

VIRAL MARKETING PROPAGATION ORIENTED TO MARKETING CONTEXT

Completed Research Paper

Li Yu

School of Information, Renmin
University of China
Beijing, P.R. China
buaayuli@ruc.edu.cn

Qiulin Li

School of Information, Renmin
University of China
Beijing, P.R. China
pilgrim_li@163.com

Xun Liang

School of Information, Renmin University of China
Beijing, P.R. China
xliang@ruc.edu.cn

Abstract

Viral marketing exploits social networks to market new product, where users are encouraged to recommend products to their friends. Propagating model is base of other researches on viral marketing, which describes how marketing information is propagated from seed users to other users. In this paper, it is found that currently widely used Independent Cascade (IC) model is not adapted to marketing context where a user will accept recommendation only when the recommendation come from a lot of his friends. Based on the finding, k -order propagating model oriented to marketing context is proposed. Two specific k -order propagating are studied, respectively General_KP and Binary_KP. Using Twitter, Friendster and Random dataset, there 384 experiments are made to show propagating results based on proposed models. The results show that influence order k has important influence on propagating process, which illustrate that k -order propagating model is key to viral marketing.

Keywords: Viral Marketing, Social Network, Influence Propagation, Web2.0

Introduction

In recent years, many social networking sites have emerged on the Web 2.0 platform. Among them are Facebook, Twitter, Xiaonei.net, Taobao, and others, where users interact with one another and explore business opportunities. Social networks have a huge potential in marketing new products, so viral marketing strategies have received a great deal of attention from scholars and enterprises (Leskovec et al. 2006; Cao et al. 2009; Li et al. 2010). In viral marketing, users are encouraged to recommend products to their friends, who would also recommend it to other friends, so that the marketing information is propagated like a contagious disease or a computer virus. Extensive research has been conducted on issues such as influence modeling, influence maximization, and pricing strategy based on influence propagation to support the application of viral marketing. In existing literature, one key element is the propagation model of viral marketing, which describes how marketing information is transmitted from seed users to other users.

The independent cascade (IC) model proposed by Kempe et al. (2003) is the most widely used model of viral marketing. It is based on the concept of disease propagation, which supposes that when a healthy user comes in constant contact with a number of neighbor users who have been infected with a disease (that is, an activated user), one of these neighbors could successfully influence the user; thus, the user will be infected with the same disease. Unfortunately, the hypothesis does not hold true in the marketing context because what usually happens is that a user accepts a friend's recommendation and buys the recommended product only if several (or at least two) of his friends recommend it. If only one friend recommended the product to him, the user will not buy it because he does not have sufficient confidence in the product. Hence, viral marketing does not conform to the hypothesis of disease propagation (i.e., the IC) model.

To accurately distinguish the propagation characteristic of viral marketing, we propose in this paper the k-order propagation model in the context of viral marketing, which extends the current IC model. The IC model can be regarded as a case of the k-order propagation model with $k=1$. Numerous experiments have been conducted to show the propagation results based on the k-order propagation model for difference k, including comparison with the IC model.

The primary contributions of this paper are as follows:

- Indicates that the current propagation model is not applicable to the field of marketing;
- Proposes the k-order propagation model oriented to viral marketing context by introducing the activated level into the propagation model;
- Conducts a series of experiments to simulate the propagation process based on the k-order propagation model;
- Compares the difference between the k-order propagation model and the IC model experimentally.

The remaining parts of the paper are organized as follows. Related work is surveyed in the next section. In Section 3, the influence propagation graph and the traditional propagation model in viral marketing are introduced. In Section 4, the k-order propagation model is proposed, and two k-order propagation models, namely, *General_KP* model and *Binary_KP* model, are presented in detail. Section 5 provides an example to show the propagation results based on *General_KP* model and *Binary_KP* model. In Section 6, experiments are performed to demonstrate the propagation results based on three datasets, and to compare the k-order propagation model with the traditional model. The experiment dataset, experiment setup, and experiment results are presented. Finally, conclusions are drawn and directions for future research are discussed.

Related Work

Viral marketing is a new marketing method that takes advantage of electronic communications (e.g., email) and social networks (e.g., Facebook and MySpace) to trigger cascade adoptions throughout the internet (Leskovec et al. 2006; Bruyn and Lilien 2008; Cao et al. 2009; Li et al. 2010). It is a very controversial field encompassing influence propagation modeling (Kempe et al. 2003; Leskovec et al. 2006), discovery of influential users (Goyal et al. 2008; Li et al. 2010; Trusov et al. 2010), pricing strategies (Arthur et al. 2009; Immorlica and Mirrokni 2010), and influence maximization (Chen et al.

2010).

An influence relationship graph among users must be created to study viral marketing from a social networking perspective. On an e-commerce or social networking site, a user often receives recommendations from friends. Those recommendations result in influence relationships between the user and his friends. Leskovec et al. (2006) first built an influence propagation graph (IPG) based on users' recommendation behaviors.

The propagation model of user influence is a basic and essential topic in viral marketing including the influence propagation graph (Kempe et al. 2003; Bruyn and Lilien 2008; Grabisch and Rusinowska 2008). Kempe et al. (2003) modeled user influence in viral marketing as a directed graph with a vertex representing a user and an edge signifying the influence relationship between two users. Many succeeding studies are based on this directed graph (Goyal et al. 2008; Arthur et al. 2009; Andrew and Toubia 2010). Kempe et al. (2003) proposed two stochastic influence-propagation models, namely, IC and linear threshold (LT), which are two of the most basic and widely studied propagation models today.

Influence maximization is a problem involving the choice of a subset of users for seeding the viral marketing process (Leskovec et al. 2007; Chen et al. 2010). Formally, given an input s , the influence maximization problem in viral marketing is to find a subset $S^* \subseteq V$, such that $|S^*| = s$ and $\delta(S^*) = \max(\delta(S) \mid |S| = s, S \subseteq V)$. This problem is NP-hard, but a constant-ratio approximation algorithm is available. Domingos and Richardson (2008) were the first to study influence maximization as an algorithmic problem. Kempe et al. (2003) were the first to formulate the problem as a discrete optimization problem, designing a natural greedy strategy for maximizing the spread of influence in viral marketing. Many studies, such as those by Leskovec et al. (2007), Chen et al. (2010), and others, have been conducted to either improve the original greedy algorithm or propose new heuristic algorithms for influence maximization.

