Recherche et Applications en Marketing, vol. 22, n° 3/2007

Modeling the social structure of online discussion groups: implications for controlling viral marketing

Alexandre Steyer
University Paris I - Sorbonne
PRISM

Renaud Garcia-Bardidia
IUT of Épinal, University of Nancy 2
CEREFIGE

Pascale Quester
University of Adélaide
PRISM and FACIREM

ABSTRACT

This research examines the process by which information diffuses within newsgroups on the Internet. Our results empirically demonstrate that these newsgroups are scale-free networks where the potential for information dissemination is important, albeit somewhat unpredictable. This leads us to reconsider the econometric foundations of forecasting methods typically used by marketers.

Key words: Internet, word-of-mouth, viral marketing, discussion groups, power law, scale-free network.

INTRODUCTION

Among the Internet's many marketing potentialities, viral marketing has undoubtedly sparked the greatest interest. It recognizes the importance of searching for information and interacting socially via the Internet, and has progressively evolved from being the subject of managerial experiments to a set of more or less stable techniques. With more effective consumer relations as its goal, it has been heralded by some as a paradigm shift (Godin, 2001). However, it is not an easily defined concept. Indeed, it suffers from a lack of academic formalization, with still relatively little research being carried out in this field compared to the existing volume of managerial literature.

Nevertheless, by studying the various approaches, it is possible to identify a certain number of common denominators. According to Bernard and Jallat (2001), one could define viral marketing as: "a marke-

Dr. Pascale Quester, associate member of the PRISM laboratory, was a guest professor during the writing of this paper. She would like to thank the University of Paris I Panthéon Sorbonne for its support during her stay.
The authors can be contacted by e-mail at the following addresses:
alexandre.steyer@univ-paris1.fr; renaud.garcia-bardidia@univ-nancy2.fr; pascale.quester@adelaide.edu.au

ting technique developed for the Internet and meant to encourage the circulation of a message among certain sites or Internet-users in order to exponentially increase its visibility and effectiveness." It would therefore depend upon this medium's specific characteristics, in terms of both the messages sent and interpersonal communications, in order to reach its objectives of exposure and persuasion. We therefore distinguish between active viral marketing, which relies on interpersonal recommendations, and passive viral marketing, which depends on the simple utilization of the product or service.

In the first case – and, indirectly, in the second case, as well – we find the classic principles that are at work in various information dissemination processes, commonly referred to as "word of mouth" (Grégory, 1996). However, they target a new type of opinion leader (Vernette and Florès, 2004), by making use of a variety of interactive situations. E-mail networks, blogs, forums and online communities therefore have become the object of specific targeting strategies (Brodin, 2000) or permission marketing (Godin, 2001). Viral marketing can therefore be considered a contextualized operationalization of certain results of consumer behavior in terms of word of mouth. However, it differs from other marketing techniques in that it proposes to potentially control specific phenomena generally considered uncontrollable (Dye, 2000), and is even more attractive due to its relatively low cost compared to more traditional advertising campaigns (Helm, 2000).

An epidemiological approach would be an appropriate way to test the supposed advantages of viral marketing with regard to exposure, persuasion and control. Indeed, there is a great tradition of models for predicting the spread of marketing innovations (Bass, 1969; Gatignon and Robertson, 1991) suited to elucidating these phenomena on the Internet. However, while this constitutes the starting point for the viral marketing metaphor, the specific characteristics of Internet networks should be integrated into this initial framework. Indeed, recent contributions made in the field of statistical physics indicate that Internet networks, when considered on a large scale, are anything but random and therefore imperfectly explained by the exponential laws that have been used up until now (Barabasi, Albert and Jeong, 2000).

These same researchers have put forward new, more appropriate network structures, in particular, the "scale-free" network structure. These networks, which have been observed on numerous occasions, give rise to distribution laws whose variance is infinite. Table 1 compares these new networks with classic, random networks. We propose to validate, for the first time, the existence of such networks in the context of Internet discussion groups. This would call into question the classic models of dissemination based on the premise of the normality of residuals and therefore of a finite variance.

The choice of discussion groups as the empirical context for this study is based on the specific characteristics of social interactions within these forums, i.e., their spontaneous development, their richness in terms of information exchanged, their frequent focalization on specific product categories, and their particular organization which facilitates the grouping together of consumers based on their affinities, passions or centers of interest. They therefore represent a typical social network associated with online word of mouth. Moreover, the network is observable *in vivo*.

The first section of this article will discuss the importance of a social network's structure for the

Table 1. – Comparison between traditional networks and scale-free networks

	Random networks	**Scale-free networks**
Type of attachment	Purely random	Preferential depending on the number of existing links
Distribution of links	Poisson distribution	Power law with an exponent of β
Variance in the distribution of links	Finite	Infinite if $\beta < 2$
Critical mass	Present	Absent

process of disseminating information. A social network's structural specificities on the Internet will be the subject of an in-depth study, from which will emerge an initial research proposal concerning the distribution in terms of the size of discussion threads in which consumers participate. A mathematical approach will then be proposed, in order to quantitatively describe the structural conditions of the viral contamination process of a population consisting of an Internet discussion group. In particular, a distribution model will be proposed for the number of contributors for each discussion thread. The following section will explain the methodological choices made and, in particular, the empirical context studied and the measures adopted. Finally, the results obtained will be discussed before concluding with their significance for marketing research and the practice of marketing.

SOCIAL NETWORK STRUCTURE AND DISSEMINATION IN MARKETING

The dissemination of innovations, or the social process by which an innovation is spread within a social system via various channels over time (Rogers, 2003), is an important issue in marketing (Goldenburg, Libai and Muller, 2001). Indeed, it affects individual behaviors and, by aggregation, the markets, and can therefore result in quasi-monopolies (David, 1985). It has therefore attracted special attention, in particular in the form of forecast models dating back to the initial work carried out by Bass (1969).

