Physica A 389 (2010) 859–870

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A study of the spreading scheme for viral marketing based on a complex network model

Jianmei Yang [a], Canzhong Yao [a], Weicheng Ma [a], Guanrong Chen [b,*]

[a] School of Business Administration, South China University of Technology, Guangzhou 510640, China

[b] Department of Electronic Engineering, City University of Hong Kong, Hong Kong SAR, China

ARTICLE INFO

Article history:
Received 17 August 2009
Received in revised form 21 October 2009
Available online 11 November 2009

Keywords:
Viral marketing
Spreading scheme
Complex network
Social network
Instant messaging system

ABSTRACT

Buzzword-based viral marketing, known also as digital word-of-mouth marketing, is a marketing mode attached to some carriers on the Internet, which can rapidly copy marketing information at a low cost. Viral marketing actually uses a pre-existing social network where, however, the scale of the pre-existing network is believed to be so large and so random, so that its theoretical analysis is intractable and unmanageable. There are very few reports in the literature on how to design a spreading scheme for viral marketing on real social networks according to the traditional marketing theory or the relatively new network marketing theory. Complex network theory provides a new model for the study of large-scale complex systems, using the latest developments of graph theory and computing techniques. From this perspective, the present paper extends the complex network theory and modeling into the research of general viral marketing and develops a specific spreading scheme for viral marking and an approach to design the scheme based on a real complex network on the QQ instant messaging system. This approach is shown to be rather universal and can be further extended to the design of various spreading schemes for viral marketing based on different instant messaging systems.

1. Introduction

Viral marketing, known also as word-of-mouth marketing, refers to a marketing mode attached to some carriers on the Internet, which can rapidly copy marketing information at a low cost [1–4]. Different from the traditional marketing methods, for example by advertisement, viral marketing spreads the marketing information through word-of-mouth among the users, which turns out to be more trustful for many consumers. Noticeably, viral marking is different from the traditional word-of-mouth marketing, and it carries some advantages over the traditional method for it is launched via web-based communication platforms such as BBS, Web Blogs, instant messaging systems, and so on [5,6]. The main reasons are multi-fold. First, the information can be easily stored in computers and the web; therefore digital word-of-mouth marketing information exists everywhere on the Internet and for a long time. Second, most people can take part in the word-of-mouth marketing process in their leisure time, so by nature it is a non-synchronous spreading mode of advertisement. Third, the anonymity feature enormously reduces the limitation induced by personal identities thereby making mutual communications on the web much easier than face-to-face in real life. Last but not least, digital word-of-mouth marketing through the Internet can spread more widely and much faster at a significantly lower cost as compared to most if not all traditional methods [7].

* Corresponding author.
E-mail address: eegchen@cityu.edu.hk (G. Chen).

0378-4371/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2009.10.034

Instant messaging system provides a tool for the involved users to identify on-line consumers and to communicate with others synchronously. Such communication based on the system is actually even more convenient than BBS and Blogs. In fact, instant messaging systems have become one of the most favorable means for word-of-mouth marketing, which has boosted a rapid development of viral marketing today.

The earliest instant messaging system, ICQ, was originated from Israel in 1996. Afterwards, it became more and more popular worldwide. The number of users of various instant messaging systems in China had reached 154 million in 2007, with QQ being the most popular platform [8].

In retrospect, the first one who wrote about viral marketing on the Internet was media critic Douglas Rushkoff, in his 1994 book Media Virus [9]. The term *viral marketing* was suggested by Jeffrey Rayport, a faculty member at Harvard Business School, in his 1996 article "The Virus of Marketing" [10]. The term afterwards was popularized by Tim Draper and Steve Jurvetson of the venture capital firm Draper Fisher Jurvetson, in 1997 [11]; they described Hotmail's email practice of appending advertisements in outgoing mails from the web users.

In real practice, the precursor of viral marketing company named Windows Live Hotmail, known as MSN Hotmail or simply Hotmail, is a free web-based email service operated by Microsoft. It is a marketing mode that every message contains a Hotmail advertisement. It grew to having 12 million accounts in its first year 1996 alone, and had more than 270 million users worldwide as of 2008 [12]. At present, PepsiCo, Tupperware Corporation and Microsoft etc. have chosen viral marketing as their main marketing mode. One may expect that more and more companies will focus their attention on this effective marketing mode, therefore it is important and even necessary to further understand and analyze viral marketing and its mechanism in a general setting, which motivates the research presented in this paper.

In the literature, there are some studies of viral marketing reported, but they are mainly referred to some concepts related to the viral marketing mode, the extent to which these might influence the success of viral marketing, some basic steps to implement the viral marketing, and so on. On the other hand, most of the existing research works were devoted to simple qualitative analysis. And there are very few references that focus on the design of spreading schemes for viral marketing, although the importance of the design steps has already been noticed recently.

Because the viral marketing campaigns were actually performed over some pre-existing and large-scale social networks, and the existing analysis methods are not powerful enough for carrying out detailed research over large-scale networks, we resort to the new complex network theory, particularly its data-based modeling and simulation methods. As it turned out, the new approach is quite effective for analyzing the difficult viral marketing scheme designs as compared to the conventional methods.

Motivated by the above observations, in this paper we bring complex network theory and modeling into the viral marketing research and propose a specific spreading scheme for viral marketing. This scheme, taking into account the numbers and locations of the initiators under different average rates of activated users, uses the instant messaging system QQ as a platform to carry out a detailed study of viral marketing over real user-group social networks. It is found that the proposed QQ-based spreading scheme for viral marketing is very effective over social networks, and that the method is rather universal which can be further extended to the viral marketing using other instant messaging systems as well.

