HAL open science

Modélisation de la structure sociale des groupes de discussion sur Internet: Implications pour le contrôle du marketing viral

Alexandre Steyer, Renaud Garcia-Bardidia, Pascale Quester

To cite this version:

Alexandre Steyer, Renaud Garcia-Bardidia, Pascale Quester. Modélisation de la structure sociale des groupes de discussion sur Internet: Implications pour le contrôle du marketing viral. Recherche et Applications en Marketing (French Edition), 2007, 22 (3), pp.29-44. 10.1177/076737010702200303 . hal-02054899

HAL Id: hal-02054899

https://hal.science/hal-02054899v1

Submitted on 25 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modélisation de la structure sociale des groupes de discussion sur Internet : implications pour le contrôle du marketing viral

Alexandre Steyer
Université Paris I - Sorbonne
PRISM

Renaud Garcia-Bardidia
IUT d'Épinal, Université Nancy 2
CEREFIGE

Pascale Quester
Université d'Adélaïde
PRISM et FACIREM

RÉSUMÉ

Nous analysons la structure des interactions informationnelles dans les groupes de discussion sur Internet. L'observation de réseaux sans échelle de variance infinie, validée ici empiriquement pour la première fois, confirme un potentiel de contagion interpersonnelle par l'information important, quoique imprévisible. Les fondements économétriques des méthodes de prévision classiques sembleraient donc mis en question.

Mots clés : Internet, bouche-à-oreille, marketing viral, groupes de discussion, loi de puissance, réseau sans échelle.

INTRODUCTION

Parmi les nombreuses potentialités marketing d'Internet, le marketing viral est sans doute celle ayant fait l'objet du plus fort engouement. Il correspond à la prise en compte de l'importance de la recherche d'information et des interactions sociales se déroulant sur ce média et est progressivement passé du stade d'expériences managériales à un ensemble plus ou moins stable de techniques. Visant une communication plus efficace vers les consommateurs, il a parfois été annoncé comme une révolution paradigmatique (Godin, 2001). Sa définition n'est pourtant pas évidente. Elle souffre en effet d'un manque de formalisation académique, les recherches en la matière étant encore rares comparées au volume de littérature managériale existant.

Dr Pascale Quester, membre associée du laboratoire PRISM, était professeur invitée pendant la rédaction de ce papier. Elle souhaite remercier l'Université Paris I Panthéon Sorbonne pour son soutien lors de ce séjour.
Les auteurs peuvent être contactés aux adresses électroniques suivantes :
alexandre.steyer@univ-paris1.fr ; renaud.garcia-bardidia@univ-nancy2.fr ; pascale.quester@adelaide.edu.au

On peut cependant identifier un certain nombre d'éléments récurrents aux approches proposées. Le marketing viral serait, selon Bernard et Jallat (2001), « *une technique de marketing développée sur Internet et destinée à favoriser la circulation d'un message vers des sites ou des utilisateurs afin d'en accroître de manière exponentielle la visibilité et l'efficacité* ». Il s'appuierait alors sur les spécificités de ce média, que ce soit en termes de messages délivrés ou de communications interpersonnelles, pour atteindre des objectifs d'exposition et de persuasion. On sépare ainsi le marketing viral actif, appuyé par des recommandations interpersonnelles, du marketing viral passif, véhiculé par la simple utilisation du produit ou du service.

Dans le premier cas, mais également indirectement dans le second, on retrouve les principes à l'œuvre de manière classique dans les processus de diffusion de l'information, plus communément nommés bouche-à-oreille (Grégory, 1996). Ils sont cependant appliqués à des leaders d'opinion d'un nouveau genre (Vernette et Florès, 2004), en utilisant des contextes d'interaction variés. Les réseaux d'e-mails, les blogs, les forums ou les communautés en ligne font alors l'objet de stratégies spécifiques de ciblage (Brodin, 2000) ou de marketing de la permission (Godin, 2001). Le marketing viral peut donc être considéré comme une opérationnalisation contextualisée de certains résultats du comportement du consommateur en matière de bouche-à-oreille. Il s'en distingue cependant par la formulation d'un contrôle potentiel d'un phénomène généralement considéré comme incontrôlable (Dye, 2000), et serait d'autant plus intéressant qu'il supposerait un coût moindre que celui nécessaire aux campagnes classiques de communication (Helm, 2000).

Une approche épidémiologique semble pertinente afin de tester ces avantages supposés d'exposition, de persuasion et de contrôle du marketing viral. Il existe en effet une grande tradition de modèles de prévision de la diffusion des innovations en marketing (Bass, 1969 ; Gatignon et Robertson, 1991) propres à éclairer ces phénomènes sur Internet. Si ceci constitue par ailleurs le point de départ de la métaphore du marketing viral, il convient cependant d'intégrer à ce premier cadre les spécificités des réseaux sur Internet. Les apports récents de la physique statistique indiquent en effet que, sur Internet, les réseaux, lorsqu'on les considère à grande échelle, sont loin d'être aléatoires et donc imparfaitement décrits par les lois exponentielles utilisées jusque-là (Barabasi, Albert et Jeong, 2000).

De nouvelles structures plus adéquates de réseaux ont été proposées par ces mêmes chercheurs, plus particulièrement, la structure de réseaux sans échelle. Dans ces réseaux, observés à de nombreuses reprises, apparaissent des lois de distribution dont la variance est infinie. Une comparaison entre ces nouveaux réseaux et les réseaux classiques aléatoires est donnée dans le Tableau 1. Nous nous proposons ici de valider, pour la première fois, l'existence de tels réseaux dans le contexte des groupes de discussion sur Internet. Ceci remettrait en cause les modèles classiques de diffusion fondés sur le postulat de la normalité des résidus et donc d'une variance finie.

Tableau 1. – Comparaison entre réseaux classiques et réseaux sans échelle

	Réseaux aléatoires	**Réseaux sans échelle**
Mode d'attachement	Purement aléatoire	Préférentiel en fonction du nombre de liens existants
Distribution des liens	Loi de Poisson	Loi de puissance avec un exposant β
Variance de la distribution des liens	Finie	Infinie si $\beta < 2$
Masse critique	Présente	Absente

Le choix des groupes de discussion comme contexte empirique pour cette recherche découle des caractéristiques spécifiques des interactions dans ces forums, c'est-à-dire leur développement spontané, leur richesse en termes d'information échangée, leur fréquente focalisation sur des catégories de produits, ainsi que leur mode de fonctionnement facilitant les regroupements de consommateurs sur la base d'affinités, de passions ou de centres d'intérêt. Ils représentent ainsi un cas typique de réseau social associé au bouche-à-oreille en ligne, de surcroît observable *in vivo*.