Among the various factors liable to influence dissemination, the form adopted by social networks plays a unique role by favoring a reduction of perceived risk (Rogers, 2003). Indeed, exposure to the adoptive behavior of close relations is an important source of vicarious learning (Bandura, 1969). Consumers' physical or geographic proximity is therefore a basic element of contagion phenomena (Whyte, 1954), thereby creating sequences or "avalanches" of adoptive behavior (Steyer, 2005). Moreover, this observation of the consequences of other people's actions can be accompanied by an exchange of information (through word of mouth), thereby encouraging interpersonal persuasion, spurred on, in particular, by opinion leaders (Lazarsfeld, Berelson and Gaudet, 1948) or by individuals more closely integrated into their particular social network (Coleman, Katz and Menzel, 1966).

Ever since Granovetter (1973), a specific current of the social sciences has revolved around this question, focusing on the relational and structural characteristics of networks and their members, as well as on their consequences (Lazega, 1998). Measuring the strength of relations (Granovetter, 1973) and the individuals' structural similarity (Burt, 1987) or centrality (Freeman, 1979) allowed for an improved understanding of network characteristics and of the positions held which are necessary for the spread of information (Valente, 1995). They allowed for a reexamination of the bases of the initial studies in this field (in particular, Coleman, Katz and Menzel, 1966), calling for a more careful distribution of the respective roles of the contagion and marketing efforts (Van den Bulte and Lilien, 2001). The existence of critical mass (Schelling, 1978), the threshold at which dissemination becomes inevitable (Rogers, 2003), has been the subject of numerous marketing studies. This has resulted in efforts to identify and then target the opinion leaders, in order to instigate the dissemination process (Lazarsfeld, Berelson and Gaudet, 1948; Robertson and Myers, 1969; Ben Miled and Le Louarn, 1994; Burt, 1999; Vernette, 2004).

While, from both a sociological and marketing point of view, important results have been obtained at micro-social and individual levels, their generalization at the macro level remains problematic. This is due, in part, to the "open" character of most relational networks formed by consumers, making it impossible to define stable and precise limits (Johnson-Brown and Reingen, 1987). In addition, problems associated with the collection and analysis of data on a large scale prevented the use of traditional sociometric measurements, replaced instead by declarative – and often biased – measures.

Therefore, it is not surprising that *a priori* random interactions between individuals are implicitly postulated in dedicated models. The probability of contact between any two consumers, whoever they may be, is therefore uniformly distributed within them (Bass, 1969; Frenzen and Nakamoto, 1993). However, it would only seem logical that consumers prefer contacts who share their interests, have a similar profession, live in the same neighborhood, or do the same activities. This "homophily" (Granovetter, 1973) would necessarily infer that they would more easily enter into direct contact with these individuals than they would with complete strangers, thereby rendering such a premise paradoxical. This observation calls for a deepening of the research that has been carried out so far, in order to reconcile the structure of network interactions with the information dissemination process. This issue is all the more relevant in the light of the multiplication of consumer interactions resulting from the emergence of the Internet. Indeed, this new means of communication enables the bringing together of individuals who, while geographically distant, are able to exchange information on the subjects that interest them, via discussion groups or blogs.

DISTRIBUTION OF INFORMATION ON THE INTERNET: STRONG SPECIFICITIES

Since it is based, in part, on broadcasting phenomena, viral marketing must adapt to the specific characteristics of the social networks formed on the Internet. In the case of discussion groups, this is facilitated by the fact that the Internet allows for an unprecedented *in vivo* observation of the word of mouth process (Brodin, 2000), including inside open, large-scale networks. From a theoretical perspective, the emergence of these specific interaction contexts has strengthened the relational dimension of communities and networks, at the expense of a more geographic basis (Wellman and Gulia, 1999). Indeed, numerous physical constraints that once limited the possibilities for contact between consumers have disappeared. It is now possible to meet individuals sharing the same passions outside one's immediate geographic area, especially as these forums are easily identifiable. Homophily is now less restricted by geographic considerations, resting primarily on affinities and, in particular, on the "proximity" of the individual's tastes and centers of interest. Moreover, it is easy to enter or leave these groups, according to the individual's specific need for information or contacts, thereby reinforcing the role of homophily with regard to their use and popularity.

These observations were the subject of an initial modeling of the Internet's component networks by Barabasi, Albert and Jeong (2000). These authors argued that this medium differs from more random networks, based on two simple observations: its growth, from only a few sites in 1960 to several billion today, and the fact that the links between sites could not be described as being equiprobable, given the deliberate nature of the favorite links connecting sites (Barabasi and Albert, 1999). Indeed, it is probable that the amount of time a particular site has been online, as well as the quality of its content and promotion, facilitates its connection to new sites providing a link to it. This principle of preferential attachment has helped explain the observed distribution of the number of hyperlinks per site as following a power law whose coefficient must be less than 3 (Barabasi, Albert and Jeong, 2000). This type of network is called a scale-free network, because the exponent of these power laws is not affected by changes in scale. Furthermore, since this value is less than 3, the variance in distribution is infinite.

These initial results have been more widely generalized by research that has since been carried out in the field of statistical physics. For example, Borgs et al. (2004) and Ebel, Mielsch and Bornholdt (2002) have examined the contacts between groups registered with Usenet (one of the main servers for discussion forums), or within e-mail networks. These networks are made up, in particular, of individuals (rather than website pages) and the results confirm the theory which holds that a power law better explains the distribution of contacts between individuals. More recently, Barabasi and Bonabeau (2003) demonstrated that similar network structures could be observed in different Internet contexts, from the relations between companies operating on the same market to the simultaneous appearance of certain actors in big Hollywood productions.