2. Literature review

2.1. Complex networks

The seminal research works of Watts and Strogatz in 1998 [13] and of Barabási and Albert in 1999 [14] have spurred a great deal of interest in studying real-world problems with complex network models. Scientists have known that networks in the real world, such as various social networks, WWW and the Internet, can be modeled by neither completely random nor completely regular network models, and that they should be and can be modeled by more general complex network models which may have some very different statistical features [13–16]. Although the concept of complex network emerged a few years later than the model of viral marketing, basic theories and many real-world applications of the complex network theory have seen very rapid developments in recent years. As a few relevant examples, the complex network theory and modeling have already been applied to analyzing the industrial competitive relationship, building a two-level complex network model of the inter-firm competitive relationship in industry and the inter-firm rivalry that this relationship determines, and modeling bank services channels and telecommunication VIP service schemes [17,18].

A complex network is composed of a set of nodes joined together with edges representing the relationship between the nodes. Complex network models have some unusual features such as large-scale characteristics, interactional relationships, and complexity and dependence on high-computing capabilities. When studying a complex network model, one not only analyzes its node-degree distribution, weighted-edge distribution, average shortest path length, and average clustering coefficient (in order to identify its graphic features), but also studies its topology index (in order to indicate every node's significance in the network), network dynamic properties, and spreading traffic over the network, etc.

From the perspective of statistics, the classification of complex networks contains two important types of models: small-world networks and the scale-free networks [13,14]. The node-degrees of a small-world network follow an exponential distribution, meanwhile it has a short average path length with a high average clustering coefficient. The node-degrees of a scale-free network follow a power-law distribution with a prominent "robust yet fragile" property known as the "Achilles' heel" phenomenon [15], which means that such a network is robust against random attacks but at the same time it is fragile to intentional attacks on its highly-connected nodes.

From the perspective of node functions and edge relationships, one may classify various complex networks into social networks, economic networks, technical networks, biological networks, and so on. A common feature of all these networks is that the scales of these networks are all very large and the relationships among their nodes are extremely complicated, verifying their generic name as "complex networks".

2.2. Viral marketing

Based on the discovery of Rushkoff and Rayport in 1997, Wilson in 2000 [19] introduced six factors that influence viral marketing, as follows: offering valuable products and services, providing efficient ways for communications, utilizing larger-scale spreading of information, making efficient use of public positivisms and behaviors, establishing communication networks, and sharing resources of the others. After many case studies, Podoshen discovered in 2006 [20] that although various company's viral marketing modes are different and heterogeneous, the actual steps of viral marketing performed by a company seems to be an integration of the design of information resources, information spreading, information release, and the consequent management. On the other hand, it is usually believed that although viral marketing needs fewer directed expenses on information transferring, the marketing mode would have some cost due to serious management of the viral marketing processes.

2.3. Viral marketing and social networks

Both the traditional word-of-mouth marketing and the viral marketing campaigns are carried out over social networks, and the existing literature of viral marketing over social networks has significantly grown as research is being further developed recently. Social network potential (SNP) is a numeric coefficient, derived through computational algorithms to represent both the size of an individual's social network and the ability of individuals to influence the whole network [21]. Alpha users were first briefly discussed in the book *3G Marketing* by Ahonen et al. in 2004 [22,23], thereafter SNP was used for the purpose of identifying alpha users. Message customization, social network structure and customers' motive are salient antecedents that determine the spreading performance of the marketing information and communications [3].

Some research on viral marketing schemes was carried out by the present authors based on QQ user-groups between 2007 and 2008 [8]. Moreover, the simulation results reported in Ref. [24] reveal that the email network of the users performing actual viral marketing activities is similar to a small-world network, regarding the information spreading behaviors.

2.4. Propagation of email viruses

A model of email virus propagation was proposed in Ref. [25]: an email user would be infected if this user opens a virus-contaminated email attachment; thus, the virus program will infect the user's computer and send itself as a new attachment to all email addresses existing in the user's computer address book. Using that model, simulations on the spread impact of the network's magnitude and the initiator's degrees of the network were analyzed [25]. The present work, to be described below, has some similarity with this email virus propagation mechanism but only in the initiation and propagation aspects; they address different networks with different mechanisms.

2.5. The present work

Based on a complex network model, this paper performs simulation studies on a spreading scheme for viral marketing over the real instant messaging system QQ in China.

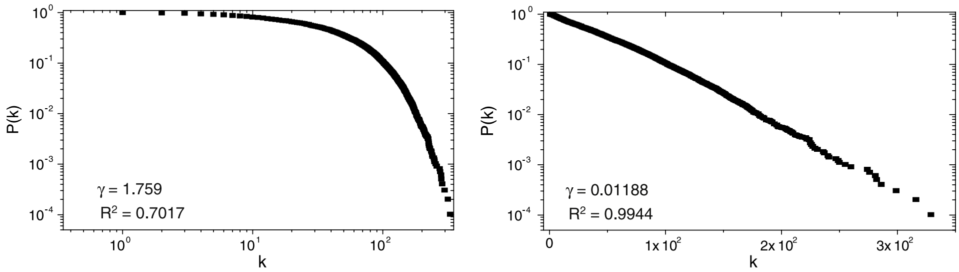
As mentioned above, Ahonen et al. [22,23] contributed quite a lot to viral marketing research in combination with the network theory and modeling. However, they only focused on the selection of the initiator in viral marketing over an ordinary social network but not on a more realistic complex network. Bampo et al. [24] used a network generated by a computer to predict the small-world and scale-free properties of some real information spreading networks. Zou [25] studied the impact of the social network's magnitude and the initiator's degree on email virus spreading. Clearly, what has been studied is not the problem about the design of spreading schemes for viral marketing. These approaches on simple networks or on computer-generated artificial networks generally cannot be applied to designing effective spreading schemes for the large-scale and complex task of viral marketing, therefore it is very important to perform real data-based studies on the complex social networks for viral marketing.