La première section de cet article traitera de l'importance de la structure des réseaux sociaux dans les processus de diffusion de l'information. Leurs spécificités structurelles sur Internet feront l'objet d'un approfondissement, dont émergera une première proposition de recherche concernant la distribution de la taille des fils de discussion dans lesquels s'engagent les consommateurs. Une approche mathématique sera ensuite proposée, afin de décrire quantitativement les conditions structurales du processus de contamination virale dans la population constituée par un groupe de discussion sur Internet. Il s'agira en particulier de proposer un modèle de la distribution du nombre de contributeurs par fil de discussion. Les choix méthodologiques, et en particulier le contexte empirique étudié et les mesures adoptées, seront justifiés dans la partie suivante. Les résultats obtenus seront enfin discutés avant de conclure sur leurs conséquences pour la recherche et la pratique du marketing.

STRUCTURE DES RÉSEAUX SOCIAUX ET DIFFUSION EN MARKETING

La diffusion des innovations, ou le processus social par lequel une innovation se répand dans un système social via différents canaux au cours du temps (Rogers, 2003), est une problématique importante en marketing (Goldenberg, Libai et Muller, 2001). Elle influence en effet les comportements individuels, et par agrégation, les marchés, pouvant ainsi conduire à des situations de quasi-monopole (David, 1985). Elle a donc fait l'objet d'une attention particulière, notamment sous la forme de modèles de prévision depuis les travaux initiaux de Bass (1969).

Parmi les différents facteurs susceptibles de l'influencer, la forme prise par les réseaux sociaux joue un rôle particulier en favorisant la réduction du risque perçu (Rogers, 2003). L'exposition aux comportements d'adoption de proches est en effet une source importante de l'apprentissage vicariant (Bandura, 1969). La proximité physique ou géographique entre consommateurs va donc être à la base de phénomènes de contagion (Whyte, 1954), créant de ce fait des séquences ou « avalanches » d'adoption (Steyer, 2005). Cette observation des conséquences des actes d'autrui peut de surcroît être accompagnée d'un échange d'information, par le biais du bouche-à-oreille, favorisant la persuasion interpersonnelle, notamment sous l'impulsion des leaders d'opinion (Lazarsfeld, Berelson et Gaudet, 1948) ou d'individus plus intégrés à leur réseau social (Coleman, Katz et Menzel, 1966).

Depuis Granovetter (1973), un courant spécifique des sciences sociales s'est centré sur cette question en se focalisant sur les caractéristiques relationnelles et structurelles des réseaux et de leurs membres, ainsi que sur leurs conséquences (Lazega, 1998). La mesure de la force des liens (Granovetter, 1973), de l'équivalence structurale des individus (Burt, 1987) ou de leur centralité (Freeman, 1979) a alors permis d'affiner la connaissance des caractéristiques des réseaux et des positions tenues nécessaires à la diffusion de l'information (Valente, 1995). Elles ont permis de réexaminer les fondements des études initiales de ce champ (notamment celle de Coleman, Katz et Menzel, 1966), appelant à plus de précaution quant à la répartition des rôles respectifs de la contagion et des efforts marketing (Van den Bulte et Lilien, 2001). L'existence d'une masse critique (Schelling, 1978), seuil au-delà duquel la diffusion devient inéluctable (Rogers, 2003), a fait l'objet de nombreuses recherches en marketing. Elle se traduit par la volonté d'identifier puis de cibler les leaders d'opinion pour démarrer le processus de diffusion (Lazarsfeld, Berelson et Gaudet, 1948 ; Robertson et Myers, 1969 ; Ben Miled et Le Louarn, 1994 ; Burt, 1999 ; Vernette, 2004).

Si les résultats obtenus à une échelle microsociale ou individuelle sont importants tant d'un point de

vue sociologique que marketing, leur généralisation au niveau macro reste problématique. Ceci découle en partie du caractère « ouvert » de la plupart des réseaux relationnels formés par les consommateurs, rendant impossible d'en identifier des limites précises et stables (Johnson-Brown et Reingen, 1987). En outre, les difficultés associées à la collecte et l'analyse de données sur une large échelle ont empêché l'utilisation de mesures sociométriques traditionnelles, remplacées par des mesures déclaratives, souvent biaisées.

Il n'est donc pas surprenant que soient implicitement postulées des interactions entre individus *a priori* aléatoires dans les modèles dédiés. La probabilité de contact entre deux consommateurs, quels qu'ils soient, y est ainsi uniformément distribuée (Bass, 1969 ; Frenzen et Nakamoto, 1993). Il semblerait pourtant logique que les consommateurs privilégient dans leurs contacts ceux qui partagent leurs intérêts, ont une profession comparable, vivent dans le même quartier ou pratiquent les mêmes activités. Cette homophilie (Granovetter, 1973) induirait qu'ils initient plus facilement des communications directes avec ces individus, plutôt qu'avec des inconnus, rendant paradoxal un tel postulat. Ce constat demande un approfondissement des recherches menées jusqu'ici, afin de mettre en rapport la structure du réseau d'interactions avec le processus de diffusion de l'information. Cette problématique est d'autant plus pertinente que les interactions entre consommateurs se sont multipliées du fait de l'émergence d'Internet. En effet, ce nouveau mode de communication permet de mettre en rapport des individus qui, bien que distants géographiquement, peuvent échanger des informations sur des sujets les intéressant, par le biais de groupes de discussion ou de blogs.