Nevertheless, to our knowledge, no study of this type has been carried out on discussion-group participation and these results have not led to any development in the study of consumer behavior, even though they would seem to considerably modify our understanding of the dissemination of innovations and preferences via this medium. Indeed, if these properties are true in all contexts, this would call into question the very concept of critical mass and the importance of the number of opinion leaders for any given market – and, incidentally, the strategies targeting these same opinion leaders in order to set off the information dissemination process (Pastor-Satorras and Vespignani, 2000). This would not be without consequence for managers and would perhaps shed light on the subject of viral marketing and its effectiveness, i.e., improving and increasing the control of interpersonal exposure and persuasion.

It would therefore seem that these types of network can be typical of human interactions and, in particular, the processes by which consumers share information on products and services. In general, the number of individuals participating in a given discussion is also likely to be distributed in accordance with the same power law. This constitutes our first proposition, P1:

> PROPOSITION 1: $P(k)$, *the probability of an Internet user participating in k discussions is proportional to* $k^{-\beta}$, with $1 < \beta < 2$.

MODELING THE STRUCTURE OF DISCUSSION GROUPS

In this section, we will attempt to model the process of viral contamination by a piece of information. To do this, we will consider the situation in which a message is first submitted by a given individual, or by a company seeking to instigate a viral marketing process by posting a particular piece of information about its product or product category in a forum. This information will then be discussed by other individuals, whose number will increase over time. When the information dissemination process has stabilized, we will measure the final number of individuals having taken part in the discussion and who have therefore been exposed to the initial information. This number measures the ultimate effectiveness of this type of viral marketing. Our approach is therefore similar to that recommended by Pastor-Satorras and Vespignani (2001) in terms of epidemiology.

In this particular case, the size of the discussion threads – veritable representations of virtual conversations and word of mouth over the Internet – can be measured in terms of the number of individuals actively participating in a given sequence of interpersonal exchanges, often referred to as a "thread". A discussion thread groups together in chronological order all forum posts dealing with a particular subject. Discussion threads are organized according to 1) the answer to a given message or 2) the answer to all of the thread's messages. It would therefore seem possible, in the context of Internet discussion groups, to use thread size to measure the exposure induced by this type of viral marketing. In this article, this measurement will be referred to as n_i, with n representing the number of individuals participating in thread i. An example of a thread is given below, in the context of a movie discussion group (see Diagram 1).

However, an Internet discussion group has structural characteristics which considerably differentiate it from a more traditional, face-to-face discussion group. First, the "initiator" submits an opinion without knowing who he/she is addressing, contrary to an interpersonal discussion, in which every message sent is directed toward one or more specific "receivers". The message submitted online is therefore liable to attract multiple, potential respondents according to its intrinsic interest. In this article, θ will represent the attractiveness of a given message. This figure is determined, for example, by the message's degree of contentiousness, the quantity of information it contains, or its more or less surprising character.

The online discussion then continues with the participation and successive contribution of a series of participants who together create a thread. Another crucial difference that sets this type of discussion apart is the permanent availability of all of the messages posted on a given subject, beginning with the initial message, contrary to face-to-face communication, in which an ongoing conversation can be entered but cannot be reviewed.

The Internet structures interpersonal interactions by acting as an amplifier. Indeed, we find at the outset

Diagram 1

Example of a discussion thread of the group fr.rec.cinema.discussion

(Review of *Snakes on a plane*, David R. Blis, 2006)

1 Francis Kain 4 Sep
2 Fernand 4 Sep
3 Francis Kain 4 sep
4 Fernand 4 sep
5 Francis Kain 4 sep
6 Fernand 4 sep
7 Francis Kain 4 sep
8 Benji 4 sep
9 Francis Kain 4 sep
10 Fernand 4
11 Benji 5 sep
12 Francis Kain 5 sep
13 Benji 5 sep
14 Benji 5 sep
15 Francis Kain 5 sep
16 Mammoud 4 sep
17 Agnès Vajda 4 sep
18 Francis Kain 4 sep
19 Aloneinthedark 8 sep
20 Didier Couderc

Messages are organized in chronological order. They are organized according to the order of responses and recipients . Messages 1 to 12 form a series as well as, messages 14 and 15 or 16 and 17. Message 18 is a new starting point on the same subject, i.e., a message for the whole group and not an answer.

This example of a thread is based on a review of a movie during its opening week, leading to questions,, comments and recommendations.

a certain message of interest θi, which is then relayed to ultimately reach n_i individuals. The goal of our model is to describe this process, analogous to the transmission of a virus within a given population.

Mathematically, n_i is a random variable whose value differs for each thread. Like all random variables, it can be defined by two measures: its expectation and its variance, $V(n_i)$. The second measure is of particular interest to us here, because it allows us to distinguish between a classic random network of finite variance and a scale-free network of infinite variance. We therefore propose to calculate its value, as well as the distribution of n_i. We will then be able to decide between a classic network and a scale-free network to describe Internet discussion groups.

To carry out our modeling, it is necessary to specify the manner in which the θi values vary. In the absence of more precise information on this latent variable, we will suppose that θ is a random variable uniformly distributed between 0 and θ max. The value θ having no influence on the observed quantities, we will suppose – in order to facilitate our calculations – that this latent variable is uniformly distributed between 0 and 1, which defines its measure *a posteriori*.

We also need a hypothesis concerning the n_i variable's evolution over time. We will suppose that the threads emerge at τ random moments, which are uniformly distributed over time (according to a Poisson process). Moreover, once the first message has been generated, we will suppose that the size of the threads will increase according to a time-based power law. This power law's exponent is not the same for all the threads, allowing us to preserve their heterogeneity.

