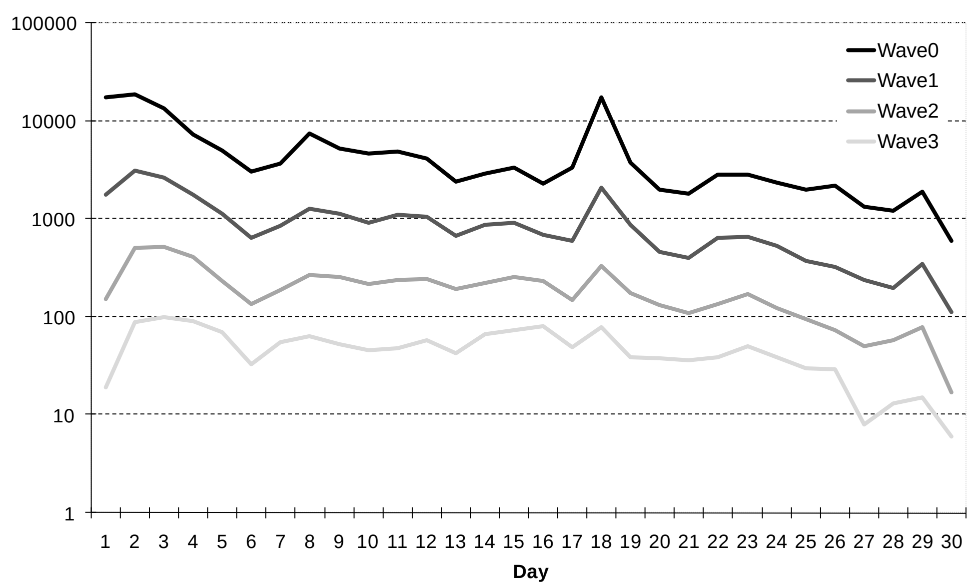

Figure 2. The Number of Registered E-Mails by Registration Wave

This viral campaign data set consists of two data tables: one, the registration recording the sender's personal information as well as registration time, and the other, the viral recording the sent time and the receiver of the invitation e-mail. We combine these two data tables to create a single data file containing the profile and invitation e-mail records of the sender and the receiver, which enables us to check whether the degree of similarity in terms of certain attributes may affect performance indicators of viral diffusion. Compared to previous studies of Brown and Reingen [1987] and Rogers [1995] who measured the degree of homophily/heterophily as a dichotomous variable by counting the number of matching attributes such as age, gender, education, and occupation, we compute three variables to represent the degree of homophily/heterophily: *gender difference* (binary), *age difference* (continuous), and *distance* (continuous) between the sender and the receiver. The age difference variable is measured as the absolute difference between the sender and the receiver's ages. The distance variable is computed by using the position information (longitude and latitude) based on their 5-digit zip code. Therefore, distance is zero if both sender and receiver reside in the region having the same zip code. In our combined data set, 30,035 dyads of sender and receiver with their residence in the metropolitan areas of France (i.e., residents in overseas such as Guadeloupe and Martinique are excluded because it causes biases when computing the dyad distance) remain with complete profile and e-mail information. The number of observations is reduced substantially because all of the non-viral initiated participants (152,400) are removed from the list as they are missing a sender and because of low registration rate of invitees (registration rate of 15.1 percent).

The volume of the viral is measured by the number of e-mail addresses provided by the participants. On average, the participants provided 1.37 e-mail addresses with the maximum of 15. Table 1 shows the frequency distribution of provided e-mail addresses.

Table 1. Frequency Table of Provided E-Mails

Number of E-Mails	Frequency	Percent (%)
0	13,666	45.50
1	7,559	25.17
2	3,530	11.75
3	1,923	6.40
4	1,048	3.49
5	723	2.41
6	458	1.52
7	328	1.09
8	239	0.80
9	179	0.60
10	127	0.42
11	117	0.39
12	93	0.31
13	32	0.11
14	9	0.03
15	4	0.01

Unlike the findings of Leskovec et al. [2007], no cascade effect was found in our data set because the average number of e-mails sent remains similar across the different waves (wave 0: 1.39, wave 1: 1.32, wave 2: 1.30, wave 3: 1.29, wave 4 and over: 1.30, with p-value=0.086). The speed of response is measured in hours based on the elapsed time between the reception time of the invitation e-mail from the sender and the time of registration by the receiver. On average, it took 50.28 hours (i.e., 2.1 days) with the minimum of 0 and the maximum of 692.16 hours (i.e., 28.8 days). Table 2 provides the characteristics of registered participants.