LA DISTRIBUTION DE L'INFORMATION SUR INTERNET : DE FORTES SPÉCIFICITÉS

Puisqu'il repose en partie sur des phénomènes de diffusion, le marketing viral doit s'adapter aux caractéristiques des réseaux sociaux constitués sur Internet. La tâche est facilitée dans le cas des groupes de discussion car Internet permet une observation sans précédent du déroulement du bouche-à-oreille *in vivo* (Brodin, 2000), y compris dans des réseaux ouverts à grande échelle. D'un point de vue théorique, l'avènement de ces contextes d'interaction spécifiques a renforcé la dimension relationnelle des communautés et des réseaux aux dépens d'une base plus géographique (Wellman et Gulia, 1999). En effet, de nombreuses contraintes physiques limitant les possibilités de contacts entre consommateurs ont disparu. Il est désormais possible de rencontrer des individus ayant les mêmes passions hors de son voisinage, et ce d'autant plus que les forums les regroupant sont facilement identifiables. Dès lors, l'homophilie n'est plus autant restreinte par des considérations géographiques, mais repose principalement sur les affinités, et plus particulièrement sur la proximité des goûts et des centres d'intérêt. Il est de surcroît facile d'entrer ou de quitter ces groupes, au gré de besoins spécifiques d'information ou de contacts, ce qui renforce le rôle de l'homophilie dans leur fréquentation.

Ces observations ont fait l'objet d'une première modélisation des réseaux constitutifs d'Internet par Barabasi, Albert et Jeong (2000). Ces auteurs suggéraient que ce média différait des réseaux plus aléatoires, sur la base de deux observations simples : sa croissance, de quelques sites à plusieurs milliards depuis 1960, et le fait que les liens entre les sites ne sauraient être décrits comme équiprobables, étant donné la nature délibérée des liens préférentiels entre sites (Barabasi et Albert, 1999). Il est en effet probable que l'ancienneté d'un site, comme la qualité de son contenu rédactionnel ou de sa promotion, facilitent le rattachement de nouveaux sites proposant un lien vers son adresse. Ce principe d'attachement préférentiel a alors permis d'expliquer la distribution observée du nombre d'hyperliens par site comme suivant une loi de puissance dont le coefficient doit être inférieur à 3 (Barabasi, Albert et Jeong, 2000). Ce type de réseaux est appelé réseau sans échelle (*scale free network*) car l'exposant de ces lois de puissance n'est pas sensible aux changements d'échelle. De plus, celui-ci étant inférieur à 3, la variance de la distribution est infinie.

Ces résultats initiaux ont été généralisés plus largement par des travaux conduits depuis dans le domaine de la physique statistique. Par exemple, Borgs *et alii* (2004) et Ebel, Mielsch et Bornholdt

(2002) ont examiné les contacts entre groupes enregistrés avec Usenet (un des principaux serveurs de forums de discussion), ou au sein de réseaux e-mails. Ces réseaux comportent notamment des individus (et non plus des pages de sites) et les résultats confirment la notion qu'une loi de puissance décrit mieux la distribution des contacts entre les individus. Plus récemment, Barabasi et Bonabeau (2003) ont révélé que des structures de réseaux similaires pouvaient être observées dans des contextes différents d'Internet, des relations entre entreprises opérant sur un même marché à l'apparition simultanée de certains acteurs dans les grandes productions hollywoodiennes.

Pour autant, à notre connaissance, aucune étude de ce type n'a été réalisée sur la participation aux groupes de discussion, et ces résultats n'ont fait l'objet d'aucun développement en comportement du consommateur, alors qu'ils semblent de nature à modifier considérablement la compréhension des phénomènes de diffusion des innovations et des préférences via ce média. En effet, si ces propriétés sont vraies dans tous les contextes, alors la notion même de masse critique et l'importance du nombre des leaders d'opinion sur le marché, et incidemment les stratégies en faisant les cibles prioritaires pour déclencher la diffusion de l'information, seraient mises en doute (Pastor-Satorras et Vespignani, 2000). Cela ne serait pas sans conséquence pour les managers et éclairerait peut-être le sujet du marketing viral et de son efficacité, c'est-à-dire l'amélioration et l'augmentation du contrôle de l'exposition et de la persuasion interpersonnelles.

Il semblerait donc que ces types de réseaux puissent être caractéristiques de l'interaction humaine et en particulier, du processus selon lequel les consommateurs partagent les informations concernant produits et services. D'une manière générale, le nombre d'individus participant à une discussion donnée est lui aussi susceptible d'être distribué suivant le même type de loi de puissance. Ceci constitue notre première proposition, P1 :

PROPOSITION 1 : *$P(k)$, la probabilité qu'un internaute participe à k discussions, est proportionnelle à $k^{-\beta}$, avec $1 < \beta < 2$.*

MODÉLISATION DE LA STRUCTURE DES GROUPES DE DISCUSSION

Dans cette section, nous cherchons à modéliser le processus de contamination virale par l'information. Nous nous plaçons dans la situation où un message est initialement émis par un individu donné, ou par une firme désirant initier un processus de marketing viral en envoyant une information particulière dans un forum traitant de son produit ou de la catégorie de produits concernée. Ces informations vont ensuite être discutées par d'autres individus dont le nombre va augmenter avec le temps. Quand le processus de diffusion de l'information est stabilisé, nous mesurons le nombre final d'individus ayant participé à la discussion, et donc exposés à l'information initiale, ce nombre mesurant *in fine* l'efficacité de ce type de marketing viral. Notre approche s'apparente donc à celle préconisée par Pastor-Satorras et Vespignani (2001) en matière d'épidémiologie.

Dans ce cas particulier, la taille des fils de discussion, véritable représentation de conversations virtuelles et du bouche-à-oreille sur Internet, peut être mesurée en nombre d'individus participant activement à une séquence d'échanges interpersonnels, souvent appelés « fils » (*threads*). Ces fils de discussion regroupent chronologiquement l'ensemble des interventions sur un sujet donné. Ils s'organisent selon deux modalités : la réponse à un message donné ou la réponse à l'ensemble des messages du fil. La taille des fils semblerait donc être à même de fournir une mesure de l'exposition induite par ce type de marketing viral, dans le contexte des groupes de discussion. Dans cet article nous noterons cette mesure n_i, le nombre d'individus participant au fil i. Un exemple de fil est donné, dans le cas du groupe cinéma, en encadré (*cf.* Encadré 1).