This gives us the following equation:

$$n_i^{(t)} = \left(\frac{t}{\tau}\right)^{\alpha_i} \qquad (1)$$

Two sorts of events are liable to occur within any discussion group. When an Internet user reads the messages, he/she can choose not to write anything or post a message in threads in which he/she has already participated. In this case, there is no contamination; in other words, no extension of group participation to a greater number of persons on a given subject. However, if an individual participates in a thread for the first time, there is contamination. This second scenario can occur at any time and can be influenced by a number of factors, such as the time, date, the emergence of various social elements, etc. In the absence of any additional information, we will use the simplest tool to model the occurrence of this type of event: the Poisson process. In order to simplify the calculations, we will choose the time unit in such a manner that the probability per time unit that a thread is explored and contaminated by a new individual is equal to 1.

Having established this random process, it now remains to model the arrival of new participants in a thread, as well as their distribution among the various threads. Here, we once again find our central argument of preferential attachment – which we have borrowed from Barabasi, Albert and Jeong (2000) – applying it, for the first time to our knowledge, to the context of Internet discussion groups.

A thread is made all the more popular – in other words, it has all the more chance of attracting new authors – when the following two conditions are met:

– the subject of the thread is interesting and stimulating. This results in a higher θ_i value.
– numerous authors participate in the thread. This results in a higher n_i value.

Therefore, the probability that a given thread attracts a new Internet user is proportional both to θ_i and n_i, or to the product of the two, $\theta_i\ n_i$. At this stage, we now have all of the hypotheses we need to construct our model, because we know both how the Internet users arrive in the discussion group and how they choose the threads in which they will participate. It is possible to construct a purely mathematical equation based on these hypotheses. This equation is explained in detail in Appendix A1. It provides us with the distribution of the size of threads and demonstrates that the probability that a given thread is a size n – in other words, that it has n contributors – is inversely proportional to $Ln(n)n^{2.25}$. The size of the threads is therefore distributed according to a quasi-power law, with an exponent of 2.25 and logarithmic correction. Our model therefore results in a theoretical distribution that has no free parameter, a particularly demanding prediction. The following section details the empirical part of this study, whose aim is to test the proposed theoretical distribution in the following manner:

> PROPOSITION 2: *The probability $P(n)$ that n consumers participate in a given discussion thread is inversely proportional to $Ln(n)\ n^{2.25}$ for a sufficient value of n.*

These two theoretical proposals concern specifically Internet discussion groups. They can therefore be empirically verified by observing one or more specific discussion groups. In the following section, we will therefore describe in detail the methodological approach chosen in this study to examine our two propositions.

EMPIRICAL CONTEXT AND METHODOLOGICAL APPROACH

In order to empirically verify our theoretical predictions, we propose to observe the communication process between members of Internet discussion groups. As we already pointed out at the beginning of this article, the emergence of the Internet as a communication tool between consumers, and, more generally, as a source of product information, has provided us with a precious opportunity to observe how information is transmitted between the different members of a social network (Brodin, 2000). It therefore constitutes a perfect laboratory for the *in situ* observation of information contagion linked to viral marketing. Indeed, the Internet provides unprecedented access to real consumer behavior as regards interpersonal communication (Brodin, 2000), thereby facilitating our understanding of information acquisition and behavioral choices (Bucklin et al., 2002). By directly observing the manner in which information is shared and spread in the real context of Internet discussion groups, it becomes possible to identify and measure the path followed by information exchanged between interacting individuals in this specific type of social network.

In order to avoid results that are too specifically linked to a single category of product, we propose to study two online discussion groups, each concerned with a separate product category, which we could expect to generate an intense communication process among consumers, but which could also take a more or less dense and divided form (Gensollen, 2004). The first category is the cellular phone: a tangible, innovative and functional product whose multiple attributes render choice difficult. The second is movies: an intangible, familiar and experiential service, where the opinions of others tend to strongly influence choices (Kruger, 1997).

The reason for choosing discussion forums, among all the various interaction contexts available to consumers on the Internet, resides in these discussion groups' characteristics. In addition to being richer than chat rooms in terms of both the volume and depth of the information exchanged, they have the added advantage – from a marketing point of view – of focusing on consumer passions, which makes them easily identifiable. For this reason, they group together exchanges between novices and experts appropriate for the study of information dissemination in a consumer context (Granitz and Ward, 1996). They therefore represent naturally favorable targets for viral marketing (Brodin, 2000).

Methodological approach

Because it is possible to observe online discussions in real time and to record their sequence, as well as the source of messages posted in these discussion groups, the Internet allows researchers to collect data on much larger social networks than were previously available for study. The data necessary for this study were provided by a Usenet.fr network manager in the form of a set of message files. The discussion groups were chosen together by the authors and expert users. The data obtained were filtered, reconstructed into threads, and organized in the form of an Access database, in order to allow for crossed queries based on the measurements used in this study.

Because it is impossible to know for sure if a consumer has read or is even aware of a given message, we must concentrate our efforts on the emission of information or the individual's deliberate contribution to the discussion group. Table 2 presents various indicators to quantitatively describe the two discussion groups studied. As shown in this table, the movie discussion group generates the greatest volume of communications. With just over 50% more participants than the other group, it generates more than twice the number of messages and discussion threads. In addition, it often generates longer messages than the group dealing with cell phones. However, the two groups are similar in terms of the proportion of messages which remain unanswered, demonstrating that in both cases, the participants construct lasting exchanges (see, in particular, Fisher, Bristor and Gainer, 1996; Galegher, Sproull and Kiesler, 1998). This runs counter to presuppositions concerning online interactions, initially perceived as being simply anonymous exchanges within completely fluid groups.