Table 2. Characteristics of Registered Participants

	Minimum	Maximum	Mean	Std. Dev.
Sender's Age	10.13	79.94	34.45	11.96
Receiver's Age	10.04	79.95	33.45	12.28
Age Difference	0.00	58.65	7.76	9.70
Distance (Km)	0.00	1214.46	81.56	173.28
Gender	**Female**	**Percent (%)**		
Sender	20,141	67.06		
Receiver	19,248	64.09		
	Same Gender	**Percent (%)**		
Gender Difference	17,968	59.82		

Empirical Models

In order to measure the speed of the viral diffusion in online referral campaign, we use the elapsed time between the time of invitation e-mail sent by the sender and the time of registration by the receiver. We also use the number of provided e-mail addresses as the proxy of the volume of the viral diffusion. Next, we provide specific empirical models applied for our data analysis.

Viral Speed (The Speed of Response)

The measure of the response speed starts from zero with the response timing having a similar characteristic of event occurrence. Therefore, we apply a Cox regression model based on non-parametric hazard rate. This model incorporates the impact of explanatory variables on hazard rate by taking an exponential form but it allows for an unspecified form for the baseline hazard function [Cox 1972]. For the Cox regression model, the hazard rate of individual (receiver) i, $h_i(t)$, at duration t is given as $h_i(t) = h_0(t)\exp(\beta X_i)$ where $h_0(t)$ is the baseline hazard and β is the impact of the explanatory variables.

Viral Volume (The Number of Provided E-Mail Addresses)

The nature of the viral is measured by the number of provided e-mail addresses as a count variable of zero and positive integers. Instead of the linear regression, we apply a form of regression based on a distribution capturing the nature of the count variable. As Table 1 shows, it has a right skewed distribution with inflation on zeros compared to a typical Poisson distribution. Therefore, we decided to apply a regression model based on the negative binomial distribution (NBD), a mixture of the Poisson distribution with the gamma distribution, providing a high degree of heterogeneity to the mean (λ) of the Poisson distribution, where γ is the shape parameter and α is the scale parameter with the mean, γ/α and the variance, $\gamma/\alpha+\gamma/\alpha^2$.

$$P(X=x)=\int \frac{\lambda \exp(\beta Z)^{x} e^{-\lambda \exp(\beta Z)}}{x!}\frac{\alpha^{\gamma}\lambda^{\gamma-1}e^{-\alpha\lambda}}{\Gamma(\gamma)}d\lambda$$

$$P(X=x)=\frac{\alpha^{\gamma}\exp(\beta Z)^{x}}{x!\Gamma(\gamma)}\int \lambda^{x+\gamma-1}e^{-(\alpha+\exp(\beta Z))\lambda}d\lambda$$

$$P(X=x)=\frac{\alpha^{\gamma}\exp(\beta Z)^{x}}{x!\Gamma(\gamma)}\frac{\Gamma(\gamma+x)}{(\alpha+\exp(\beta Z))^{\gamma+x}}\int \frac{(\alpha+\exp(\beta Z))^{\gamma+x}\lambda^{x+\gamma-1}e^{-(\alpha+\exp(\beta Z))\lambda}}{\Gamma(\gamma+x)}d\lambda$$

$$P(X=x)=\frac{\Gamma(\gamma+x)}{\Gamma(\gamma)x!}\frac{\alpha^{\gamma}\exp(\beta Z)^{x}}{(\alpha+\exp(\beta Z))^{\gamma+x}}=\frac{\Gamma(\gamma+x)}{\Gamma(\gamma)x!}\left(\frac{\alpha}{\alpha+\exp(\beta Z)}\right)^{\gamma}\left(\frac{\exp(\beta Z)}{\alpha+\exp(\beta Z)}\right)^{x}$$

The impact of explanatory variables ('Z') on the number of provided e-mail addresses is assessed through the exponential function. Its mean is proportional to $e^{\beta Z}$, $E(X) = \frac{\gamma e^{\beta Z}}{\alpha}$

IV. RESULTS

We used Stata 9.0 (www.stata.com) to estimate the empirical models we developed in the previous section. We next provide empirical results which we found from the analysis of our suggested models.