Un groupe de discussion par Internet comporte cependant des caractéristiques structurelles qui le différencient fortement d'un groupe de discussion traditionnel en face-à-face. La première différence réside dans le fait que l'initiateur émet une opinion sans savoir à qui il l'adresse, contrairement à une discussion interpersonnelle où tout message émis l'est en fonction d'un (ou de) récepteur(s) spécifique(s). Le message émis en ligne est donc susceptible d'attirer des répondants potentiels multiples en fonction de

Encadré 1

Un exemple de fil de discussion du groupe fr.rec.cinema.discussion

(Critique *Des serpents dans l'avion* (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 sep
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 Alone inthedark 8 sep
20 Didier Couderc

Les messages sont organisés par ordre chronologique. Ils s'enchaînent par ordre de réponse ainsi que par destinataire. Les messages 1 à 12 constituent une suite, de même que les messages 14 et 15 ou 16 et 17. Le message 18 constitue un nouveau point de départ sur le même sujet, c'est-à-dire un message posté vers le groupe et non une réponse.

Cet exemple de fil repose sur une critique initiale d'un film après sa sortie, donnant lieu à questionnements, commentaires et recommandations .

son intérêt intrinsèque. On notera dans cet article θ l'attrait de ce message qui est fonction, par exemple, de son degré de polémique, de la quantité d'information qu'il contient ou de son caractère plus ou moins surprenant.

La discussion en ligne se poursuit ensuite par l'intervention et la contribution successive d'une série de participants qui construisent un fil. Une autre différence cruciale de ce mode de discussion est alors la disponibilité permanente de l'intégralité des messages émis sur le sujet, et ce depuis le message initial, contrairement à la communication en face-à-face où une conversation peut être prise en cours sans possibilité de remonter en arrière.

Internet structure les interactions interpersonnelles en jouant un rôle d'amplificateur. En effet, nous trouvons en entrée un certain message, d'intérêt θi, qui est relayé jusqu'à atteindre n_i individus. L'objectif de notre modélisation est de caractériser ce processus, analogue à la transmission d'un virus dans une population donnée.

Mathématiquement, n_i est une variable aléatoire qui prend une valeur pour chacun fil. Comme toute variable aléatoire, elle peut être caractérisée par deux indicateurs, son espérance et sa variance, $V(n_i)$. Ce dernier indicateur nous intéresse ici plus particulièrement car il permet de distinguer un réseau classique aléatoire, à variance finie, d'un réseau sans échelle, à variance infinie. Nous nous proposons donc de calculer sa valeur ainsi que la loi de distribution de n_i. Nous serons alors en mesure de trancher entre un réseau classique et un réseau sans échelle pour décrire les groupes de discussion.

Pour modéliser, il est nécessaire de préciser comment varient les valeurs θi. En l'absence d'information plus précise sur cette variable latente, nous supposons que θ est une variable aléatoire distribuée uniformément entre 0 et θ max. L'unité de θ n'ayant pas d'influence sur les quantités observées, nous supposons, pour faciliter les calculs, que cette variable latente est distribuée uniformément entre 0 et 1, ce qui définit son unité *a posteriori*.

Il est également nécessaire de disposer d'une hypothèse relative à l'évolution dans le temps de la variable n_i. Nous supposons que les fils naissent à des instants aléatoires τ répartis uniformément dans le temps (selon un processus de Poisson). Une fois le premier message généré, nous supposons en outre que la taille des fils augmente suivant une loi de puissance du temps. L'exposant de cette loi de puissance n'est pas le même pour tous les fils, ce qui permet d'en conserver l'hétérogénéité.

On a donc la relation :

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

Deux sortes d'événements sont susceptibles d'intervenir à l'intérieur du groupe de discussion. Lorsqu'un internaute lit les messages, il peut soit ne rien écrire, soit écrire dans des fils où il s'est déjà exprimé auparavant. Dans ce cas, il n'y a pas de contamination, c'est-à-dire pas d'extension de la participation au groupe à un plus grand nombre de personnes sur un sujet donné. Par contre, si un individu participe à un fil auquel il n'avait pas encore participé, il y a contamination. Cette deuxième sorte d'événement peut survenir à n'importe quel moment dans le temps, et être influencé par de nombreux facteurs comme l'heure, la date, la survenue d'éléments sociaux, etc. En l'absence d'information supplémentaire, nous utiliserons l'outil le plus simple pour modéliser la survenue de ce type d'événement, à savoir le processus de Poisson. Afin de simplifier les calculs, nous choisissons l'unité de temps de telle sorte que la probabilité par unité de temps qu'un fil soit exploré par un nouvel individu et qu'il y participe soit égale à 1.

Ce processus aléatoire étant posé, il reste à modéliser les arrivées de nouveaux auteurs dans un fil ainsi que leur répartition entre les différents fils. Nous retrouvons ici notre argument central d'attachement préférentiel que nous empruntons à Barabasi, Albert et Jeong (2000) en l'appliquant, pour la première fois à notre connaissance, au contexte original des groupes de discussion.

Un fil est d'autant plus préféré, c'est-à-dire a d'autant plus de chance de voir arriver un nouvel auteur, que deux conditions sont réunies :

- le fil a un sujet intéressant et stimulant. Ceci se traduit par une valeur élevée de θ_i ;
- de nombreux auteurs animent ce fil. Ceci se traduit par une valeur élevée de n_i.

Ainsi donc, la probabilité pour un fil d'accueillir un nouvel internaute est proportionnelle à la fois à θ_i et à n_i soit au produit des deux, $\theta_i\ n_i$. Nous disposons, à ce stade, de l'ensemble des hypothèses nécessaires pour construire notre modèle puisque nous savons, à la fois, comment les internautes arrivent dans le groupe de discussion et comment ils choisissent les fils où ils vont intervenir. À partir de ces hypothèses, un calcul purement mathématique est possible. Il est détaillé en Annexe A1. Ce calcul aboutit à la distribution de la taille des fils et montre que la probabilité pour qu'un fil ait une taille n, c'est-à-dire soit composé de n contributeurs, est inversement proportionnelle à $Ln(n)n^{2,25}$. La taille des fils est donc distribuée selon une quasi-loi de puissance, d'exposant 2,25, avec correction logarithmique. Notre modèle aboutit donc à une distribution théorique qui ne comporte aucun paramètre libre, prévision particulièrement exigeante. La prochaine section détaille la partie empirique qui vise à tester la distribution théorique proposée de la façon suivante :

PROPOSITION 2 : *La probabilité $P(n)$ que n consommateurs participent à un fil de discussion est inversement proportionnelle à $Ln(n)\ n^{2,25}$ pour n suffisamment grand.*

Ces deux propositions théoriques concernent spécifiquement les groupes de discussion sur Internet, elles peuvent donc être vérifiées empiriquement par l'observation d'un ou plusieurs groupes de discussion spécifiques. Nous détaillons donc dans la section suivante l'approche méthodologique adoptée dans cette étude pour examiner nos deux propositions.