Influence Propagation Graph and Propagation Model

Influence Propagation Graph

In viral marketing, the purchasing decisions of users are heavily influenced by recommendations and referrals from their friends. The influence relationship among users can result in influence propagation. Theoretically, it is almost impossible to obtain completely accurate data to describe the influence relationship among users. However, such a relationship can be estimated through users' interactive behavior. For example, if Tom always buys a product after knowing that his friend John has bought the same product, we can believe that Tom is influenced by John in purchasing certain products. In particular, John has an influence on Tom if the following two conditions are satisfied: (i) Tom and John have been friends in a social network before they buy a product, and (ii) the time of John's purchase of the product is earlier than that of Tom's. When many products are involved, we can reasonably believe that John has a strong influence on Tom. Based on the above idea, the influence propagation graph can be built, as detailed by Leskovec et al. (2007). An influence propagation graph is a directed graph $G = (U, E, W)$, where the vertices $U = \{u_i \mid i = 1, 2, \dots, N\}$ represent individuals, the edges $E = \{\langle u_i, u_j \rangle \mid i, j = 1, 2, \dots, N\}$ represent relationships, the orientations of the edges indicate the direction of influence, and $W = \{w(u, v) \mid u, v \in U\}$ denotes the influence strength or probability of an individual's influence on another individual.

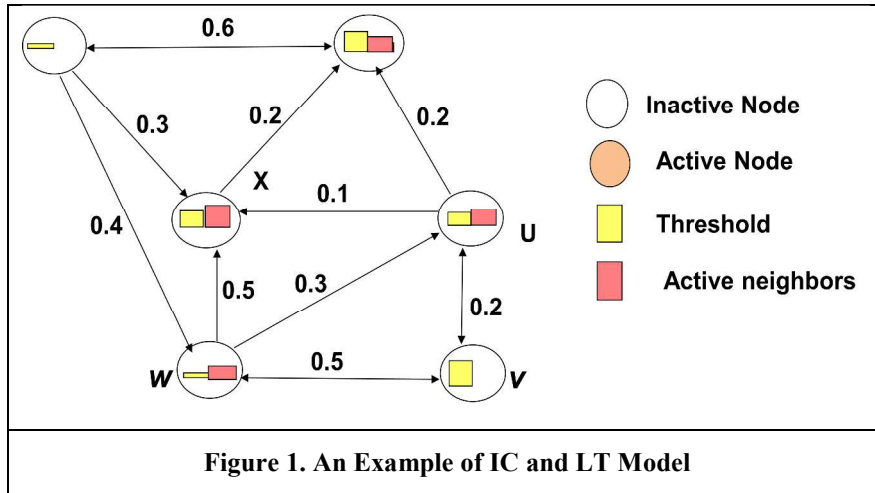
IC model and LT Model

In the IC model, influence is propagated by activated users independently activating their inactive out-neighbor users. If we consider an inactive user u and the set of its activated in-neighbors $Neighbor_u$, to predict whether user u will activate, we need to determine $p_u(Neighbor_u)$. According to the IC model, once an in-neighbor user successfully infects the user u , he will be infected. Based on the idea, the joint influence probability of $Neighbor_u$ on user u is computed,

$$p_u(Neighbor_u) = 1 - \prod_{v \in Neighbor_u} (1 - w_{v,u}) \quad (1)$$

According to the LT model, if $p_u(Neighbor_u) \geq \theta_u$, where θ_u is the activation threshold of user u , user u is

activated. An example is shown in Figure 1.



K-order Propagation Model in Viral Marketing

The k-order propagation model for viral marketing is proposed in this section to accurately describe the difference of the propagation phenomena between viral marketing and disease spreading.

Definition

Before proposing the model, several definitions are given.

Definition 1 Seed User/Node: In the same way that epidemics are caused by initial infections, marketing information propagation in viral marketing generally starts from some people who have been initially influenced to buy the product. These people are called *Seed Users*, or *Seed Nodes* in a social network.

Definition 2 Activated Level: This term is used to show the probability that a user is activated at a certain time.

Definition 3 Activated User: If the probability that a user will be influenced is more than the fixed threshold value θ , the user is an activated user.

Definition 4 Influence Probability: This term is defined as the probability that a user would influence another user. The value comes from the influence propagation graph.

Definition 5 Influence Strength: This term refers to the extent to which a user affects another user. It depends on the user's activated level and influence probability on other users. The influence strength of user u on his neighbor user v , $Influ_Strength(u,v)$, is equal to the activated level of user u by his influence probability on user v ,

$$Influ_Strength(u,v) = w_{u,v} * Activated_Level(u) \quad (2)$$

Definition 6 Influence Spread: For given users S , the influence spread is defined as the expected number of activated nodes by users S , denoted as $\delta(S)$.

Definition 7 Influence Order: For a viral marketing network, influence order is defined as the minimal number of activated users guaranteed to successfully influence their common friend, denoted as k .

For example, an obstinate user may buy a recommended product only when he has received recommendations from five friends (that is to say, $k=5$). By contrast, an easily persuaded user will buy the recommended product even if only one friend has recommended it to him, that is, $k=1$.

In this paper, it is hypothesized that all users have the same k in a viral marketing network. Although this hypothesis is not very reasonable (and will be deleted in our future research) compared with the IC model, it is more applicable to the marketing context. Correspondingly, the propagation process based on the k -order model is more complicated than the one based on the current model, and makes our research more challenging. In fact, when k is equal to 1, the proposed k -order model is degraded into the IC model. That

is, the proposed k -order model extends the IC propagation model from 1 to k .