Viral Speed (The Speed of Response)

In consistent with Hypothesis 1, the speed of the viral provides quite different results compared to the viral volume. The larger the difference between the sender and the receiver, the quicker the event (i.e., response) occurs. Therefore, the response time shortens as the characteristics of the sender and the receiver become heterogeneous. In particular, the impact of the distance is substantial as one unit increase in the log of the distance raises 3.2 percent of the hazard rate. If a tie is separated by the distance of 100 Km, the response becomes 14.7 percent quicker (i.e., ln(100) x 0.032) than a tie located at the same zip code. As for the impact of the receiver profile on the speed of response, younger and male receivers tend to respond faster than older and female ones (the effect of age is significant at 10 percent). Additionally, another factor, the number of e-mail addresses provided by the receiver, " Number of e-mail addresses by receiver" in Table 3, also affects the response speed which may explain the intrinsic motivation of providing e-mail. If the receiver is a heavy viral generator, he or she tends to respond quickly. Therefore, our analysis supports *Hypothesis 1 The Speed of Viral Diffusion Hypothesis*, which parallels the findings of Van den Bulte and Stremersch [2004] who explained the relative increase of viral impact (explained by q/p ratio, the relative weight of imitation (q) factor compared to innovation (p) in the case of enhanced heterogeneity of consumers.

Table 3. Viral Speed (Response Time)

	Coeff.	S.E.	Wald	Sig.
Age	-0.001*	0.000	2.797	0.094
Gender (male=1)	0.037**	0.013	8.473	0.004
Gender difference	0.088**	0.012	50.749	0.000
Age difference	0.001**	0.001	4.793	0.029
ln (Distance+1)	0.031**	0.003	151.271	0.000
Number of e-mail addresses by sender	-0.003	0.002	1.781	0.182
Number of e-mail addresses by receiver	0.017**	0.003	36.983	0.000

Note: Significance levels: ** = $p < 0.05$, * = $p < 0.10$.

Viral Volume (The Number of Provided E-Mail Addresses)

As reported in Table 4, the value of the shape parameter, γ, of the negative binomial distribution is smaller than 1, which indicates the distribution of the mean of provided e-mail addresses is very heterogeneous without having a mode. This is mainly due to a large number of receivers who did not provide any e-mail address. As explanatory variables are added one by one, the negative binomial distribution demonstrates its stability in terms of the distribution shape in a range of 0.712 to 0.725. To test whether the impact of explanatory variables is significant or not, we proceeded the log-likelihood ratio (LR) test of each step where the threshold of the LR is 3.84 at the confidence level of 95%, χ^2 (d.f.=1, α=5%).

The characteristics of the receiver affect the viral volume significantly as shown in Table 4. Young participants tend to provide more e-mail addresses than old participants and females more than male participants. Different from the result of the viral speed, none of the differences in the tie characteristics such as gender, age and distance has impacts on the viral volume. Only the viral behavior of the sender affects positively that of the receiver (See Table 4). It means that a receiver who was invited by a heavy viral-generating sender tends to send more e-mail address to others. This phenomenon is similar to the assortativity rule in social networks where the nodes (i.e., individuals) with many edges (i.e., links) tend to be connected to each other. Catanzaro et al. [2003] find that scientists who have written papers with many colleagues tend to be connected with others (as a co-author) having similar characteristics in a scientist network. In a similar manner, we expect the same effect in case of the viral volume: heavy viral generators are likely to be connected to each other. Therefore, we conclude that *Hypothesis 2 The Volume of Viral Diffusion Hypothesis*, is not supported and the assortativity rule in terms of volume is confirmed by our findings.

Table 4. Viral Volume (The Number of Provided E-Mail Addresses)							
Gamma	0.712	0.722	0.722	0.722	0.722	0.722	0.725
Alpha	0.521	0.376	0.382	0.384	0.385	0.387	0.422
Sum LL	-48167.17	-48075.48	-48069.52	-48068.93	-48067.77	-48067.64	-48042.71
LR		183.39	11.92	1.18	2.32	0.26	49.85
Age		-0.010**	-0.009**	-0.009**	-0.010**	-0.010**	-0.009**
Gender (male=1)			-0.057**	-0.063**	-0.062**	-0.061**	-0.065**
Gender difference				0.020	0.019	0.019	0.024
Age difference					0.001	0.001	0.001
ln (distance+1)						0.002	0.000
Number of e-mail addresses by sender							0.020**

Note: Significance levels: ** = $p < 0.05$, * = $p < 0.10$.