CONTEXTE EMPIRIQUE ET APPROCHE MÉTHODOLOGIQUE

Afin de vérifier de façon empirique nos prévisions théoriques, nous proposons d'observer le processus de communication ayant lieu entre membres de groupes de discussion sur Internet. Comme nous l'avons déjà noté en introduction, l'émergence de l'Internet comme outil de communication entre consommateurs, et plus généralement de recherche d'information sur les produits, fournit une opportunité précieuse d'observer comment l'information est transmise entre différents membres d'un réseau social (Brodin, 2000), et donc constitue un laboratoire parfait pour l'observation *in situ* de la contagion informationnelle liée au marketing viral. Internet permet en effet un accès sans précédent au comportement réel du consommateur en matière de communication interpersonnelle (Brodin, 2000), facilitant ainsi notre compréhension du phénomène d'acquisition de l'information et de choix de comportement (Bucklin *et alii*, 2002). En observant directement la manière dont l'information est partagée et diffusée dans le contexte réaliste des groupes de discussion sur Internet, il devient possible d'identifier et de mesurer le chemin suivi par l'information entre individus interagissant dans ce type particulier de réseau social.

Afin d'éviter des résultats trop spécifiques à une seule catégorie de produits, nous proposons d'examiner le cas de deux groupes de discussion en ligne, ayant trait chacun à une catégorie de produits différente dont on pourrait s'attendre à ce qu'elle génère un processus intense de communication entre consommateurs, mais pouvant prendre une forme plus ou moins dense et partagée (Gensollen, 2004). La première est le téléphone portable, un produit tangible, innovant et fonctionnel dont les multiples attributs rendent le choix difficile. La seconde est le cinéma, un service intangible, familier et expérientiel où l'opinion des autres tend à fortement influencer les choix effectués (Kruger, 1997).

Le choix des forums de discussion parmi l'ensemble des contextes d'interaction disponibles sur Internet pour les consommateurs réside, quant à lui, dans les caractéristiques de ces groupes. Plus riches que les salons de discussion (ou *chatrooms*) en termes d'information échangée tant en volume qu'en profondeur, ils possèdent de surcroît l'avantage d'un point de vue marketing d'être centrés sur des passions de consommateurs les rendant facilement identifiables. Ils regroupent donc de ce fait des échanges entre novices et experts propres à l'étude de la diffusion de l'information dans un contexte de consommation (Granitz et Ward, 1996). Ils s'inscrivent ainsi naturellement parmi les cibles propices au marketing viral (Brodin, 2000).

Approche méthodologique

Du fait de la capacité qu'ont les discussions en ligne d'être observées en temps réel et de l'opportunité qui existe d'enregistrer la séquence ainsi que la source des messages postés sur ces groupes de discussion, Internet permet la collecte de données concernant des réseaux sociaux de bien plus grande échelle que ce que les chercheurs ont pu effectuer dans leurs recherches précédentes. Les données nécessaires à ce travail ont été fournies par un gestionnaire du réseau Usenet.fr sous la forme d'un ensemble de fichiers de messages. Le choix des groupes retenus a fait l'objet d'une réflexion conjointe des auteurs et d'utilisateurs experts. Les données obtenues ont été filtrées, reconstituées en fils et organisées sous la forme d'une base de données Access, afin de permettre des requêtes croisées à la base des mesures utilisées dans cette recherche.

Puisqu'il est impossible de s'assurer qu'un consommateur a bien lu ou pris connaissance d'un message, nous devons concentrer nos efforts sur l'émission de l'information ou la contribution délibérée de l'individu au groupe de discussion. Différents indicateurs sont présentés dans le Tableau 2 pour décrire quantitativement les deux groupes de discussion inclus dans notre étude. Comme le montre le Tableau 2, le groupe de discussion ayant trait au film génère le plus grand volume de communications. Pour un peu plus de 50 % de participants en plus, il produit plus du double de messages et de fils de discussion. Il donne aussi souvent lieu à des messages plus longs que le groupe concerné par les téléphones portables. Cependant, les groupes sont comparables en matière de proportion de messages laissés sans réponses, démontrant que dans les deux cas, les participants construisent des échanges durables (*cf.* notamment Fisher, Bristor et Gainer, 1996 ; Galegher, Sproull et Kiesler, 1998), ce qui va à l'encontre des présupposés concernant les interactions en ligne perçues initialement comme de simples échanges anonymes au sein de groupes totalement fluides.

Choix des mesures

Traditionnellement, les mesures d'interaction sociale utilisées pour décrire la circulation de l'information comprennent les relations, leur direction, la force des liens et la centralité (Valente, 1995 ; Rogers, 2003), mesurées typiquement par des réponses auto-rapportées par le biais de questionnaires ou d'entretiens. La possibilité de mesurer la contribution et la place de chaque individu dans le groupe est l'un des bénéfices les plus utiles de l'observation directe des groupes de discussion en ligne. Selon Granitz et Ward (1996), les individus peuvent être classés en fonction de leur centralité et de leur appartenance à différents sous-groupes, les experts étant plus actifs, connectés et jouissant de ce fait de plus de capital social.

Nous avons deux propositions théoriques à tester. L'une concerne la distribution des fils par auteurs et l'autre la distribution des auteurs par fil. Pour tester la première, nous avons mesuré empiriquement le nombre de fils auxquels chaque individu a contribué. Nous avons calculé alors la distribution inverse cumulée et testé son ajustement avec une loi de puissance sur un graphique doublement logarithmique (*cf.* Figures 1 et 2). Pour la seconde, c'est le nombre d'auteurs participant aux différents fils constituant un forum de discussion qui nous intéresse, puisque chaque fil représente une conversation spécifique sur un thème particulier ayant trait au produit concerné. Nous avons mesuré pour chaque fil son nombre de participants, n, et testé l'ajustement par une régression du type $Ln[p(n)]+Ln[Ln(n)]=-aLn[n]+b$. La seconde

Tableau 2. – Profil général et indicateurs pour les deux groupes de discussion en ligne

Groupe	**Téléphones portables**	**Films**
Nombre de jours d'observation	244	245
Nombre de contributeurs de messages	1 257	1 952
Nombre de messages	8 036	19 964
Nombre de fils initiés	1 597	3 025
Nombre de réponses	6 439	16 939
Ratio réponses/messages	.801	.848
Ratio messages sans réponses/messages	.046	.045
Nombre de lignes	183 213	666 388

proposition est donc validée si a n'est pas significativement différent de 2,25. Les petits fils de 1 ou 2 auteurs sont exclus de cette analyse (*cf.* Figures 3 et 4), par souci de parcimonie, ce qui ne modifie pas les résultats obtenus.