It has been commonly experienced that although the macroscope properties of two network models can be similar, for example possessing the same small-world features and even identical power-law degree distributions, their microscope structures can be very different. Therefore, instead of discussing mathematical graph models generated by computer, it is more practical to consider real-world networks with information spreading dynamics using real data sets. The present paper takes on this latter approach to study the design of a spreading scheme for viral marketing, as detailed below.

3. Modeling and analysis of the QQ user-group network

3.1. Network model

We randomly drew ten-thousand (9867 to be exact) individuals from the list of all QQ users as a sample for study. Our research objective is to explore marketing related to general merchandise but not any specific consumer groups, for



(a) Log–log plot. (b) Mono-log plot.

Fig. 1. Cumulative degree distribution.

which random sampling of consumers is reasonable. On the other hand, because marketing communication on QQ is mainly through QQ groups, the group numbers are indexed in the data set. In constructing a network model for the QQ user-groups, a node represents a QQ user labeled by its index integer (namely, 1, 2, 3, ...), and an edge links any two QQ users who have a common group index number. Since some individuals join several groups, different groups may be connected through them. Of course there exist some isolated groups if they do not have common members. The constructed group network of QQ users, named the *QQ network* hereafter, has totally 9867 nodes and 223,375 edges.

3.2. Analysis

Intensive numerical analysis has been performed based on real data for the QQ network described above, using some common measures for complex networks [15].

Fig. 1 shows the cumulative degree distribution of the network. It reveals that the R^2 of the exponential distribution is larger than that of the power-law distribution, so it is concluded that the degree distribution in this network is exponential.

Computational results of other essential indices of the network are as follows.

The average density of the network is 0.046, so it is a sparse network.

The total of 9867 nodes forms 210 connected subgraphs. The largest connected subgraph consists of 9081 nodes and has an average shortest path length 3.1405. These topological indices show an interesting phenomenon that in the QQ network there exist a large number of isolated small subgraphs and a few large-scale subgraphs. In terms of connected numbers of users, the QQ network has a better connectivity as compared to the numbers of the subgraphs.

The average clustering coefficient of the QQ network is found to be 0.719 and the average clustering coefficient of the maximum connected subgraphs is 0.697. Therefore, the QQ network has relatively short average distances and a relatively high clustering coefficient, indicating that it is a small-world network.

On the other hand, the average degree in the QQ network is 45 with the standard deviation being 41 (having minimum 1 and maximum 329, respectively). The degrees of user #3324 and user #1306 are larger than 300. The degree centrality of the network is 2.88%, with minimum and maximum of the farness being 7,776,761 and 97,333,864, respectively.

The farness of user #3324 and user #1306 is minimum, implying that their closeness centrality is maximum. The minimum of node betweenness is 0 while the maximum is 241,224, whereas the betweenness of user #3324 and user #6070 is maximum.

4. Simulations on the QQ-based spreading scheme for viral marketing

In this section, a QQ-based spreading scheme for viral marketing is proposed based on extensive simulations, which show that it is very effective over social networks and is rather universal in the sense that it can be extended to many similar network topologies.

The objective is to design a QQ-based spreading scheme for viral marketing, when a desirable average activation probability of users is given and fixed.

The criterion used for judging the goodness of the scheme is the total number of activated (infected) users when the spreading process reaches the stable state: the higher this number, the better the scheme.

4.1. Assumptions

The simulation variables include the total number L_0 of initiators (initially infected nodes) and their degrees, betweenness and average activation probability.

At the beginning, the initiators transmit some kind of virus (marketing information) to all their friends (users) in each of such QQ groups. These group-friends are all contaminated (covered) by the virus received, but only when they have been activated (namely, have been infected and then started to spread the virus) will they spread the virus to their individual friends. The possibility of such friends being repeatedly contaminated and activated by the same virus is excluded from the present study.

The probability of user i being activated is P_i and assume that this probability follows a normal distribution with the expectation $E(P_i)$ being the average activation probability.

Let n and m be the numbers of users who are activated or contaminated in the stable state, respectively, excluding the initiator itself, due to one single virus source (i.e., one single node with degree $D = i$). Then, since there might be many users with degree $D = i$, let n and m be the average values of n and m, respectively, for all the initiators whose degrees $D = i$.

The final results (outputs of the scheme) are N (the average number of activated uses) and M (the average number of contaminated users), obtained by averaging many simulation runs that have the same inputs created by using different random number generator seeds.

4.2. Simulation models

Two simulation models are considered: (I) it is about the position of a single initiator ($L_0 = 1$) in the QQ network; (II) it is about the total number of initiators (their positions are fixed).

4.2.1. Simulation model I

The degree D and the betweenness B of the single initiator show the role of the initiator in the QQ network. Assuming that $L_0 = 1$, it is to find the relationships between N or M with D or $E(P_i)$. There might be a number of users that have the same degree D, so simulation is performed for every initiator with degree $D = i$, for $i = 1, 2, \ldots,$ and then the average value of these results are taken as the simulation output of one run. A total of 50 runs were simulated for each single user of degree $D = i$, for $i = 1, 2, \ldots,$ until the process is ended.

The procedure is proposed as follows, which provides a step-by-step procedure for numerical simulations, where $E(P_i)$ is assumed to be given and fixed.