V. DISCUSSION AND CONCLUSIONS

In conclusion, we present the overall findings and implications of this research, consider some of its limitations, and identify a number of future directions for valuable research on this topic that will yield new managerial knowledge about the online viral diffusion.

Findings

In this paper, we examine two key indicators—viral diffusion speed and volume—determining the overall performance of the viral diffusion process from the perspective of the tie characteristics at the individual level. Our analysis results reveal that the tie characteristics influence the volume and speed of the viral diffusion in different patterns. In previous social network research, the notion of the tie strength and the tie similarity called homophily has been considered as synonymous [Gatignon and Robertson 1985; Rogers 1995]. The impact of weak tie on the diffusion of information was proposed by Granovetter [1973], and Van den Bulte and Stremersch [2004] proved the positive impact of consumer heterogeneity on the diffusion through their meta analysis.

As a consequence, we assume that the tie heterogeneity may lead to a high performance of the viral diffusion. We decompose the diffusion into its response speed and volume. Based on three categories—age, gender, and distance, the degree of the tie heterogeneity was measured and it positively affects the speed of response but no significant impact on the volume. As a result, we conclude that the tie heterogeneity affects the performance of the viral diffusion mainly by speeding up the response of recipient.

Implications

We broaden the research on the viral diffusion supporting Van den Bulte and Stremersch's [2004] recent meta-analysis. Our findings open a venue for the diffusion research to examine two probable explanations. The one question is whether the sequence of contacts among individuals may influence the diffusion speed even though the aggregate level heterogeneity remains the same. To illustrate, we use an example of four people: two men and two women. If a piece of information is diffused in a sequence of man → woman → man → woman, its diffusion should be faster than another sequence of man → man → woman → woman because our results show that the heterogeneity of dyads accelerates the diffusion. It shows the limitation of explanation given by previous diffusion studies that explore the impact of heterogeneity at the aggregate level. The difference of the diffusion performance by the nature of sequence implies the importance of information that the firm tries to diffuse. If a piece of information can attract the attention of recipients having different characteristics, it can be diffused faster than others even though the aggregate level heterogeneity remains the same. At the organizational level, Hansen [1999] explains indirectly the speed of information diffusion between working units where the tie strength interacts with the complexity of knowledge. Having weak ties between working units speeds up the viral diffusion when knowledge is not complex, but slows down when knowledge is transferred in highly complex formats. The other question is a potential gap between the tie similarity and strength measures. In spite that most of the previous studies consider these concepts synonymous, the tie similarity may not directly lead to the tie strength. Especially in previous studies, the tie strength was a structural issue like the case of the mobile telecommunications network in which the tie strength was measured by the duration of conversation [Onnela et al. 2007]. In our paper, the similarity of the tie is not related to the network structure. It describes just the degree of homogeneity between the sender and the receiver.

In the case of the viral volume, a positive impact is found only for the viral volume of the sender (i.e., "Number of e-mail addresses by sender" in Table 4). However, no impact of the differences in profile characteristics (i.e., demographically and geographically) of the tie appears. This is similar to the previous findings about heavy viral-generating individuals called "market mavens" defined as those who have a high propensity to provide general shopping and marketplace information [Feick and Price 1987]. Feick and Price [1987] provide scales distinct from demographical information to measure the tendency of being market maven. Rosen [2000] classify well-connected individuals called *hubs* into four different types depending on the number of links (e.g., regular or mega) and the quality of diffused information (e.g., expert or social). Any of major criteria is related to demographical characteristics. The notion of assortativity related to viral-generating behavior supports our findings [Catanzaro et al. 2004].

Based on our findings, managers need to pay attention to the following points in order to maximize the performance (e.g., the total volume of viral in a given time period) of viral marketing campaign. As heavy viral generators tend to be connected to each other, it is crucial to pinpoint heavy viral generators in the early stage of the campaign. If it is difficult to select them *a priori*, one can design a reward scheme (e.g., doubling the chance to play as one provides another e-mail address) that can be more attractive to heavy viral generators than to lighter ones. As for the speed of viral diffusion, our findings suggest that a tie having different characteristics tends to respond quicker upon invitation e-mails. To speed this up, it is recommended to provide rewards that can attract more heterogeneous ties (e.g., a gift to a couple living a remote place, a gift to old/young generation) than homogeneous ones as well as rewards that can be easily transferable to people having different demographic/geographic characteristics.