RÉSULTATS EMPIRIQUES ET CONSÉQUENCES SUR LA DIFFUSION DE L'INFORMATION

Les Figures 1 et 2 décrivent les observations empiriques du nombre de fils fréquentés par auteur de messages, et ce pour les deux catégories de produits examinées dans cette étude. On constate, en accord parfait avec notre première proposition, que la distribution de cette mesure suit une loi de puissance, et ceci dans les deux groupes de discussion. Nos résultats confirment donc bien que dans le cas des deux catégories de produits examinées dans cette étude, le nombre de fils auxquels les consommateurs internautes participent suit une loi de puissance, d'exposant $\beta = 2{,}23$ pour le groupe téléphone et $\beta = 2{,}15$ pour le groupe cinéma. Ceci signifie que la probabilité que n consommateurs participent à un fil de discussion donné est proportionnelle à $n^{-2{,}23}$ pour le groupe téléphone et $n^{-2{,}15}$ pour le groupe cinéma. L'ajustement est excellent avec des coefficients de détermination très proches de 1. En accord avec la théorie des réseaux sans échelles (*scale free*) de Barabasi, Albert et Jeong (2000), l'exposant est bien compris entre 2 et 3. La première proposition, P1, est donc vérifiée empiriquement.

Les Figures 3 et 4 présentent quant à elles les observations empiriques de la taille des fils (mesurée en nombre d'auteurs y contribuant), dans les deux catégories de produits considérées dans notre étude. Ces figurent portent, en ordonnées, le logarithme de la fréquence des fils en fonction du logarithme de leur nombre de contributeurs en abscisses. Ce type de graphique a été choisi car il permet de mettre en évidence des distributions de probabilité en loi de puissance ainsi la qualité de l'ajustement dans les deux cas (coefficients de détermination autour de 0,9), la représentation graphique obtenue étant une simple droite. Les exposants obtenus sont de 2,5 pour le groupe téléphone et de 2,0 pour le groupe cinéma. Ces valeurs ne sont pas significativement différentes de la valeur théorique prévue par notre modèle (2,25), comme le confirment les tests de Student ($t = 1{,}4$ pour le téléphone et 1,2 pour le cinéma). Notre seconde proposition, P2, est donc validée, le nombre d'auteurs par fil suivant bien la loi de distribution attendue (voir équation 5 en Annexe).

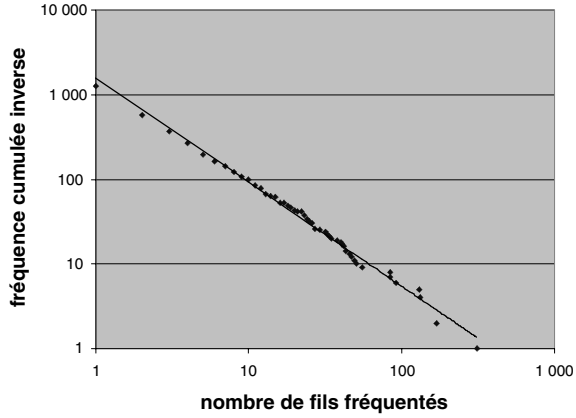

Figure 1. – Fils fréquentés par auteur (groupe portable)

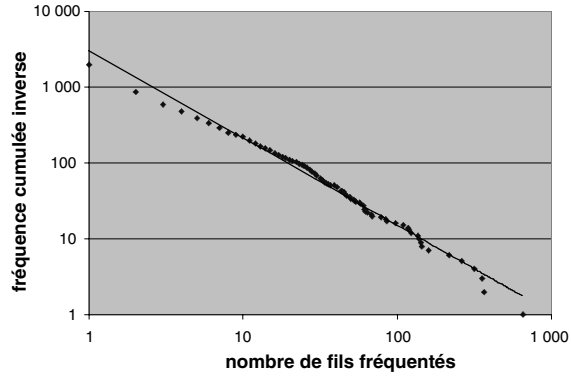

Figure 2. – Fils fréquentés par auteur (groupe cinéma)

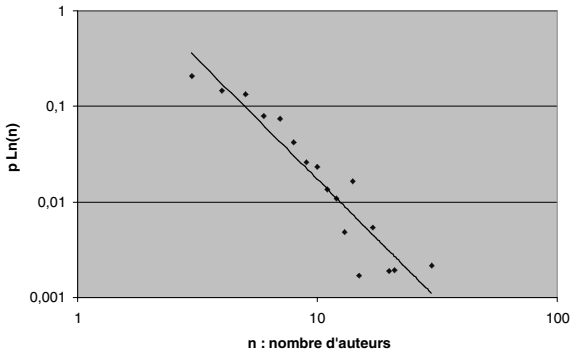

Figure 3. – Distribution des auteurs par fil (cas du portable)

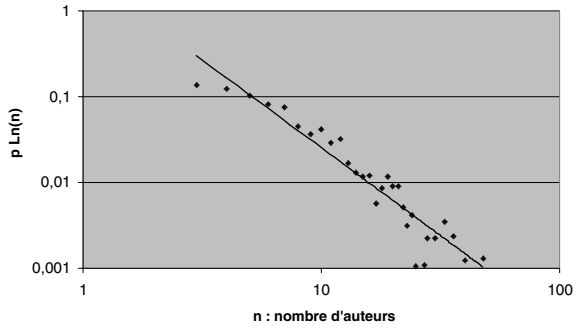

Figure 4. – Distribution des auteurs par fil (cas du cinéma)