The procedure is as follows:

For each i, which corresponds to the degree of certain node in the network ($i = 1, 2, \ldots, 329$), do the following:

(1) Randomly choose one node of degree $D = i$ as the initiator; if there is no node with degree i, go to step 8.
(2) Let all the initiator's group-friends be contaminated by the virus (information) from the initiator.
(3) Generate a normally-distributed random number as the probability P_i for friend i to be activated, where the expectation of the normal distribution is $E(P_i)$ in the experiment, and the variance is set as 0.0009 if $E(P_i)$ is less than 0.005, and is set as 0.01 otherwise. Then, for each of its group-friends, generate a random number with a uniform distribution in [0,1]: if P_i is larger than the random number, the corresponding friend is activated. Repeat the same process for all the initiator's group-friends.
(4) Let every activated friend be a new initiator, and then return to step 2.
(5) The simulation process continues until the number of activated users reaches a stable state. Then, calculate n and m, respectively.
(6) Take another node with degree $D = i$ as the new initiator, and then return to step 2.
(7) After the above-described process has been completed for every user with degree $D = i$, calculate the average number $\underline{n}$ of n, and the average number $\underline{m}$ of m.
(8) Let $D = i + 1$ and, if D is not larger than the maximum node-degree in the network, return to step 1; otherwise, stop.

4.2.2. Simulation model II

Users in the QQ network are located at different positions. Assume that the position of every initiator is given and fixed. The degree value of an initiator roughly reflects the role of the initiator in the network. This degree value of each initiator is determined by referencing the simulation results of Model I above. The number of initiators will then be determined through simulations as further discussed below.

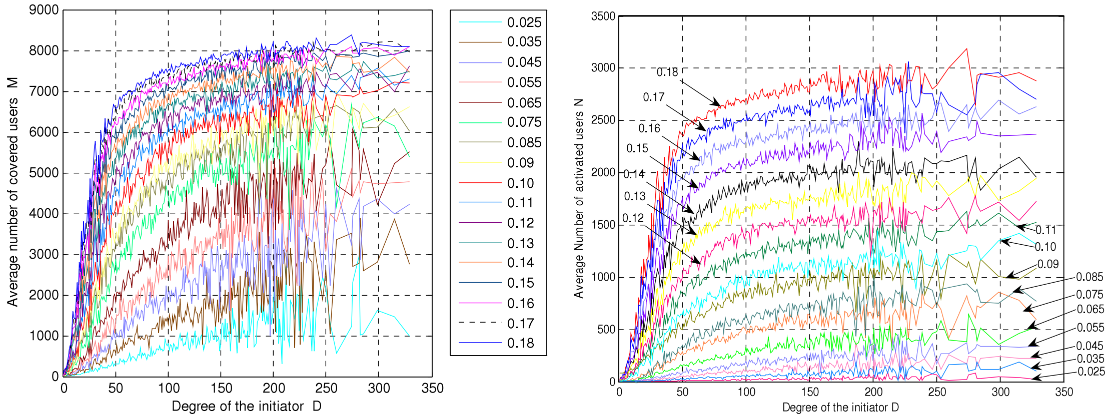
In every run of the simulation, randomly choose L_0 initiators whose degree values are equal to the assigned value. Perform simultaneous simulations for all initiators, based on the model described in Model I above. These initiators are chosen randomly, so they will not exhaust all possible cases (which are excessively too many), and consequently the number of runs of the simulation exceeds that in Model I above. Finally, calculate the average result of 500 runs of the simulations performed.

4.3. Simulation results and findings: Model I

Simulations for Model I all reached their stable states in no more than 5 time step.

Fig. 2 shows the results of the simulations, where (1) shows the relationships of the average number of contaminated users versus the degree of the initiator and the average activation probability; (2) shows the relationships of the average number of activated users versus the degree of the initiator and the average activation probability.

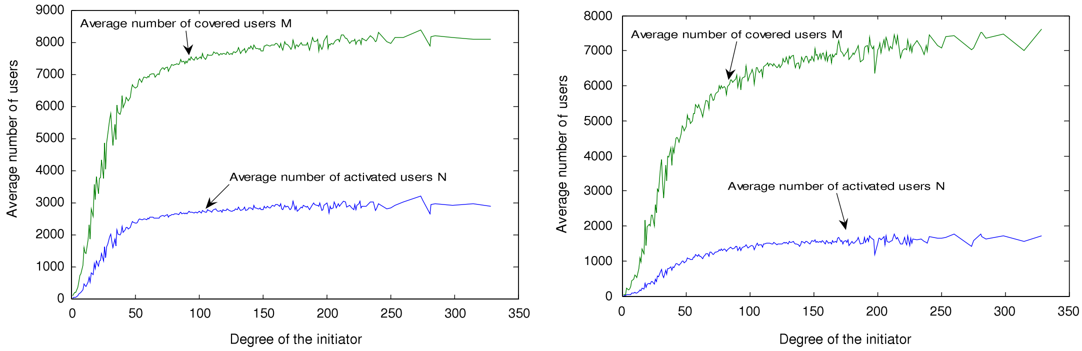
It can be seen from Fig. 2 that when there is only one initiator:



(a) Relationships of M versus D and $E(P_i)$.

(b) Relationships of N versus D and $E(P_i)$.

Fig. 2. Comparison of relationships of M and N versus D and $E(P_i)$ ($L_0 = 1$).



(a) $E(P_i) = 0.18$.

(b) $E(P_i) = 0.12$.

Fig. 3. Another comparison of relationships of M and N versus D and $E(P_i)$ ($L_0 = 1$).

(a) When the users' average activation probability is given and fixed, the higher the degree of the initiator, the higher the number of users eventually be contaminated and then be activated. Here, the numbers of average contaminated and activated users both have limiting or stationary values.

(b) When the degree of the initiator is given and fixed, the higher the users' average activation probability, the larger the number of users eventually be contaminated and then be activated. Likewise, the numbers of average contaminated and activated users both have limiting or stationary values. But, when the degree is lager than 250, the fluctuations of the numbers of average contaminated and activated users are larger.