Limitations and Future Research

Our research provides the following directions for future research based on the limitations of this study. First, it is necessary to analyze the impact of the tie structure and characteristics on the viral diffusion at the same time. Our tie characteristics approach is limited to assess the structural impact as our data set is from a single viral campaign event from which it is limited to assess the tie strength based on the structure of social networks of the senders and the receivers. Second, our analysis results from a single campaign, and therefore more research need to be conducted to generalize our findings. Third, due to imperfect randomization and lack of control, our methods based on quasi-experiments may suffer from internal validity in that there may be several competing hypotheses with the experimental manipulation that can act as explanations for the observed results. Therefore, either lab experiments or empirical results from multiple events are recommended to validate our findings. Finally, as we focus on the impact of tie characteristics on the volume and speed of the viral diffusion, we do not try to simulate methods to demonstrate the impact of changing the conditions on the total viral volume. It would be interesting to simulate the whole process of the viral diffusion by changing conditions such as initial seeding volume and seeding target characteristics in relation to reward design.

ACKNOWLEDGMENTS

The authors thank Thibault Munier, the managing director, of '1000Mercis.com' for providing campaign data. This work was supported by a Research Grant from Korea University.

REFERENCES

Anderson, E. W. (1998). "Customer Satisfaction and Word of Mouth," *Journal of Service Research* 1(1), pp. 5-17.

Bemmaor, A. C. and J. Lee. (2002). "The Impact of Heterogeneity and Ill-Conditioning on Diffusion Model Parameter Estimates," *Marketing Science* 21(2), pp. 209 -220.

Brown, J. J. and H. Reingen. (1987). "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research*, 14(3), pp. 350-362.

Campbell, D. T. (1988). "Quasi-Experimental Designs," in E. S. Overman (ed.), *Methodology and Epistemology for Social Science: Selected Papers*, Chicago, IL: University of Chicago Press, pp. 191-221.

Catanzaro, M., G. Caldarelli, and L. Pietronero. (2004). "Assortative Model for Social Networks," *Physical Review E* 70(037101), pp.1-4.

Chevalier, J. A. and D. Mayzlin. (2006). "The Effect of Word of Mouth on Sales : Online Book Reviews," *Journal of Marketing Research* 43(3), pp. 345-354.

Cox, R. D. (1972). "Regression Models and Life Tables (with Discussions)," *Journal of the Royal Statistical Society Series B* 34(2), pp. 187-220.

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

Elliot, S. (2007). “Faith, Hopewood, and Charity,” *The New York Times*, July 10, 2007.

Feick, L. F. and L. L. Price. (1987). “Market Maven: A Diffuser of Marketplace Information,” *Journal of Marketing* 51(1), pp. 83-97.

Frenzen, J. K. and H. L. Davis. (1990). “Purchasing Behavior in Embedded Markets,” *Journal of Consumer Research* 17(1), pp. 1-12.

Gatignon, H. and T. S. Robertson. (1985). “A Propositional Inventory for New Diffusion Research,” *Journal of Consumer Research* 11(4), pp. 849-867.

Godes, D. and D. Mayzlin. (2004). “Using Online Conversations to Study Word of Mouth Communication,” *Marketing Science* 23(4), pp. 415-428.

Granovetter, M. S. (1973). “The Strength of Weak Ties,” *American Journal of Sociology* 78(6), pp. 1360-1380.

Hansen, M. T. (1999). “The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits,” *Administrative Science Quarterly* 44(1), pp.82-111.

Helm, S. (2000). “Viral Marketing: Establishing Customer Relationships by 'Word-of-Mouse',” *Electronic Markets* 10(3), pp. 158-161.

Herr, P. M., F. R. Kardes, and J. Kim. (1991). “Effects of Word-of-Mouth and Product Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective,” *Journal of Consumer Research* 17(4), pp.454-462.

Kalika, M., N. B. Charki, and H. Issac. (2008). “An Empirical Investigation of E-Mail Use versus Face-to-Face Meetings: Integrating the Napoleon Effect Perspective,” *Communications of the AIS* 22(27), pp. 501-514.

Katz, E. and P. F. Lazrasfeld. (1955). *Social Influences: The Part Played by People in the Flow of Mass Communications*, Glencoe, IL: Free Press.

Lakshmipathy, V. (2007). “Social Computing Goes Mobile,” *Forrester Research*, December 3, 2007.