Les conséquences sur la diffusion de l'information et son contrôle par les managers sont alors importantes. Pour que celles-ci deviennent apparentes, il faut au préalable expliciter les conséquences mathématiques des résultats empiriques obtenus. En effet, lorsque nous voulons calculer la variance de ces distributions, nous obtenons des intégrales divergentes, c'est-à-dire des intégrales de loi de puissance dont les primitives ne tendent pas vers zéro en l'infini (*cf.* Annexe A2). Qu'il s'agisse du nombre d'auteurs par fil, ou de fils par auteur, les variances théoriques sont donc infinies. Or, les managers sont habitués et ont été formés à un raisonnement économétrique dominant, gaussien, selon lequel les lois de distribution ont des variances finies et où, par application du théorème central limite, elles convergent vers la loi normale. Une manifestation pratique de ce raisonnement est l'ubiquité des intervalles de confiance comme support d'aide à la décision. Face au phénomène de communication par Internet et de marketing viral, ces raisonnements ne semblent plus pertinents. En particulier, les managers se doivent d'éviter de sur-interpréter des écarts, même extrêmes, de leurs observations en ligne. En effet, le phénomène de bruit et l'apparition de variations de grande amplitude pourraient en toute vraisemblance n'être le résultat que de processus aléatoires plutôt que de phénomènes sous-jacents. Agir précipitamment sur des variations fallacieusement interprétées pourrait alors induire des résultats contre-productifs ou totalement inappropriés.

En effet, puisque les variances sont infinies, le théorème central limite ne peut plus s'appliquer. Les groupes de discussion se structurent donc suivant des processus tellement aléatoires et variant tellement qu'un intervalle de confiance ne peut plus les tenir captifs. Le manager doit donc s'attendre à tout moment à des surprises, dans un sens ou dans l'autre, diffusion très rapide de l'information ou arrêt brutal. Contrairement aux médias classiques, le pilotage de la diffusion de l'information est donc très risqué. Si l'information est positive, nos résultats doivent encourager les managers à utiliser le marketing viral de façon multiple afin d'arriver à générer de plus grandes « avalanches », c'est-à-dire de multiplier des foyers potentiels de contamination par l'information. Si au contraire, l'information est négative, il deviendra quasiment impossible de prédire et donc de maîtriser la diffusion de l'information.

Autre conséquence managériale majeure : l'absence de masse critique qui caractérise les réseaux sans échelle du type de ceux constitués par les deux groupes de discussion Internet étudiés dans notre recherche, remet en cause la stratégie centrale utilisée par les managers (selon laquelle la diffusion d'un produit innovant sera facilitée par l'identification, puis le ciblage, de leaders d'opinion). Le rôle clé de ces individus, et leur contribution nécessaire au processus de diffusion, sont ici remis en question, en tout cas pour ce qui concerne la diffusion de l'information, qui pourtant semblerait essentielle à l'éventuelle adoption de nouveaux produits. Connaissant

les difficultés rencontrées par les managers qui cherchent à cibler les leaders d'opinion, cette constatation peut justifier un changement de stratégie dans l'élaboration des campagnes de diffusion et éviter ainsi des démarches aussi coûteuses que peu fructueuses.

Dans le cas particulier de phénomènes de rumeurs ou de crises auxquelles le manager peut se voir confronté, l'absence de masse critique demande une stratégie de réponse très différente de celle habituellement préconisée dans le cas du bouche-à-oreille négatif. En effet, plutôt que de se contenter de surveiller l'évolution de la rumeur, sujette à une extinction spontanée pour peu qu'elle n'atteigne pas le seuil critique dans le cas du bouche-à-oreille, et que l'on risquerait alors plutôt de diffuser à tort par une réaction prématurée, le *buzz* négatif demande lui une réaction immédiate de la part de la compagnie ciblée. Faute d'y mettre fin de façon rapide, ce *buzz* négatif est en effet susceptible de prendre rapidement des proportions telles que la réputation de la marque en serait sensiblement érodée, et ceci sans qu'un seuil critique n'en prédise l'occurrence. Pour prévenir de telles éventualités, il est souhaitable et sans doute nécessaire pour les managers de superviser de façon continuelle le contenu des discussions en ligne, afin de pouvoir réagir dès ses premières ébauches, à toute diffusion de cette information en ligne.

CONCLUSION : APPORTS, LIMITES ET VOIES DE RECHERCHE

Cette recherche a validé empiriquement, et pour la première fois, deux prévisions originales concernant spécifiquement les groupes de discussion. Premièrement, comme dans beaucoup de réseaux sociaux, les interactions sont dominées par des lois de puissance strictes, dans le cas du nombre de fils par auteur, ou des lois corrigées par des logarithmes, dans le cas du nombre d'auteurs par fil. Ces deux lois, postulées et validées dans cet article, ne présentent pas seulement un intérêt mathématique ou descriptif. D'un point de vue théorique, ces résultats apportent une meilleure connaissance de la forme agrégée prise par les interactions sociales dans les groupes de discussion. Or, bien que celles-ci jouent un rôle non négligeable dans les processus de diffusion, leur topologie est trop souvent considérée soit comme purement aléatoire (notamment dans les modèles de prévision issus des travaux de Bass), soit comme normalisée (Frenzen et Nakamoto, 1993). Les distributions validées possèdent de surcroît des propriétés dont les conséquences sont importantes au regard des propositions du marketing viral.

Ainsi, l'idée selon laquelle cet ensemble de techniques permet une augmentation « exponentielle » de la visibilité de produits ou de services « viraux » (Godin, 2001) est validée tant par la variance infinie de la contribution en messages que par l'absence de masse critique induite (Steyer, Garcia-Bardidia et Quester, 2006). D'un point de vue managérial, il serait donc possible de saturer le marché potentiel, ou la cible du marketing viral, d'une exposition à un sujet donné via les réseaux sociaux portés par ce média, ici matérialisés par des groupes de discussion. En revanche, le marketing viral se différencie des politiques de communication fondées sur le bouche-à-oreille en ce qu'il permettrait un contrôle accru tant des interactions entre consommateurs que de leur issue en termes de diffusion. Ce n'est au final qu'à cette condition qu'il en constitue une approche spécifiquement plus efficace. Le premier point n'est pas sans poser des questions éthiques importantes, en particulier lorsqu'on envisage de rémunérer, même symboliquement, des « leaders d'opinion » (Godin, 2001), ou des actions clandestines dans les groupes, forums ou blogs (Dellarocas, 2006). L'absence de masse critique induite par les résultats obtenus indique cependant une imprévisibilité totale des processus de diffusion au moyen des modèles classiques, ainsi qu'un rôle ambigu des leaders d'opinion, cibles privilégiées du marketing viral (Vernette et Florès, 2004), la diffusion pouvant atteindre son maximum quels que soient les individus à leur origine (Steyer, Garcia-Bardidia et Quester, 2006).