(c) When the users' average activation probability is less than 0.09 and the initiator's degree is less than 100, the increase rate in the number of contaminated or activated users is very high, whereas when the initiator's degree is larger than 100, the increase rate is low. Therefore, 100 is a critical threshold and also a stability threshold for the initiator's degree used in the scheme, so one may choose a user with degree 100 to be the initiator in this QQ network. Correspondingly, when the users' average activation probability is 0.09 or higher, 50 is the stability threshold for the initiator's degree. This implies that when the users' average activation probability is increasing, the degree of the initiator can be decreased.

Fig. 3 demonstrates the relationships described above more clearly. The curves denote the relationships of the average number of contaminated users versus the degree of the initiator, and the relationships of the average number of activated users versus the degree of the initiator, respectively, under the condition that the average activation probability remains unchanged (0.18 or 0.12). From Fig. 3, one can see the relationships of the average number of contaminated users versus the degree of the initiator, and the average number of activated users versus the degree of the initiator, respectively, both following the same distribution. Therefore, one may focus on analyzing the changes in the average number of activated users.

In the following, we further explain the threshold value $D = 50$ when $E(P_i) \geq 0.09$ ($L_0 = 1$). Fig. 4 indicate the scopes of curve of N versus D when D is in the range of 1–50 and in the range of larger than 50, respectively ($E(P_i) = 0.18$ and $L_0 = 1$). Table 1 supplies the corresponding numerical data. From Fig. 4 and Table 1, one can see that when the degree is in the range of 1–50, the slope is 54.258 ($R^2 = 0.9619$), while when the degree is in the range of larger than 50, the slope

Table 1
Threshold value $D = 50$ ($E(P_i) = 0.18$, $L_0 = 1$).

Degree of initiator	1	10	20	30	40	50	60	70	80	90	100
Average number of activated users N	25.8411	462.1942	1171.899	1937.9494	2253.8416	2362.7581	2510.6833	2573.9423	2618.2653	2664.8372	2711.1538
Increase of N per 10 degree	–	436.3531	709.7048	766.0504	315.8922	108.9165	147.9252	63.259	44.3230	46.5719	46.3166

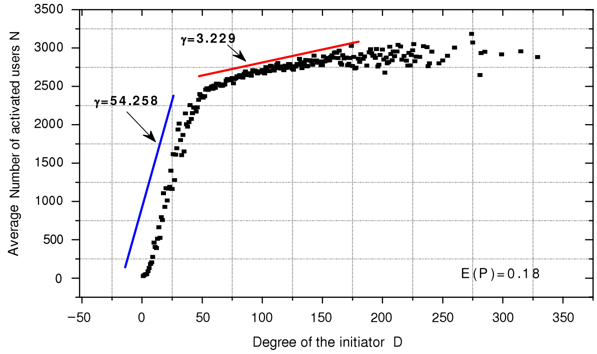


Fig. 4. Slopes indicate the threshold values: $D = 50(E(P_i) = 0.18, L_0 = 1)$.

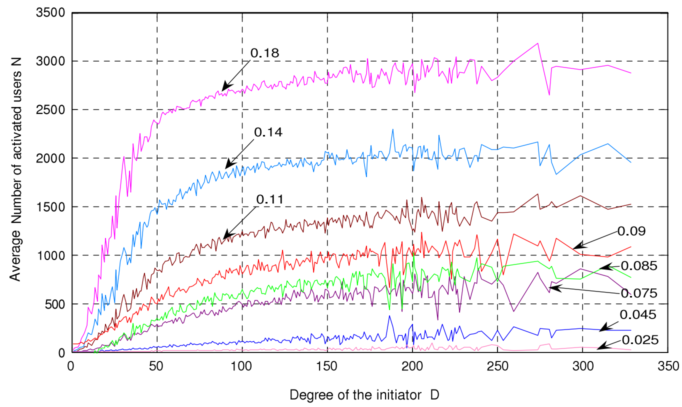


Fig. 5. Threshold $E(P_i) = 0.09$ ($L_0 = 1$).

is 3.229 ($R^2 = 0.8280$). These support the claim of threshold value $D = 50$ for the case of $E(P_i) = 0.18$, $L_0 = 1$. We have tested many different cases with $E(P_i) \geq 0.09$ and $L_0 = 1$, and found that the results are similar.

Next, we further explain the threshold value $E(P_i) = 0.09$ ($L_0 = 1$) by simulation figures and numeric. For clarity, selecting a few curves from Fig. 2(b) to draw Fig. 5. Table 2 supplies more numerical data on N versus $E(P_i)$ ($L_0 = 1$ and $D = 50$). From Fig. 5 and Table 2, one can see that under the condition of $D = 50$ ($L_0 = 1$) when the activated rate $E(P_i)$ is increasing from 0.085 to 0.09 (here, increment is only 0.005 rather than 0.01, unlike the increments in Table 2), the number of activated users N is increasing from 367 to be more than 500, the increment of N is larger than the others when $E(P_i)$ is between 0.025 and 0.085; therefore, we use 0.09 as the threshold for $E(P_i)$.

4.4. Simulation results and findings: Model II

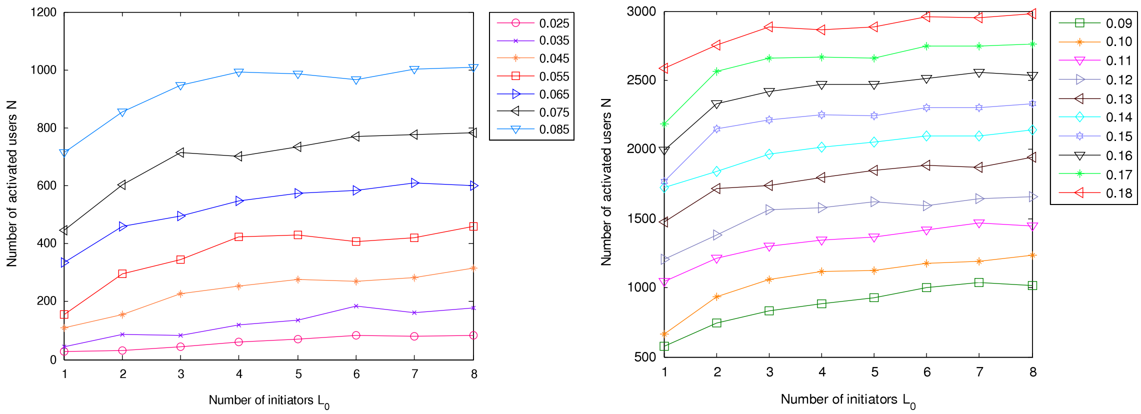
Simulations for Model II all reached their stable states in no more than 5 time steps.