Choosing the indicators

Social interaction indicators traditionally used to describe the circulation of information include relations, their direction, the strength of the connections, and centrality (Valente, 1995; Rogers, 2003), typically measured by self-reported answers to questionnaires or interviews. One of the most useful benefits of directly observing online discussion groups is the

Table 2. – General profile and indicators for the two online discussion groups

Group	**Cell phones**	**Movies**
Number of days observed	244	245
Number of message contributors	1,257	1,952
Number of messages	8,036	19,964
Number of threads started	1,597	3,025
Number of responses	6,439	16,939
Reponses/messages ratio	.801	.848
Messages with no responses/messages ratio	.046	.045
Number of lines	183,213	666,388

ability to measure the contribution and location of each individual in the group. According to Granitz and Ward (1996), the participants can be classified according to their centrality and their membership in different subgroups, with the expert users being more active and connected, thereby benefiting from more social capital.

We have two theoretical proposals to test. One concerns the distribution of threads by author and the other concerns the distribution of authors by thread. To test the first, we empirically measured the number of threads to which each individual contributed. We then calculated the cumulative inverse distribution and tested its adjustment using a power law on a doubly logarithmic plot (see Figures 1 and 2). For the second, we are interested in the number of authors participating in the various threads making up a discussion forum, because each thread represents a specific conversation on a specific topic dealing with the product concerned. We measured the number of participants for each thread, *n*, and tested the adjustment via a regression of type $Ln[p(n)] + Ln[Ln(n)] = -aLn[n] + b$. The second proposal is therefore validated if *a* does not differ significantly from 2.25. The small threads of 1 or 2 authors were excluded from this analysis (see Figures 3 and 4). This was done in order to keep things simple, without modifying the results obtained.

EMPIRICAL RESULTS AND CONSEQUENCES ON THE DISSEMINATION OF INFORMATION

Figures 1 and 2 show the empirical observations of the number of threads used by authors of messages, and for the two product categories examined in the study. We can observe, in perfect agreement with our first proposition, that the distribution of this indicator follows a power law and for both discussion groups. Our results therefore confirm that in the case of the two product categories examined in this study the number of threads in which cyber consumers take part follows a power law with the exponent $\beta = 2.23$ for the cell phone group and $\beta = 2.15$ for the movies group. This means that the probability of *n* consumers taking part in a given thread is proportional to $n^{-2,23}$ for the phone group and $n^{-2.15}$ for the movie group. Adjustment is excellent with coefficients of determination very close to 1. In keeping with the scale-free networks theory of Barabasi, Albert and Jeong (2000), the exponent is indeed between 2 and 3. The first proposition, P1, is therefore verified empirically.

Figures 3 and 4 present empirical observations concerning the size of thread (measured by the number

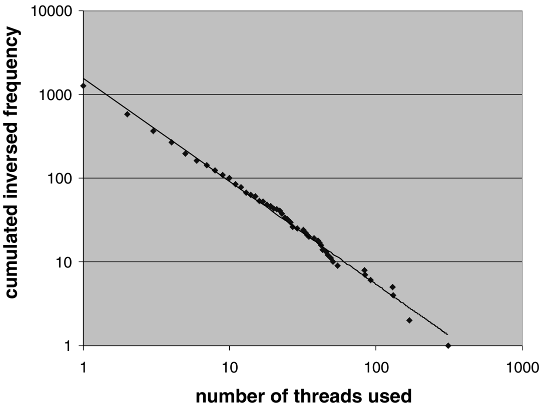

Figure 1. – Threads used by author (cell phone group)

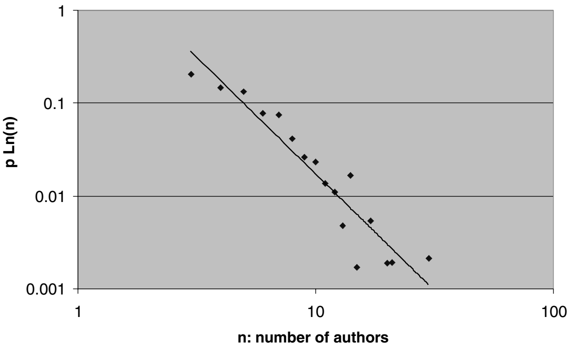

Figure 3. – Distribution of authors by thread (cell phone group)

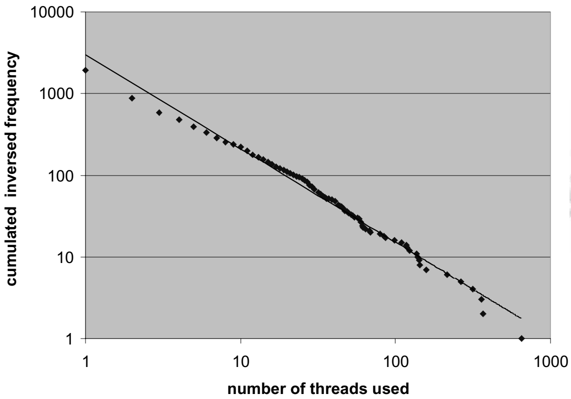

Figure 2. – Threads used by author (movie group)

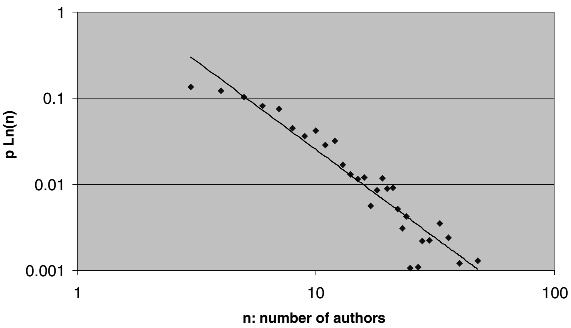

Figure 4. – Distribution of authors by thread (cinema group)

of contributors) for both product categories considered in our study. These figures show, in ordinates, the logarithm of the frequency of threads in function of the logarithm of the number of contributors in abscissas. This type of graph was chosen because it enables to highlight the distributions of probability in power laws as well as the quality of adjustment in both cases (coefficients of determination near 0.9), the resulting graphic representation being a simple straight line. The resulting exponents are 2.5 for the cell phone group and 2.0 for the movie group. These values are not significantly different from the theoretical value forecast by our model (2.25), as confirmed by Student's t test ($t = 1.4$ for the telephone and 1.2 for the movie). Our second proposition, P2, has therefore been validated, the number of contributors per thread following the expected law of distribution (see equation 5 in Appendix A1).