Au final, les résultats obtenus et leurs conséquences managériales en termes de diffusion de l'information ne confirment qu'en partie les propositions du marketing viral. Dès lors, la gratuité de ce type de techniques, quoique remise en cause par le recours croissant aux agences spécialisées en marketing viral, prend tout son sens et confirme l'intérêt et l'urgence de la formation d'un corpus de recherche spéci-

fique. Cependant d'un point de vue méthodologique, notre étude démontre comment le manager averti peut développer et collecter des mesures quantitatives qui, si elles ne permettent pas de prédictions réelles, pourraient toutefois faciliter la comparaison entre groupes de discussion et identifier les opportunités d'intervention dans les fils. En particulier, des informations positives ont sur Internet le potentiel de se propager à l'infini, pour peu que suffisamment d'avalanches puissent être déclenchées. Bien entendu, nos résultats n'ayant été validés que dans le contexte de groupes de discussion ayant pour sujet deux catégories de produit spécifiques, il sera nécessaire de faire précéder toute généralisation de nos conclusions d'études similaires répétées dans d'autres contextes.

Dans ce cadre, le passage d'un modèle de diffusion de l'information à l'étude de la diffusion des préférences constituerait une étape ultérieure nécessaire (Godes et Mayzlin, 2004). Pour ce faire, il conviendrait d'intégrer une mesure reflétant la valence de l'information partagée par les participants dans les deux groupes de discussion choisis. Ont ainsi été négligées l'orientation positive ou négative des messages, dont on sait qu'elles influencent la diffusion du bouche-à-oreille et son efficacité (Richins, 1983), et par extension, la possibilité de déformation du contenu des messages au fur et à mesure de la construction du fil.

Enfin, nos résultats soulignent l'importance de l'attrait du sujet dans la croissance des fils et indiquent qu'une technique donnée de marketing viral ne peut *a priori* se substituer aux capacités intrinsèques du produit à devenir un sujet attrayant ou polémique, point notamment rappelé par Godin (2001). Le rôle joué par l'implication, qui conditionne la recherche d'information menée par les consommateurs et donc le processus d'achat, et le rôle social occupé par les différents contributeurs, par exemple sous l'angle du leadership d'opinion ou d'une nethnographie (Kozinets, 2002), méritent un approfondissement puisque ce travail sur le message ou sur la cible constitue deux stratégies alternatives de marketing viral.

RÉFÉRENCES BIBLIOGRAPHIQUES

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

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

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

Barabasi A.L., Albert R. et 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. et 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. et 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. et Saberi A. (2004), Exploring the community structure of newsgroups, papier présenté à la 10e Conférence internationale "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. et 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. et 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, papier de recherche, MIT Sloan School of Management, Cambridge, MA.

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

Ebel H., Mielsch L.-I. et 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. et Gainer B. (1996), Creating or escaping community? An exploratory study of internet consumers' behaviour, in K.P. Corfman et J. Lynch (coord.), *Advances in Consumer Research*, 23, Provo, UT, Association for Consumer Research, 178-182.

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

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

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

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

Gensollen M. (coord.) (2004), Biens informationnels et communautés médiatées, *Revue d'Économie Politique*, numéro hors-série « Marchés en ligne et communautés d'agents », mars, Paris, Dalloz, 9-40.

Godes D. et 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. et 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. et Ward J.C. (1996), Virtual community: a socia-cognitive approach, in K.P. Corfman et J. Lynch (coord.), *Advances in Consumer Research*, 23, Provo, Utah, Association for Consumer Research, 161-166.

Granovetter M. (1973), The strengh 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. et 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, Thèse de doctorat en sciences de gestion, Université de Bourgogne, Dijon.

Lazarsfeld P.F., Berelson B.R. et Gaudet H. (1948), *The people choice*, 2e édition, DC, Columbia University Press.

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

Pastor-Satorras R. et 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. et 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*, 5e édition, 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. et 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. et Lilien G.L. (2001), Medical innovation revisited: social contagion versus marketing effort, *American Journal of Sociology*, 106, 5, 1409-1435.

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

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

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

En tenant compte de la nécessaire normalisation des probabilités, nous obtenons le modèle

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

Bien que le phénomène étudié soit de nature discontinue, nous utilisons une modélisation continue pour simplifier les calculs. Pour calculer $\sum_k n_k\theta_k$ nous remplaçons donc la somme par une intégrale, soit

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

Qui donne, après calcul des primitives :

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

avec

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

En utilisant l'équation 3 dans l'équation 2, nous obtenons l'équation différentielle suivante :

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

Cette équation différentielle permet une cohérence avec l'équation (1) à condition de prendre :

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

Cette dernière relation peut être incorporée dans l'équation (4) afin d'obtenir l'équation implicite suivante :

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

Dont la solution est $\omega = 1{,}25$.

Munis de cette valeur de ω, nous sommes en mesure de calculer la distribution de la taille des fils en auteurs. Pour un fil donné i, son temps initial t étant distribué uniformément, nous déduisons de l'équation 1 que n_i est distribué inversement proportionnellement à sa puissance $1 + \frac{\omega}{\theta_i}$. En moyennant suivant les valeurs de θ, on obtient la distribution de la taille des fils, n :

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

Quand n est assez grand, cette intégrale se comporte comme :

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

1) $\langle k^2 \rangle = \int_1^\infty k^2 \frac{dk}{k^\beta}$ est divergente car $2 < \beta < 3$ et donc $V(k) = \infty$.

2) De même $\langle n^2 \rangle = \int_1^\infty n^2 \frac{dn}{Ln(n)\, n^{2,25}} = \int_1^\infty \frac{dn}{Ln(n)\, n^{0,25}}$ est une intégrale de Bertrand, elle aussi divergente. On a donc aussi $V(n) = \infty$.