Fig. 6 demonstrates the simulation results for the average numbers of activated users on the condition that the average activation probability is less than (larger than or equal to) 0.09 and the initiator's degree equals 100 (50), respectively. Under the condition that the degree of the initiator is fixed:

(a) When the average activation probability is unchanged, if the total number of initiators increases, then the average number of activated users also increases.
(b) When the average activation probability is less than or equal to 0.035, the increase rate in the number of activated users resulting from an increase in the total number of initiators is low, and there are no obvious thresholds.

Table 2
Threshold $E(P_i) = 0.09$ ($D = 50$, $L0 = 1$).

	$D = 50$																
$E(P_i)$	0.025	0.035	0.045	0.055	0.065	0.075	0.085	0.09	0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18
N	8.9355	22.2258	47.2903	83.2419	142.7903	289.4839	367.6774	512.0908	643.0806	842.6935	997.7581	1271.129	1421.2258	1751.4194	1938.371	2263.6744	2362.7581



(a) $D = 100$, $E(P_i) < 0.09$.

(b) $D = 50$, $E(P_i) \geq 0.09$.

Fig. 6. Relationships of N versus L_0.

Table 3
The value of N when $L_0 = 1$, $D = 100$, $E(P_i) < 0.005$.

$E(P_i)$	0.005	0.004	0.003	0.002	0.001	0.0009
N	1.7692	1.5000	1.4231	1.2308	1.0000	0.6385

(c) When the average activation probability is between 0.045 and 0.085 (less than 0.09), the total number of initiators equal to 3 (degree = 100) is the stability threshold of the increase rate.
(d) When the average activation probability is between 0.09 and 0.018 (degree = 50), the total number of initiators equal to 2 is the stability threshold of the increase rate.

It should be emphasized that the findings described in Sections 4.3 and 4.4 can only be revealed through intensive simulations.

4.5. Discussions

4.5.1. Threshold value of $E(P_i)$

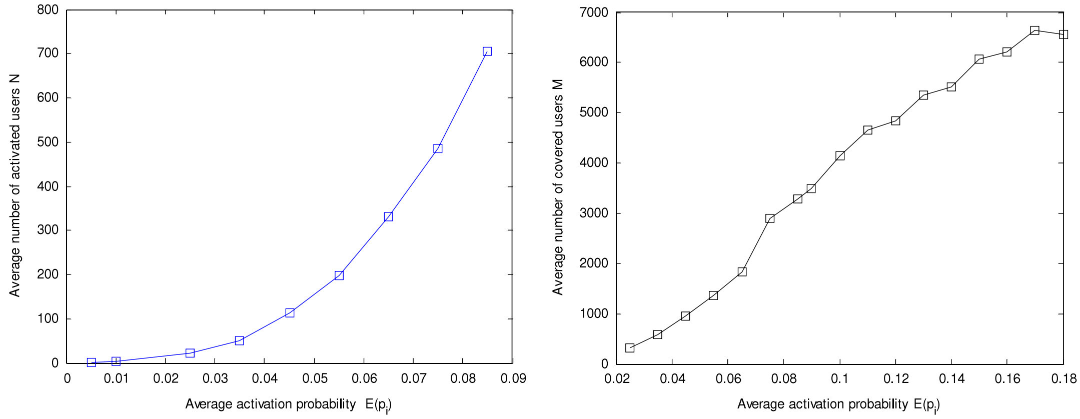
Referring to the theory of the SIS epidemic model [15], it can be pointed out that for exponential networks, the numbers of infection (activation) rate in the network will become zero only when the infection (activation) rate is larger than a certain threshold. Such a threshold value does not exist in power-law networks with non-limited magnitudes. The simulated QQ-based epidemic model here is different from the SIS model, but the threshold of $E(P_i)$ deserves some discussions.

Table 3 and Fig. 7 demonstrate the relationships of the average number of activated users N versus the average activation probability $E(P_i)$ when there is only one initiator with a given and fixed degree. From Table 3 and Fig. 7(a), one can see that when the average activation probability is lower than 0.001, there still exists an activated user, whereas when the average activation probability is 0.001, the average number of activated users is 1. Therefore, 0.001 is the threshold for the average activation probability. Note that this study focuses on the information spreading in viral marketing rather than disease prevention, so the small number of activated users and the corresponding activation threshold value are not very significant in the marketing applications concerned here.

4.5.2. Betweenness

The node betweenness B also indicates the role of the initiator in the network [2]. For Model I, there is only one initiator, simulations were performed on the betweenness B instead of the degree D. Here, the simulation procedure is the same as that using the degree, but the number of nodes of the same betweenness is smaller. Fig. 8(a) shows the relationships of the average number of activated users versus the betweenness and the average activation probability. Fig. 8(b) and Table 4 compares the maximum numbers of activated users when the criterion for choosing the initiator is betweenness and degree, respectively.