The consequences for the dissemination of information, and the ability of managers to control it, are therefore important. In order to reveal them, the mathematical consequences of the empirical results obtained must first be explained. Indeed, when we wish to calculate the variance in these distributions, we obtain divergent integrals, i.e., integrals of power laws whose primitives do not extend from 0 to infinity (*cf.* Appendix A2). Whether we are referring to authors per thread or threads per author, the theoretical variances are infinite. Yet, managers are used to, and have been trained in, the dominant, Gaussian, econometric model where distribution laws have finite variance and also where, by applying the central limit theorem, they converge toward normal distribution. A practical example of this logic is the ubiquity of confidence intervals used as decision-making tools. With the advent of Internet communications and viral marketing, this type of logic no longer seems relevant. In particular, managers must avoid over-interpreting variances, even if they are extreme, in their online observations. In fact, phenomena such as noise and the appearance of wide variations could in all likelihood be simply the result of random processes rather than underlying phenomena. Acting hastily on erroneous interpretations of variations could lead to counter-productive or totally inappropriate results.

Indeed, since the variances are infinite, the central limit theorem can not be applied. Discussion groups are therefore structured according to such random and various processes that a confidence interval cannot capture them. Managers must therefore expect surprises either way: rapid dissemination or a sudden stop. Contrary to traditional media, managing the dissemination of information is very risky. If the information is positive, our results should encourage managers to use viral marketing in multiple ways to generate large "avalanches", i.e., multiple sources of potential contamination by information. If, on the contrary, the information is negative, it would become almost impossible to predict, and thus master, the dissemination of information.

Another major managerial effect is the absence of critical mass which characterizes scale-free networks like those formed by Internet discussion groups and the subject of our research. This calls into question the central strategies used by managers (according to which the distribution of an innovative product is facilitated by the identification and subsequent targeting of opinion leaders). The key role of these individuals, and their necessary contribution to the dissemination process, is called into question, at least in terms of dissemination of information, which seems essential for the adoption of new products. Aware of the problems managers face in trying to target opinion leaders, this observation can justify a change in strategy in the design of communication campaigns and thus avoid costly and unproductive approaches.

In the specific case of rumors or crises that a manager can face, the absence of critical mass requires a response strategy that is highly different from what is usually prescribed in the event of negative word of mouth. Indeed, instead of just monitoring the evolution of the rumor, subject to spontaneous extinction if it does not reach a critical threshold in the case of word of mouth, and that could actually be spread by a premature reaction, negative *buzz* requires an immediate reaction on the part of the company concerned. If it is not nipped in the bud, negative *buzz* is in fact capable of rapidly taking on proportions that could significantly erode the reputation of a brand and without reaching a critical threshold indicating this outcome. To prevent such occurrences, it is desirable, and even necessary, for managers to continuously monitor the content of online discussions in order to react at the first signs of dissemination of this type of information online.

CONCLUSION: CONTRIBUTIONS, LIMITATIONS AND AVENUES FOR FUTURE RESEARCH

This research has empirically validated, and for the first time, two original forecasts specifically concerning discussion groups. First, like in many social networks, interactions are dominated by strict power laws, in the case of the number of threads per author, or laws adjusted by logarithms, as in the case of number of authors per thread. These two laws, postulated and validated in this article, are not purely mathematical or descriptive. From a theoretical standpoint, these results contribute to a better understanding of the aggregate form taken on by social interactions within discussion groups. Yet, even if these play a significant role in the dissemination process, their typology is too often considered either as purely random (in particular forecast models based on the work of Bass), or as standardized (Frenzen and Nakamoto, 1993). These validated distributions also possess properties with important consequences for viral marketing propositions.

Thus, the idea that this set of techniques produces an "exponential" increase in visibility of "viral" products and services (Godin, 2001) is validated by both the infinite variance of the contribution of messages as well as the absence of induced critical mass (Steyer, Garcia-Bardidia and Quester, 2006). From a managerial standpoint, it would therefore be possible to saturate a potential market, or the target of viral marketing, by exposing it to a given subject via social networks supported by this medium, in this case materialized by discussion groups. On the other hand, viral marketing is different from communication policies based on word of mouth in that it allows for heightened control of both interactions among consumers and their outcomes in terms of broadcasting. In the end, it is only on this condition that it represents an approach that is specifically more efficient. The first point raises important ethical questions, in particular when the plan is to remunerate, even symbolically, "opinion leaders" (Godin, 2001), or organize undercover actions in groups, forums or blogs (Dellarocas, 2006). The absence of critical mass induced by the results obtained indicates, however, total unpredictability of dissemination processes using traditional means, as well as the ambiguous role of opinion leaders, the target of choice for viral marketing campaigns (Vernette and Florès, 2004) since dissemination can achieve its maximum level independent of the individuals involved and their origins (Steyer, Garcia-Bardidia and Quester, 2006).