In the following, two specific k -order propagation models are studied, namely, General k -order Propagation Model (*General_KP*) and Binary k -order Propagation Model (*Binary_KP*). In the *General_KP* model, it is hypothesized that the influence strength of all activated users on their neighbor user is related to his activated level and influence probability on their out-neighbor users, whereas in the *Binary_KP* model, the influence strength of all activated users on their neighbor user is only related to his influence probability on their out-neighbor users. In both models, it is supposed that the influence strength of all non-activated users on other users is zero because their activated level is zero.

General K -order Propagation Model (*General_KP*)

In *General_KP* model, it is key to compute the activated level of a user. In this paper, for user i , his activated level, $Activated_Level(i)$, is determined by the number of users directing to it and the weight of these directed edges. Formally,

$$Activated_Level(i) = 1 - \sum_{s=0}^{K-1} C_{n_i}^s \prod_s w_{ji} \prod_{n_i-s} (1 - w_{hi}), \quad e_{ji}, e_{hi} \in G, j \neq h \quad (3)$$

$$Activated_Level(u) = 1 - \sum_{s=0}^{K-1} C_{n_u}^s \prod_s I_{ju} \prod_{n_u-s} (1 - w_{iu}), \quad e_{ju}, e_{hu} \in G, j \neq i$$

$$Activated_prob(u) = 1 - \sum_{m=0}^{K-1} prob(u, m)$$

$$prob(u, m) = \sum_{\forall (Y_AIN, N_AIN) \in PARTITION(m, u)} \left(\prod_{v \in Y_AIN} Infl(u, v) \prod_{v \in N_AIN} (1 - Infl(v, u)) \right)$$

$$PARTITION(m, u) = \{(Y_AIN, N_AIN) | Y_AIN \cup N_AIN = AIN(u), |Y_AIN| = m\}$$

$$Activated_Level(u) = \begin{cases} 1 & Activated_prob(u) \geq \theta_u \\ 0 & Activated_prob(u) < \theta_u \end{cases}$$

$$Activated_{Level}(u) = \alpha * Activated_prob(u)$$

The algorithm for computing activated level is described in the following.

Input: Network $G=(V,E,W)$, $|V|=N$, Seed Users: S , Threshold: θ
Output: Activated Level: $Active_Level[]$, Activated User: $Activated_User$

```

1  for i = 0 to i = N - 1
2    if  $v_i \in S$  then  $Active\_Level[i] = 1.0$ 
3    else  $Active\_Level[i] = 0.0$ 
4  end for
5  for i = 0 to i = N - 1
6     $Active\_Level'[i] = Active\_Level[i]$ 
7  end for
8
9  for j = 0 to N - 1
10 if  $v_j \in G$  and  $v_j \notin S$ 
11  $Waitlist_j = \emptyset$  //  $Waitlist_j$  store activated nodes who indirect to the node j
12   for i = 0 to i = N - 1
13     if  $e_{ij} \in G$  and  $w_{ij} \geq \theta$ 

```

```

14      $v_i \Rightarrow \text{waitlist}_j$ 
15     end if
16     end for
17      $\text{Active\_level}[j] = 1 - \sum_{s=0}^{K-1} C_{n_j}^s \prod_s \text{Influ}(l, j) \prod_{n_j-s} (1 - \text{Influ}(m, j)), \quad v_l, v_m \in \text{waitlist}_j, l \neq m$ 
18     end if
19     end for
20
21     for  $i = 0$  to  $i = N - 1$ 
22          $\text{Active\_Level}[t] \neq \text{Active\_Level}'[t]$ 
23     end for
24
25     for  $i = 0$  to  $i = N - 1$ 
26         if  $\text{Active\_Level}[i] \geq \theta$ 
27             add user  $i$  to  $\text{Activated\_User}$ 
28         end if
29     end for

```

Figure 2. Algorithm on *Activated_Level* for k-order Propagation Model

According to *General_KP* model, activated level of users in viral marketing network could be higher and higher with propagating from a generation to next generation when activated user nodes form a loop. Let's look an example shown in Figure 3, which is a part of viral marketing network. In the figure, the user nodes L, M, N are activated while other nodes are not presented in the figure.

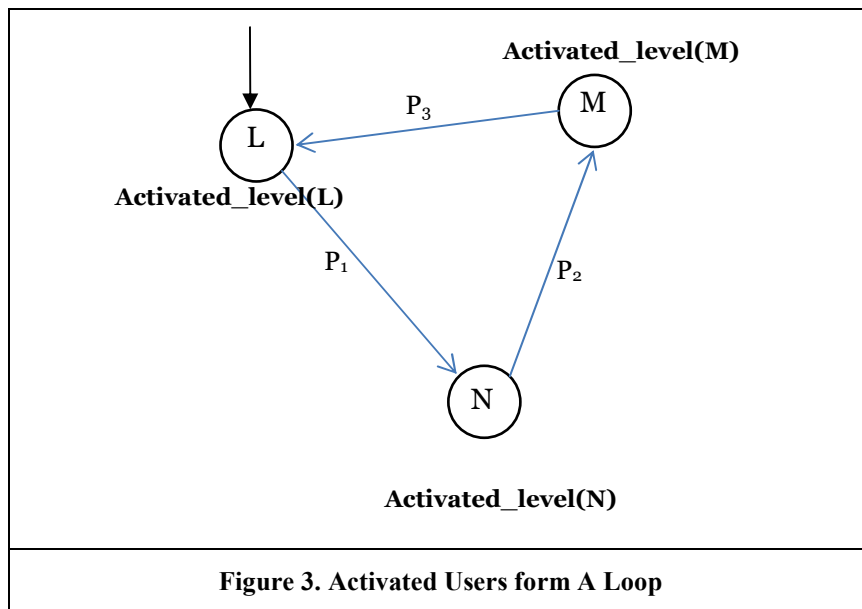


Figure 3. Activated Users form A Loop

When $\text{Activated_level}(L)$ increases Δ , $\text{Activated_level}(N)$ will increase, which will cause an increase in $\text{Activated_level}(M)$, which will result in an increase in $\text{Activated_level}(L)$. That is, when there is a loop in the network, such as $L \rightarrow N \rightarrow M \rightarrow L$ in Figure 3, the activated level of the node from the loop will strengthen circularly. However, we find that although the activated level of a user increases from one generation to the next, the value is convergent, as proven by the following theorem.