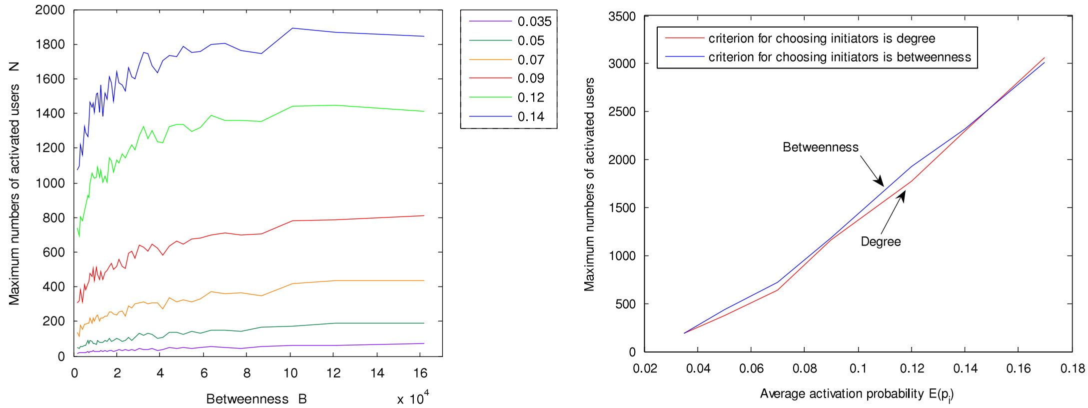
Fig. 8 also shows that, under the condition that the average activation probability is given and fixed, the changes in the number of activated users versus betweenness are the same as the number of activated users versus the degree. Generally speaking, the maximum number of activated users, when the initiator is chosen using betweenness, is a little larger than that when degree is used. Since the computation of betweenness is more complicated, it is convenient to choose the initiator based on its position determined by its degree.



(a) $L_0 = 1, D = 100, E(P_i) < 0.09$.

(b) $L_0 = 1, D = 50, E(P_i) \geq 0.09$.

Fig. 7. Relationships of N versus $E(P_i)$.



(a) Average number of activated users versus betweenness and the average activation probability.

(b) Comparison of maximum numbers of activated users for B and D respectively.

Fig. 8. Comparison results.

Table 4
Comparison of maximum numbers of activated users for D and B.

$E(P_i)$	0.035	0.05	0.07	0.09	0.12	0.14	0.17
N_{max} (for D)	187	372	643	1159	1777	2302	3068
N_{max} (for B)	189	435	726	1182	1928	2315	3054

Furthermore, one can see from Fig. 8(a) that, while the average activation probability is below 0.09, the stable threshold is betweenness $= 2*10^4$, whereas when the average activation probability is above 0.09, the stable threshold is betweenness $= 1 * 10^4$.

4.5.3. *Activation probability*

From the definition of the activation probability P_i, one may note that different users have different activation probabilities at the same time; meanwhile, users have different activation probabilities at different times. In traditional infection propagation models on exponential networks, the activation probabilities of all users are simply set to be identical, and the activation probabilities of the users with the same degree on scale-free networks are also set to be identical [15]. In comparison, the activation probability used in the numerical simulations of this paper is more practical therefore the obtained results are more trustful.

Generally speaking, in the study of complex systems, since the concerned problems are usually too complicated to provide analytic solutions, a typical approach is to drastically simplify the underlying mathematical models, which consequently reduce the reality of the analytic results. In comparison, numerical simulations do not have such drawback, therefore are deemed to be a more effective approach to deal with complex systems.

4.6. Simulation summary

The QQ network discussed in this paper is a small-world network, as verified above. The main findings from the numerical simulations for the proposed marketing scheme using this network are summarized as follows:
(i) The QQ network is a sparse network and the distribution of the node-degrees follow an exponent law. There are many isolated components (subgraphs) on the network, while most users are reachable by the others. The largest component of the network has prominent small-world features for it has a large average clustering coefficient with a short average path length. Every pair of users may contact with each other through no more than two hops. Besides, the degree and the betweenness of user #3324 are both maximum while the average length between user #3324 and the other users is minimum, implying that user #3324 is the center of the QQ network, from the viewpoint of either relationship numbers or centrality of the network.
(ii) When the average activation probability is no larger than 0.18, the average number of activated users will not increase infinitely as the number of the initiators and the degrees of the initiators increase. They finally reach the stable state. Among about ten-thousand users in a network, the maximum number of activated users reaches about 3000 while the state remains stable. On the other hand, the average number of activated users reaches a stable state after about 5 time steps in any case, so the convergence speed is indeed very fast.
(iii) There are a number of factors affecting the average number of activated users over the QQ network. The average activation probability is the most important factor. When the average activation probability is no higher than 0.045, the number of activated users is small, so the positions and the numbers of the initiators are insignificant. However, when the average activation probability is larger than 0.045, the degrees, the betweenness and the number of the initiators all may have positive effects on the increase of the average number of the activated users. When the number of activated users reaches the stable state, the degree, the betweenness and the number of the initiators all have thresholds, and the higher of the average activation probability, the lower the thresholds. As a matter of fact, the new spreading scheme for viral marketing is designed mainly based on these threshold values.
(iv) From the above analysis, it can be seen that the new QQ-based spreading scheme for viral marketing was designed for non-specified users. The initiator cannot be chosen with high centrality in the network, unexpectedly. When the average activation probability is in between 0.045 and 0.085, one should choose the initiator with degree about 100. When the average activation probability is higher than 0.09, one should choose the initiator with degree about 50 instead. When the average activation probability is in between 0.045 and 0.085, however, one should choose 3 initiators (if their degrees are 100), whereas when the average activation probability is in between 0.09 and 0.018, one should choose only 2 initiators (if their degrees are 50). This means that the higher the average activation probability, the smaller the total number of initiators will be needed with lower degrees. Of course, one may also choose and fix a larger number of initiators so as to develop other spreading schemes for viral marketing.