Leskovec, J., L. A. Adamic, and B. A. Huberman. (2007). “The Dynamics of Viral Marketing,” *ACM Transactions on the Web* 1(1) Article 5, pp. 1-39.

McPherson, J. M. and L. Smith-Lovin. (1987). “Homophily in Voluntary Organizations: Status, Distance and the Composition of Face-to-Face Groups,” *American Sociology Review* 52(3), pp. 370-379.

Meyer, B. D. (1995). “Natural and Quasi-Experiments in Economics,” *Journal of Business and Economic Statistics*, 13(2), pp. 151-162.

Nail, J. (2005). “The Consumer Advertising Backlash Worsens,” *Forrester Research*, January 5, 2005.

Neuborne, E. (2001). “Viral Marketing Alert!” *Business Week*, March 19, 2001.

Newman, M. E. J. and J. Park. (2003). “Why Social Networks Are Different from Other Types of Networks,” *Physical Review E* 68(03612), 1-9.

Onnela, J. P., J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A. L. Barabasi. (2007). “Structure and Tie Strengths in Mobile Communication Networks,” in *Proceedings of National Academy of Science USA* 104(18), pp. 7332-7336.

Parameswaran, M. and A. B. Whinston. (2007). “Social Computing: An Overview,” *Communications of the AIS* 19(37), pp. 762-780.

Reichheld, F. (2006). *The Ultimate Question*, Boston, MA: Harvard Business School Press.

Rogers, E. M. (1995). *Diffusion of Innovations*, *4th edition*, New York, NY: The Free Press.

Rosen, E. (2000). *The Anatomy of Buzz*, New York, NY: Random House.

Subramani, M. R. and B. Rajagopalan. (2003). “Knowledge-Sharing and Influence in Online Social Networks via Viral Marketing,” *Communications of the ACM* 46(12), pp. 300-307.

Van den Bulte, C. and S. Stremersch. (2004). “Social Contagion and Income Heterogeneity in New Product Diffusion: A Meta-Analytic Test,” *Marketing Science* 23(4), pp. 530-544.

Walker, J., S. Wasserman, and B. Wellman. (1994). “Statistical Models for Social Support Networks,” in Wasserman, S. and J. Galaskiewicz (eds.), *Advances in Social Network Analysis*, Thousand Oaks, CA: Sage., pp. 53-78.

Watson-Manheim, M. B. and F. Bélanger. (2007). “Communication Media Repertoires: Dealing with the Multiplicity of Media Choices,” *MIS Quarterly* 31(2), pp. 267- 293.

Watts, D. J. and S. H. Strogatz. (1998). “Collective Dynamics of ‘Small-World’ Networks,” *Nature* 393, pp. 440-442.

ABOUT THE AUTHORS

Janghyuk Lee is an assistant professor of Marketing at Korea University Business School (KUBS) in Seoul, Korea. He received his Ph.D. degree from ESSEC, France and taught at University of Reading (UK), HEC School of Management (France), and KAIST Business School (Korea) before joining KUBS. His research focuses on the marketing impacts on customer equity and individual level information diffusion in large-scale social networks. His research work has been published at *Journal of Services Marketing, Marketing Science, Recherche et Applications en Marketing, and Seoul Journal of Management.*

Jong-Ho Lee is an assistant professor of Marketing at Korea University Business School (KUBS) in Seoul, Korea. He earned his Ph.D. degree in Strategy and Marketing from University of Cambridge, UK and his MS degree in Marketing Research from the University of Wisconsin-Madison. His research interests focus on understanding performance of innovation especially in high tech companies. His research papers have appeared in *Industrial Marketing Management* and conferences such as IPDM (13^{th} & 14^{th}) and ICIS (28^{th}).

Dongwon Lee is an assistant professor of Information Systems at Korea University Business School (KUBS) in Seoul, Korea. He earned his Ph.D. degree in MIS from University of Minnesota and his MS degree in MIS from the University of Arizona. His research interests include pricing strategies in e-commerce, competition in telecommunications market, knowledge management and sharing, and adoption of new technologies. His research papers have appeared in *Journal of Management Information Systems*, *International Journal of Electronic Commerce*, *Information Systems Frontiers*, *Electronic Markets*, *Journal of Global Information Technology Management*, *ACM Crossroads*, and conferences such as ICIS, HICSS, Jerusalem Summer School in Economic Theory, and INFORMS CIST, and INFORMS Marketing Science Conference.

Communications of the Association for Information Systems

ISSN: 1529-3181