In the end, the results obtained and their managerial consequences in terms of disseminating information only confirm part of the propositions of viral marketing. This is where the cost (free) of this type of technique, although this is changing due to increased use of agencies specialized in viral marketing, takes on its full significance and confirms the interest and urgent need for a corpus of specific research. However, from a methodological point of view, our study shows how a well-informed manager can develop and collect quantitative measurements that, even if they do not allow him to make real forecasts, could facilitate comparison of discussion groups and identify opportunities for intervening in the threads. In particular, positive information has the potential of spreading infinitely via the Internet, if enough avalanches are triggered. Of course, our results have only been validated within discussion groups on two specific product categories. Any generalization of our conclusions would have to be preceded by similar and repeated studies in other contexts.

In this framework, moving on from an information dissemination model to a study of the dissemination of preferences would be a necessary second step (Godes and Mayzlin, 2004). To do this, a measurement reflecting the valence of the information shared by the participants in the two discussion groups selected would have to be integrated. The positive or negative orientation of messages was not included and we know that this influences the spreading of word of mouth and its efficiency (Richins, 1983), and by extension, the possibility of deformation of message content as the thread grows.

Finally, our results underline the importance of the attraction of the subject in the growth of threads and indicate that a given viral marketing technique cannot *a priori* replace the intrinsic capability of the product to create interest or stimulate opinion, a point that has been raised by Godin (2001). The role played by involvement, which conditions the search for information by consumers and therefore the buying process, and the social role of the different contributors, for example in terms of opinion leadership or virtual ethnography (Kozinets, 2002), merit further study since work on the message or the target represents two alternative strategies to viral marketing.

REFERENCES

Bandura A. (1969), *Principles of behavior modification*, New York, Holt, Rinehart & Winston.

Barabasi A.L. and Albert R. (1999), Emergence of scaling in random networks, *Science*, 286, 5439, 509-512.

Barabasi A.L. and Bonabeau E. (2003), Scale free networks, *Scientific American*, 288, 5, 60-69.

Barabasi A.L., Albert R. and Jeong H. (2000), Scale free characteristics of random networks: the topology of the worldwide web, *Physica A*, 281, 2115, 69-77.

Bass F.M. (1969), A new product growth model for consumer durables, *Management Science*, 15, 215-227.

Ben Miled H. and Le Louarn P. (1994), Analyse comparative de deux échelles de mesure du leadership d'opinion : validité et interprétation, *Recherche et Applications en Marketing*, 9, 4, 23-51.

Bernard G. and Jallat F. (2001), Blair Witch, hotmail et le marketing viral, *L'Expansion Management Review*, 100, www.lexpansion.com/EMR/1448.34.57704.html.

Borgs C., Chayes J.T., Mahdian M. and Saberi A. (2004), Exploring the community structure of newsgroups, Paper presented at the 10th International Conference in "Knowledge, diversity and data mining", Seattle, Washington.

Brodin O. (2000), Les communautés virtuelles : un potentiel marketing encore peu exploité, *Décisions Marketing*, 21, 46-56.

Bucklin R., Lattin J.M., Ansari A., Gupta S., Bell D., Coupey E., Little J.D., Mela C., Montgormery A. and Steckel J. (2002), Choice and the internet: from clickstream to research stream, *Marketing Letters*, 13, 3, 245-258.

Burt R.S. (1987), Social contagion and innovation : cohesion vs. structural equivalence, *American Journal of Sociology*, 92, 6, 1287-1335.

Burt R.S (1999), The social capital of opinion leaders, *The Annals*, 566, 37-54.

Coleman J.S., Katz E. and Menzel H. (1966), *Medical innovation; a diffusion study, Indianapolis*, Bobbs-Merrill Co.

David P.A. (1985), Clio and the economics of QWERTY, *American Economic Review*, 75, 323-337.

Dellarocas C.N. (2006), Strategic manipulation of internet opinion forums: implications for consumers and firms, Working Paper, MIT Sloan School of Management, Cambridge, Massachusetts.

Dye R. (2000), The buzz on buzz, *Harvard Business Review*, 78, 6, 139.

Ebel H., Mielsch L.-I. and Bornholdt S. (2002), Scale free topology of e-mail networks, *Physical Review*, http://www.theo-physik.uni.kiel.de/~bornhol/pre035103.pdf.

Fisher E., Bristor J. and Gainer B. (1996), Creating or escaping community? An exploratory study of internet consumers' behaviour, Eds. K.P. Corfman and J. Lynch, *Advances in Consumer Research*, 23, Provo, Utah, Association for Consumer Research, 178-182.

Freeman L. (1979), Centrality in social networks: conceptual clarification, *Social Networks*, 1, 215-239.

Frenzen J.K. and Nakamoto K. (1993), Structure, cooperation and the flow of market information, *Journal of Consumer Research*, 20, 3, 360-75.

Galegher J., Sproull L. and Kiesler S. (1998), Legitimacy, authority and community in electronic support groups, *Written Communication*, 15, 4, 493-530.

Gatignon H. and Robertson T.S. (1991), Innovative decision processes, in T.S. Robertson and H.H. Kassarjian (Eds.), *Handbook of Consumer Behavior*, Englewood Cliffs, NJ, Prentice-Hall, 316-348.

Gensollen M. (Ed.) (2004), Biens informationnels et communautés médiatées, *Revue d'Économie Politiqu*e, Special Issue "Marchés en ligne et communautés d'agents", March, Paris, Dalloz, 9-40.

Godes D. and Mayzlin D. (2004), Se servir de conversations en ligne pour étudier le bouche-à-oreille, *Recherche et Applications en Marketing*, 19, 44, 89-111.

Godin S. (2001), *Les secrets du marketing viral*, Paris, Maxima.

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

Granitz N.A. and Ward J.C. (1996), Virtual community: a socio-cognitive approach, in K.P. Corfman and J. Lynch (Eds.), *Advances in Consumer Research*, 23, Provo, Utah, Association for Consumer Research, 161-166.