5. Some proposals of spreading schemes for viral marketing based on the simulation results

From the simulations, one can see that the positions of the initiators in the QQ network, the number of users, and the average activation probability of users, are all important factors influencing the effects of QQ-based viral marketing. Among these factors, the average activation probability is the most important one. Therefore, when an enterprise develops an individual QQ-based spreading scheme for viral marketing, it should first forecast the average activation probability on the basis of the marketing merchandise and the features of the marketing objective, and then choose the total number and positions of the initiators based on the average activation probability.

Moreover, the enterprise should routinely study its marketing scheme in order to raise the average activation probability. In our opinion, a user usually spreads information without asking for rewards because he believes that the information distributed has its own value already and is beneficial to him in the end.

As discussed in Section 2.2, viral marketing is a process and our research is only one step in the whole process, namely, the information spreading part. It is believed that the average activation probability determines the effects of viral marketing but there are very few research reports on how to improve the average activation probability, leaving an important topic for future study.

Additionally, it has been assumed in this paper that the initiator should be activated definitely. In reality, to guarantee this one has to consider the initiator's characteristics so as to see whether he has very low immunity against certain kinds of viruses (marketing information). Such analysis should be carried out for individual marketing case studies.

After the implementation of the spreading scheme for viral marketing, one should track its effects thereby to timely learn how the marketing information is spreading and to find out how to improve the performance and to accumulate more experience for further actions.

6. Conclusions

In this paper, we have suggested a spreading scheme for viral marketing for general products aiming at non-specified users. The simulations were performed based on real data that describe a complex network of group-friend relationships in the QQ system, giving meaningful results and conclusions that may be insightful for real-world marketing applications in the future.

For the QQ-based spreading scheme for viral marketing targeting general products fore non-specified users, the findings and the scheme of this paper are constructive, which can be used as direct references. For products aiming at specified users, one may choose some particular QQ user-groups who are interested in those products, sample these groups, establish a corresponding QQ complex network model, and then simulate a corresponding spreading scheme on this complex network. In the case of viral marketing through other types of instant messaging systems, the ideas presented in this study can also be applied to design corresponding marketing schemes, as well as model constructions and simulation studies. In this sense, the new method proposed in this paper is somewhat universal.

Acknowledgement

This research was supported by the National Natural Science Foundation of China (Grant 70773041).

References

[1] Baidu Encyclopedia, Word-of mouth marketing, 2006, http://baike.baidu.com/view/932115.html?wtp=tt.
[2] J. Chevalier, D. Mayzlin, The effect of word of mouth on sales: Online book reviews, Journal of Marketing Research 26 (2006) 345–354.
[3] J.E. Phelps, R. Lewis, L. Mobilio, D. Perry, Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email, Journal of Advertising Research 11 (2004) 333–348.
[4] A. Steyer, R. Garcia-Bardidia, P. Quester, Online discussion groups as social networks: An empirical investigation of word-of-mouth on the internet, Journal of Interactive Advertising 1 (2004) 51–59.
[5] D.S. Sundaram, K. Mitra, C. Webster, Word-of-mouth communication: A motivational analysis, Advances in Consumer Research 25 (1998) 527–531.
[6] M. Bartlett, Blog, the new word of mouth: A look at what's being said about credit unions in cyberspace, Credit Union Journal 13 (2006) 24.
[7] L. Yinan, Study on construction and application of word-of mouth marketing in new market environment, Shandong University, Shandong, 2006.
[8] W. Ma, J. Yang, The application of the instant messages system in the viral marketing, South China University of Technology, Guangzhou, 2008.
[9] R. Douglas, Media Virus, Random House Publishing Group, 1994.
[10] J.F. Rayport, The Virus of Marketing, Fast Company, 1996.
[11] A. Montgomery, Applying quantitative marketing techniques to the Internet, Interfaces 31 (2001) 90–108.
[12] Hotmail Staff, We heard you loud and clear. Microsoft. http://mailcall.spaces.live.com/blog/cns!CC9301187A51FE33!29123.entry.
[13] D. Watts, S. Strogatz, Collective dynamic of small world network, Nature 393 (1998) 440–442.
[14] A.L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[15] X. Wang, X. Li, G. Chen, Theory and Application of Complex Networks, Tsinghua University Press, Beijing, 2006, 72–95.
[16] M.E.J. Newman, The structure and function of complex networks, SIAM Review 45 (2003) 167–256.
[17] J.M. Yang, L.P. Lu, W.D. Xie, et al., On competitive relationship networks: A new method for industrial competition analysis, Physica A 382 (2007) 704–714.
[18] J.M. Yang, W. Wang, G. Chen, A two-level complex network model and its application, Physica A 388 (2009) 2435–2449.
[19] R.F. Wilson, The six simple principles of viral marketing, http://www.wilsonweb.com/wmt5/viral2principles2clean.htm.
[20] J.S. Podoshen, Word of mouth, brand loyalty, acculturation and the American Jewish consumer, Consumer Marketing 23 (2006) 266–282.
[21] B. Gerstley, Advertising research is changing, 2003.
[22] T.T. Ahonen, T. Kasper, S. Melkko, 3G Marketing, John Wiley & Sons, 2004.
[23] T.T. Ahonen, A. Moore, Communities Dominate Brands, Futuretext, 2005.
[24] M. Bampo, M.T. Ewing, D.R. Mather, D. Stewart, M. Wallace, The effects of the social structure of digital networks on viral marketing performance, Information Systems Research 19 (2008) 273–290.
[25] C.C. Zou, D. Towsley, W. Gong, Email virus propagation modeling and analysis. Technical report TR2CSE203204, University of Massachusetts, Amherst, USA, 2003.