Theorem: According to the *General_KP* model, activated level of each user in viral marketing network is convergent.

Proof: Let's consider two kinds of situations.

Situation 1: A node only appears in the single loop.

Assuming there are n nodes (v_1 to v_n) in a loop, the probability of each node successfully affecting the next node is P_i ($i=1, \dots, n$). When $activated_level(V_1)$ increases Δ_1 , $activated_level(V_i)$ increases Δ_1 , then

$$\Delta_1 active_level(V_i) < \Delta_1 active_level(V_1) \times \prod_{s=1}^{i-1} P_s$$

and $activated_level(V_j)$ increases Δ_1 , then

$$\Delta_1 active_level(V_j) < \Delta_1 active_level(V_i) \times \prod_{s=i}^{j-1} P_s < \Delta_1 active_level(V_1) \times \prod_{s=1}^{j-1} P_s$$

Supposing that after the first circulation, $activated_level(V_1)$ increases Δ_2 , then

$$\Delta_1 active_level(V_1) < \Delta_1 active_level(V_j) \times \prod_{s=j}^n P_s < \Delta_1 active_level(V_1) \times \prod_{s=1}^n P_s$$

Here denoting $\prod_{s=1}^n P_s = P$ ($0 < P < 1$), then

$$\Delta_m active_level(V_1) < \Delta_1 active_level(V_1) \times P^m,$$

Because P is between 0 and 1, $\Delta_m active_level(V_1)$ dramatically decreases when m is increasing. So we get

$$\Delta active_level(V_1) < \Delta_1 active_level(V_1) \times (1 + P + P^2 + \dots + P^m) < \Delta_1 active_level(V_1) \times \frac{1 - P^{m+1}}{1 - P},$$

When $m \rightarrow \infty$, then the limit of increment of $activated_level(V_1)$ is small than $\Delta_1 active_level(V_1) \times \frac{1}{1-P}$.

Situation 2: A node appears in several circles.

When in a huge and complex social network, nodes may not just appear in one circle, but several circles. Assuming node V_1 appears in l loops, and the product of affecting probabilities in each loop is denoted as P , that is to say, $\prod_{s=1}^n P_s = P$. When the first loop is stable, $activated_level(V_1)$ increases $\frac{P}{1-P}$. Then the

second loop is stable, $activated_level(V_1)$ increasing $\frac{P}{1-P} + \frac{P^2}{1-P}$. Until the m -th loop is stable, $activated_level(V_1)$ increasing $\frac{P}{1-P} + \frac{P^2}{1-P} + \dots + \frac{P^m}{1-P} = \frac{P}{1-2P} \left(1 + \frac{P^l}{1-P}\right)$, $l \rightarrow \infty$

The limit of $\Delta Active_level(V_1)$ is small than $\frac{P}{1-2P}$. So the activated levels of all the nodes are convergent. ■

Binary K-order Propagation Model (Binary_KP)

In the *Binary_KP* model, it is hypothesized that all activated users have the same influence on their neighbors, whereas inactivated users have no influence at all. This hypothesis means the influence of a user on other users is binary.

$$activated_level(u) = \begin{cases} 1 & \text{if } u \text{ is activated} \\ 0 & \text{if } u \text{ is inactivated} \end{cases} \quad (4)$$

An Example

In this section, a sample viral marketing network is used to illustrate the propagation process based on the k -order propagation model proposed in this paper. Specifically, we show the propagation result for influence order $k=2$ and the activated level threshold $\theta=0.3$. In Figure 4, nodes A, B, and C are selected as seed users. To clearly describe the propagation process, this example assumes that propagation between

any two user nodes occurs at the same time, denoted as t .

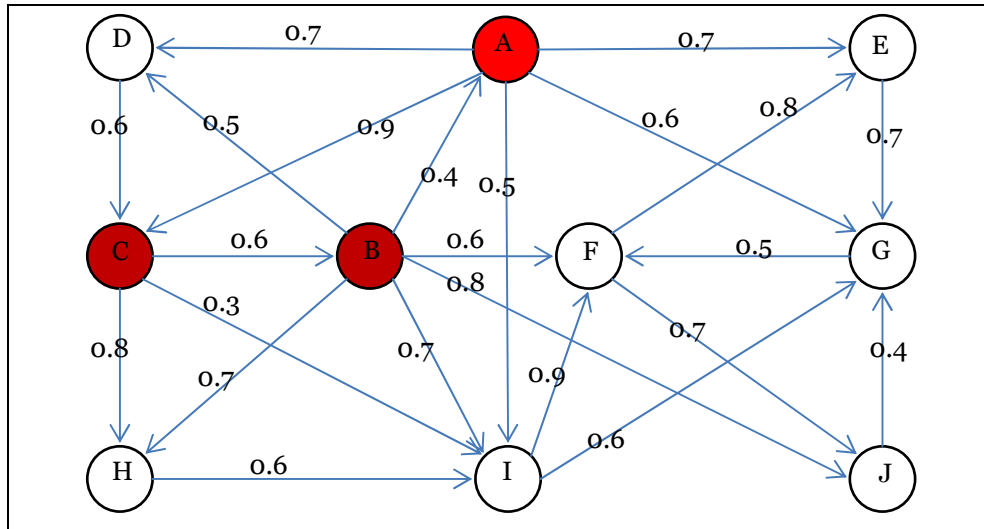


Figure 4. An Example for 2-order Propagating (Nodes A, B and C are seed nodes)

We consider propagating process respectively based on Binary_KP model and General_KP model.