Granovetter M. (1973), The strength of weak ties, *American Journal of Sociology*, 78, 6, 1360-1380.

Grégory P. (1996), *Marketing*, Paris, Dalloz.

Helm S. (2000), Viral marketing: establishing customer relationships by word-of-mouse, *Electronic Markets*, 10, 3, 158-161.

Johnson-Brown J. and Reingen P.H. (1987), Social ties and word-of-mouth referral behaviour, *Journal of Consumer Research*, 14, 3, 350-362.

Kruger A. (1997), Traitement de l'information et décision de choix d'un film : influence des caractéristiques individuelles et de la communication inter-personnelle, Doctoral Dissertation in Management Science, University of Bourgogne, Dijon.

Lazarsfeld P.F., Berelson B.R. and Gaudet H. (1948), *The people choice*, 2nd Edition, DC, Columbia University Press.

Lazega E. (1998), *Réseaux sociaux et structures relationnelles*, Paris, PUF.

Pastor-Satorras R. and Vespignani A. (2001), Epidemic spreading in scale free networks, *Physical Review Letters*, 86, 14, 3200-3203.

Richins M.L. (1983), Negative word-of-mouth by dissatisfied consumers: a pilot study, *Journal of Marketing*, 47, 1, 68-78.

Robertson T.S. and Myers J.H. (1969), Personality correlates of opinion leadership and innovative buying behavior, *Journal of Marketing Research*, 6, 2, 164-168.

Rogers E.M. (2003), *Diffusion of innovations*, 5th Edition, New York, Free Press.

Schelling T.S. (1978), *Micromotives and macrobehaviour*, New York, WW Norton and Co.

Steyer A. (2005), Géométrie de l'interaction sociale : le modèle de diffusion en avalanches spatiales, *Recherche et Applications en Marketing*, 20, 3, 3-20.

Steyer A., Garcia-Bardidia R. and Quester P. (2006), Online discussion groups as social networks: an empirical investigation of word of mouth on the Internet, *Journal of Interactive Advertising*, 6, 2, http://www.jiad.org/vol6/no2/steyer/index.htm.

Valente T. (1995), *Network models of the diffusion of innovations*, Cresskill, NJ, Hampton Press.

Van den Bulte C. and Lilien G.L. (2001), Medical innovation revisited: social contagion versus marketing effort, *American Journal of Sociology*, 106, 5, 1409-35.

Vernette E. (2004), Targeting women's clothing fashion opinion leaders in media planning: an application for magazines, *Journal of Advertising Research*, 44, 1, 90-107.

Vernette E. and Florès L. (2004), Communiquer avec les leaders d'opinion en marketing : comment et dans quels médias ?, *Décisions Marketing*, 35, 23-37.

Wellman B. and Gulia M. (1999), Virtual communities as communities: net surfers don't ride alone, in P. Kollock and M. Smith (Eds.), *Communities in cyberspace*, New York, Routledge, 167-195.

Whyte W.H. Jr. (1954), The web of word of mouth, *Fortune*, 50.

APPENDIX A1 – Mathematical demonstration of the proposed model

Taking into account the necessary standardization of probabilities, we obtain the model

$$\frac{dn_i}{dt} = \frac{n_i\theta_i}{\sum_k n_k\theta_k} \tag{2}$$

Even though the phenomenon studied is discontinuous, we will use a continuous model in order to simplify the calculations. To calculate $\sum_k n_k\theta_k$ we replace the sum by the integral, hence

$$E\left(\sum_k n_k\theta_k\right) = \int_0^1 d\theta \int_0^t d\tau \left(\frac{t}{\tau}\right)^{\alpha(\theta)} \theta$$

This gives, after calculation of the primitives:

$$E\left(\sum_k n_k\theta_k\right) = \int_0^1 d\theta \frac{t}{1-\alpha(\theta)}\theta = \omega t \tag{3}$$

with
$$\omega = \int_0^1 \frac{\theta}{1-\alpha(\theta)} d\theta \tag{4}$$

By using equation 3 in equation 2, we obtain the following differential equation:

$$\frac{dn_i}{dt} = \frac{n_i\theta_i}{\omega t} \tag{2'}$$

This differential equation enables coherency with equation (1) on condition that:

$$\alpha_i = \frac{\theta_i}{\omega}$$

This final relation can be incorporated in equation (4) in order to obtain the following implicit equation:

$$1 = \int_0^1 \frac{d\theta}{\omega/\theta - 1},$$

The solution of which is $\omega = 1.25$.

Using the value of ω we can calculate the distribution of the size of threads in terms of authors. For a given thread i, its initial time t being uniformly distributed, we deduct from equation 1 that n_i is distributed inversely, proportionally to its power $1 + \frac{\omega}{\theta_i}$. By averaging according to the values of θ, we obtain the distribution of the size of threads, n:

$$p(n) \propto \int_0^1 \frac{du}{u} \frac{1}{n^{1+\omega/\theta}}$$

When n is large enough, this integral behaves like:

$$p(n) \propto \frac{1}{n^{2,25}} \frac{1}{Ln(n)} \tag{5}$$

APPENDIX A2 – Calculation of the variance in distributions

1) $\langle k^2 \rangle = \int_1^\infty k^2 \frac{dk}{k^\beta}$ is divergent since $2 < \beta < 3$ and therefore $V(k) = \infty$.

2) In the same way $\langle n^2 \rangle = \int_1^\infty n^2 \frac{dn}{Ln(n)\, n^{2,25}} = \int_1^\infty \frac{dn}{Ln(n)\, n^{0,25}}$ is a Bertrand integral, it is also divergent. So we also have $V(n) = \infty$.