Propagating Result based on Binary_KP Model

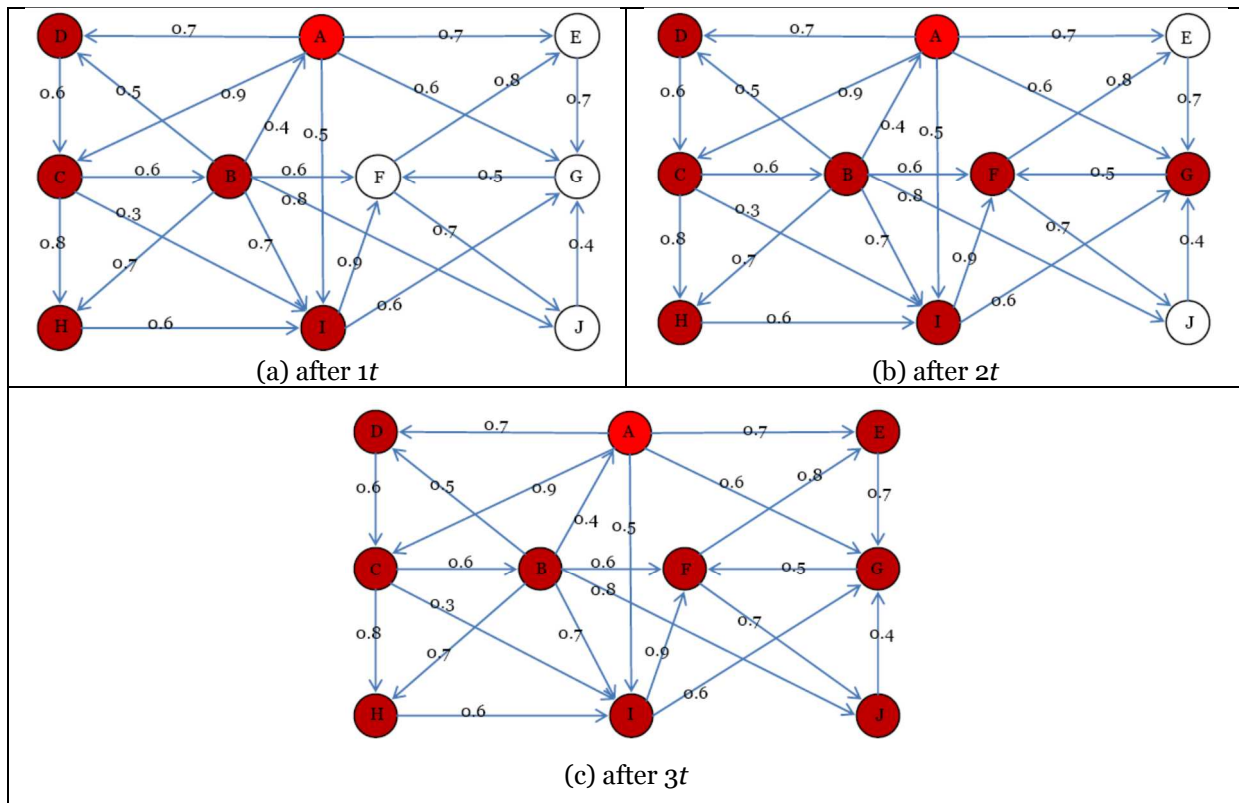


Figure 5. Propagating Result Based Binary_KP Model

In the table 1, activated level of all users for everyday is renewed.

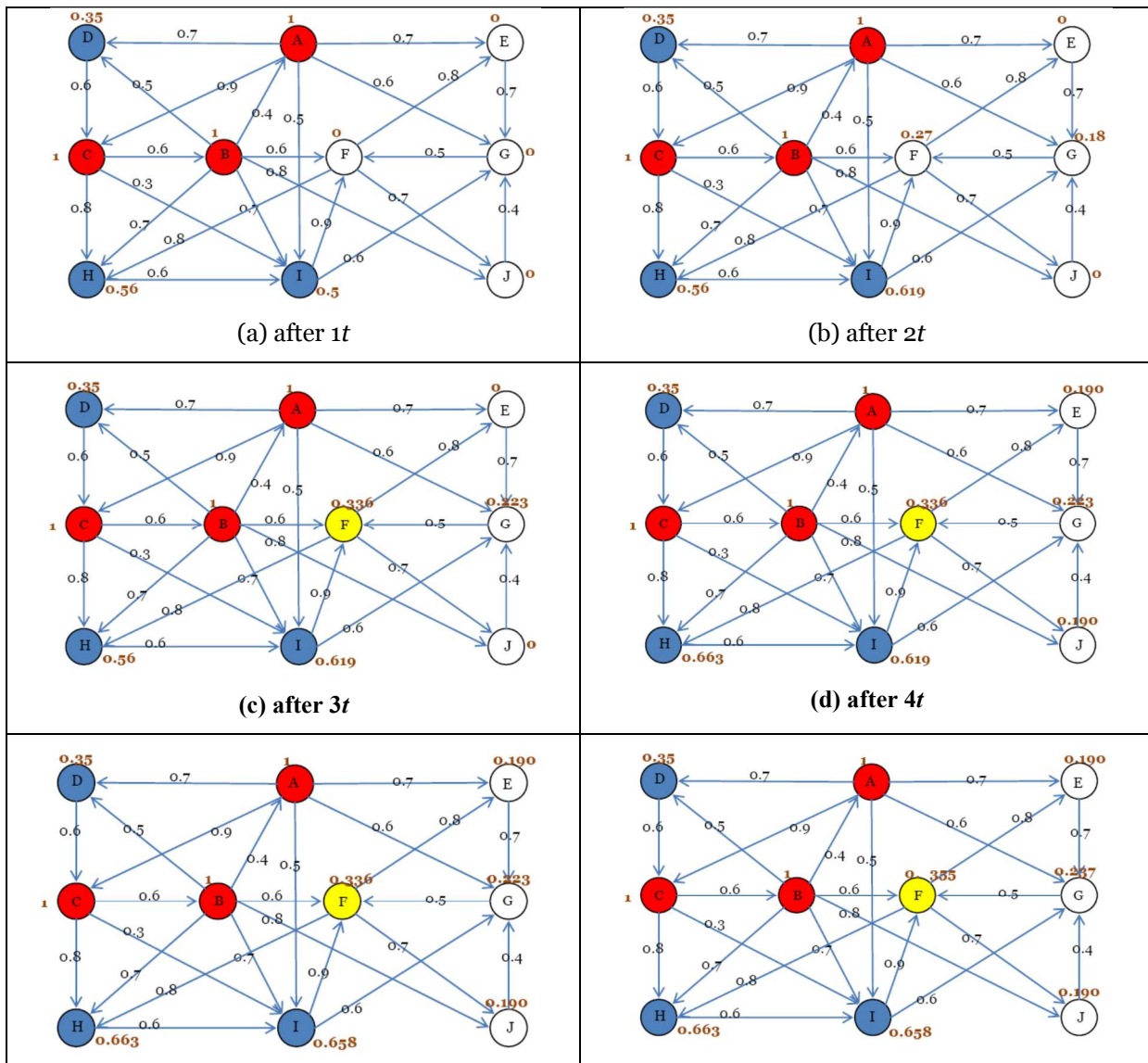
Table 1. Activated Level of All Users Based on *Binary_KP* Model

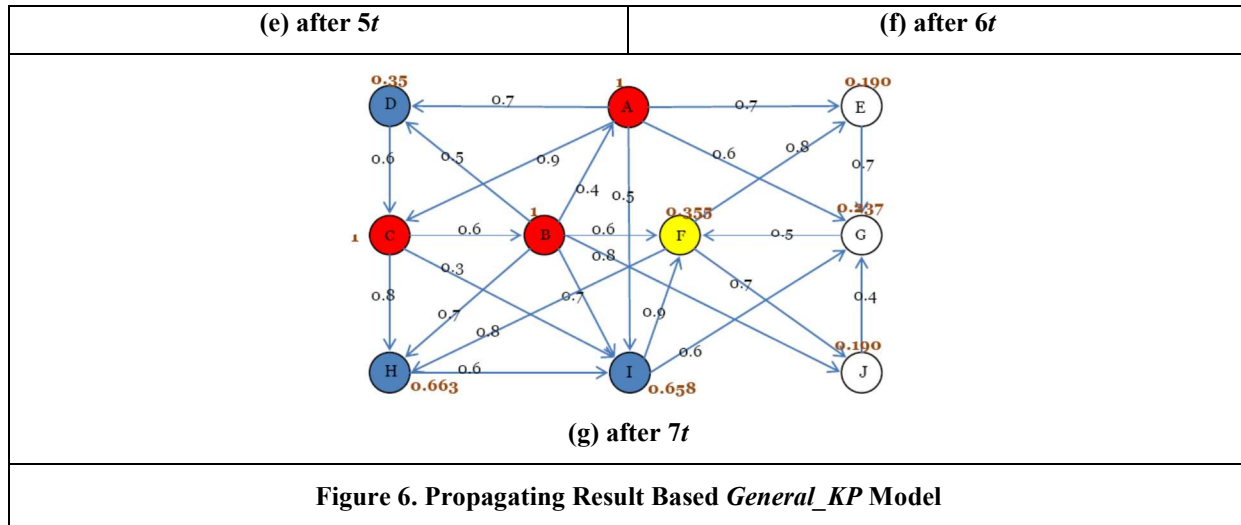
Time	A	B	C	D	E	F	G	H	I	J
initial	1	1	1	0	0	0	0	0	0	0
after 1t	1	1	1	1	0	0	0	1	1	0
after 2t	1	1	1	1	0	1	1	1	1	0
after 3t	1	1	1	1	1	1	1	1	1	1

As shown in this example, if *Binary_KP* model is used, starting from three seed users, all users are activated after three days, when the propagating process ended.

Propagating Result Based on *General_KP* Model

Propagating result Based on *General_KP* Model are shown in Figure 6. As shown in the figure, when after 7t, activated level of all users does not change, including seed user A, B and C, user D, H, I and F are activated. Since no new user is activated, the propagating process ends. Detail activated level of all users after every unit time is shown in Table 2.





Time	A	B	C	D	E	F	G	H	I	J
Initial	1	1	1	0	0	0	0	0	0	0
after 1t	1	1	1	0.35	0	0	0	0.56	0.5	0
after 2t	1	1	1	0.35	0	0.27	0.18	0.56	0.619	0
after 3t	1	1	1	0.35	0	0.336	0.223	0.56	0.619	0
after 4t	1	1	1	0.35	0.190	0.336	0.223	0.663	0.62	0.190
after 5t	1	1	1	0.35	0.190	0.336	0.223	0.663	0.658	0.190
after 6t	1	1	1	0.35	0.190	0.355	0.237	0.663	0.658	0.190
after 7t	1	1	1	0.35	0.199	0.355	0.237	0.668	0.658	0.199
after 8t	1	1	1	0.35	0.199	0.355	0.237	0.668	0.658	0.199

Experiment

In this section, a lot of experiments are made to show the propagation process based on k -order propagation model.

Experiment Dataset

In our experiment, we want to show propagating results process based on k -order propagation model for various social networks with different sparsity of edge, so three kinds of datasets are used to make experiments, respectively *Twitter*, *Friendster* and *Random* dataset. **Twitter** is a social news website. It can be viewed as a hybrid of email, instant messaging and sms messaging all rolled into one neat and simple package. It's a new and easy way to discover the latest news related to subjects you care about. In the dataset, there are 11316811 nodes and 85331846 edges. **Friendster** is a social networking website. The service allows users to contact other members, maintain those contacts, and share online content and media with those contacts. This is the data set crawled by Stephen Boohar (stephen.boohar@asu.edu) on Nov, 2010 from Friendster. It includes 100199 nodes and 14067887 edges. In *Twitter* and *Friendster* dataset, both of them have very few edges. In fact, the sparsity of edges in *Twitter* is $6.6629E-07$ while the sparsity of edges in *Friendster* is 0.00140122 . If the edge is too few, it is hard to propagate from seed nodes even for $k=1$. Considering the reason, we extract sub-dataset with 1000 user nodes including dense edge from both original *Twitter* and *Friendster* dataset. The sparsity of edge in corresponding sub-dataset is respectively 0.01 for *Twitter* and 0.05 for *Friendster*. In addition to *Twitter* and *Friendster* dataset,

Random dataset is randomly generated including 1000 nodes, in which the sparsity of edge is 0.1. For each dataset, the weight of edge is generated randomly ranging from 0 to 1.

Experiment Setup

By using above three dataset, two group experiments are made. The first group experiment is made to compare propagating results for the different number of seed nodes. In this group experiment, activated level threshold is fixed as 0.7. We simulate the propagating based on k -order propagation model when the number of seed node (*Seed_Num*) is respectively 10, 20, 50 and 100. The second experiment is made to study influence of activated level threshold on propagating results. In this group experiment, number of seed nodes is fixed as 20. We simulate the propagating based on k -order propagation model when activated level threshold θ is respectively 0.2, 0.4, 0.6 and 0.8.

For each experiment, two k -order propagation models, *Binary_KP* model and *AL_KP* model are test for different influence order k . If $k=1$, then *Binary_KP* model became traditional propagation model.

Altogether, 384 experiments are made. In each experiment, we show the number of nodes which are finally activated from initial in-activated state. We also want to know that how long time it will be spent to finish whole propagating process when no new node is activated. In this paper, we suppose that it spend one unit's time to propagate between two connected nodes.

Experiment Result

The experiment results on first group experiment are shown in Table 3 for activated level threshold $\theta = 0.7$ while experiment result on second group experiment are shown in Table 4 for number of seed nodes *SeedNum* = 30.

As shown in Table 3, with number of seed nodes increasing, more nodes will be activated. For $k = 2$ in *Binary_KP* and Twitter, when number of seed nodes is respectively 10, 20, 50 and 100, the number of activated nodes is respectively 10, 20, 60 and 968. Additionally, in the whole, with influence order k increasing the number of activated nodes decreases for both *Binary_KP* model and *General_KP* model. It shows that influence order k has great influence on propagating result. When k increases to a limit number, it is hard to propagating and no new node is activated. For example, in Twitter, for *Seed_Num*=10 in both *Binary_KP* model and *General_KP* model, number of final activated nodes is 997 for $k=1$ while the number is 10 for any k more than 1. If the sparsity of edge in dataset is dense, k is more. As shown for *Random*, for *Seed_Num*=10 in both *Binary_KP* model and *General_KP* model, no node is activated when k is 3. There also shown in this experiment that it spend more time to end propagating when there are more seed nodes.

Model	DataSet	Twitter				Friendster				Random			
	Seed_Num	10	20	50	100	10	20	50	100	10	20	50	100
$k=1$ for <i>Binary_KP</i>	Activated_Num	997	997	997	997	1000	1000	1000	1000	1000	1000	1000	1000
	Time	6	5	5	4	3	2	2	2	2	2	2	2
$k = 1$ for <i>General_KP</i>	Activated_Num	997	997	997	997	1000	1000	1000	1000	1000	1000	1000	1000
	Time	7	5	5	4	3	2	2	2	2	2	2	2
$k = 2$ for <i>Binary_KP</i>	Activated_Num	10	20	60	968	1000	1000	1000	1000	1000	1000	1000	1000
	Time	0	0	2	12	7	4	2	2	4	3	2	2
$k = 2$ for <i>General_KP</i>	Activated_Num	10	20	58	183	17	1000	1000	1000	1000	1000	1000	1000
	Time	0	0	1	9	1	4	3	2	4	3	2	2

<i>k</i> =3 for <i>Binary_KP</i>	<i>Activated_Num</i>	10	20	50	101	10	20	1000	1000	10	1000	1000	1000
	<i>Time</i>	0	0	0	1	0	0	4	3	0	4	2	2
<i>k</i> =3 for <i>General_KP</i>	<i>Activated_Num</i>	10	20	50	101	10	20	1000	1000	10	1000	1000	1000
	<i>Time</i>	0	0	0	1	0	0	4	3	0	5	2	2
<i>k</i> =4 for <i>Binary_KP</i>	<i>Activated_Num</i>	10	20	50	100	10	20	57	1000	10	20	1000	1000
	<i>Time</i>	0	0	0	0	0	0	2	3	0	0	3	2
<i>k</i> =4 for <i>General_KP</i>	<i>Activated_Num</i>	10	20	50	100	10	20	56	1000	10	20	1000	1000
	<i>Time</i>	0	0	0	0	0	0	2	3	0	0	3	2

Denote: *Seed_Num* represents the number of seed nodes which are initially used to propagate; *Activated_Num* represents the final number of activated nodes including initial seed nodes after propagating; *Time* represents how long time the propagating process will finish when no new node is activated.

As shown in Table 4, activated level threshold has great influence on propagating results. In the whole, the number of activated nodes will decrease with activated level threshold increasing. For example on *Twitter* based on *General_KP* model with *k* =2, number of activated nodes is 62 for $\theta = 0.2$ while the value is 31 for $\theta = 0.8$. There also shown in Table 4, it spend more time to finish propagating process when θ is larger. It is explained that when θ is larger, it is hard to propagate.

Model	DataSet	<i>Twitter</i>				<i>Friendster</i>				<i>Random</i>			
	θ	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8
<i>k</i> =1 for <i>Binary_KP</i>	<i>Activated_Num</i>	1000	999	998	994	1000	1000	1000	1000	1000	1000	1000	1000
	<i>Time</i>	3	4	4	6	2	2	2	2	2	2	2	2
<i>k</i> =1 for <i>General_KP</i>	<i>Activated_Num</i>	1000	999	998	994	1000	1000	1000	1000	1000	1000	1000	1000
	<i>Time</i>	4	4	5	7	2	2	2	2	2	2	2	2
<i>k</i> =2 for <i>Binary_KP</i>	<i>Activated_Num</i>	997	45	34	31	1000	1000	1000	1000	1000	1000	1000	1000
	<i>Time</i>	7	3	1	1	2	3	3	4	2	2	2	3
<i>k</i> =2 for <i>General_KP</i>	<i>Activated_Num</i>	62	39	34	31	1000	1000	1000	1000	1000	1000	1000	1000
	<i>Time</i>	3	1	1	1	2	3	3	4	2	2	2	3
<i>k</i> =3 for <i>Binary_KP</i>	<i>Activated_Num</i>	31	30	30	30	1000	1000	1000	32	1000	1000	1000	1000
	<i>Time</i>	1	0	0	0	3	4	5	1	2	3	3	4
<i>k</i> =3 for <i>General_KP</i>	<i>Activated_Num</i>	31	30	30	30	1000	1000	1000	32	1000	1000	1000	1000
	<i>Time</i>	1	0	0	0	4	5	7	1	2	3	3	4
<i>k</i> =4 for <i>Binary_KP</i>	<i>Activated_Num</i>	30	30	30	30	1000	37	31	30	1000	1000	1000	35
	<i>Time</i>	0	0	0	0	6	4	1	0	3	4	5	4
<i>k</i> =4 for <i>General_KP</i>	<i>Activated_Num</i>	30	30	30	30	40	34	31	30	1000	1000	1000	33
	<i>Time</i>	0	0	0	0	3	2	1	0	3	4	6	3

In order to understand the difference between *Binary_KP* and *General_KP*, a detailed propagating result based on them are compared for *k*=3, *Seed_Num*=30, $\theta = 0.6$. As shown in Table 5, at any time, there are more activated nodes for *Binary_KP* than *General_KP*. For example, at time 3, there are 137 activated nodes for *Binary_KP* while there are only 62 activated nodes for *General_KP*. It is reasonable because activated level of all activated nodes for *Binary_KP* is one, larger than activated level of corresponding node for *General_KP* model.

Table 5. Comparing Between <i>Binary_KP</i> and <i>General_KP</i> for $k=3$, $Seed\ Num=30$, $\theta=0.6$ and <i>Friendster</i> Dataset		
Time	<i>Binary_KP</i>	<i>General_KP</i>
0	30	30
1	41	41
2	62	51
3	137	62
4	622	90
5	1000	183
6	1000	686
7	1000	1000

Conclusion

Viral marketing is a kind of important social commerce application based on social network where users are encouraged to recommend products to their friends, so marketing information is propagated like a contagious disease or computer virus. The propagation model describes the propagating process of marketing information, which is based and important for viral marketing. In this paper, it is found that traditional propagation model is not adapted to marketing context where it is often that a user receive recommendation from his friends only when the same recommendation come from a lot of his friends. In this paper, two specific k -order propagation model is proposed, respectively *Binary_KP* and *General_KP*. The two models are more adapted to marketing context. In fact, it also extend traditional propagation model which is same as k -order propagation model with $k=1$ for *Binary_KP*. A lot of experiments are made to show propagating result based on our proposed two models. The experiment shows that influence order for viral marketing. In the future, we will study on viral marketing based on k -order propagation model with different activated level threshold for different users.

Acknowledgements

This work was supported by National Natural Science Foundation of China under Grant No.71271209, Humanity and Social Science Youth Foundation of Ministry of Education of China under Grant No.11YJC630268, the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China.

References

- Leskovec J, Adamic LA, Huberman BA. 2006. The dynamics of viral marketing. *Proceedings of the 7th ACM conference on Electronic Commerce*. pp. 228-237.
- Bruyna AD, Lilien GL. 2008. A multi-stage model of word-of-mouth influence through viral marketing. *International Journal of Research in Marketing*. (25:3), pp. 151-163.
- Andrew S, Toubia O. 2010. Deriving value from social commerce networks. *Journal of Marketing Research*. (47:2), pp. 215-228.
- Cao JW, Knotts T, Xu J, Chau M, 2009. Word of mouth marketing through online social networks. *Proceedings of American Conference on Information System*. pp. 291-292.
- Kiss C, Bichler M. 2008. Identification of influencers-measuring influence in customer networks. *Decision Support Systems*. (46:1), pp. 233-253.
- Glen J, Jennifer W. 2002. SimRank: A measure of structural-context similarity. *Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining*. pp. 538-543.
- Goyal A, Bonchi F, Lakshmanan LVS. 2008. Discovering leaders from community actions. *Proceeding of the 17th ACM conference on information and knowledge management*. pp. 179-182.

- Kempe D, Kleinberg J, Tardos E. 2003. Maximizing the spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining*. pp. 137-146.
- Trusov M, Boapati AV, Bucklin RE. 2010. Determining influential users in Internet social networks. *Journal of Marketing Research*, (47:4), pp. 643-658.
- Ioannis A, Hector GM, Chichao C. 2008. SimRank++: Query rewriting through link analysis of the click graph. *Proceedings of the VLDB endowment*. pp. 235-242.
- Arthur D, Motwani R, Sharma A, Xu Y. 2009. Pricing strategies for viral marketing on social networks. *Proceedings of the 5th International Workshop on Internet and network economics*. pp. 101-112.
- Immorlica N, Mirrokni V. 2010. Optimal marketing and pricing over social networks. *Proceedings of the 19th International World Wide Web conference*. pp. 1349-1350.
- Li Y, Lin C, Lai C. 2010. Identifying influential reviewers for word-of-mouth marketing. *Electronic Commerce Research and Applications*. (9:4), pp. 294-304.
- Chen W, Wang C, Wang Y. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*. pp. 1029-1038.
- Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance NS. 2007. Cost-effective outbreak detection in networks. *Proceeding of the 13th ACM International Conference on knowledge discovery and data mining*. pp. 420-429.