Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection

Hao Ma, Haixuan Yang, Michael R. Lyu and Irwin King
Dept. of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
{hma, hxyang, lyu, king}@cse.cuhk.edu.hk

ABSTRACT

Social Network Marketing techniques employ pre-existing social networks to increase brands or products awareness through word-of-mouth promotion. Full understanding of social network marketing and the potential candidates that can thus be marketed to certainly offer lucrative opportunities for prospective sellers. Due to the complexity of social networks, few models exist to interpret social network marketing realistically. We propose to model social network marketing using *Heat Diffusion Processes*. This paper presents three diffusion models, along with three algorithms for selecting the best individuals to receive marketing samples. These approaches have the following advantages to best illustrate the properties of real-world social networks: (1) We can plan a marketing strategy sequentially in time since we include a time factor in the simulation of product adoptions; (2) The algorithm of selecting marketing candidates best represents and utilizes the clustering property of real-world social networks; and (3) The model we construct can diffuse both positive and negative comments on products or brands in order to simulate the complicated communications within social networks. Our work represents a novel approach to the analysis of social network marketing, and is the first work to propose how to defend against negative comments within social networks. Complexity analysis shows our model is also scalable to very large social networks.

Categories and Subject Descriptors: J.4 [**Computer Applications**]: Social and behavioral sciences; H.m [**Information Systems**]: Miscellaneous

General Terms: Algorithms, Theory, Measurement

Keywords: Social Network, Marketing, Heat Diffusion

1. INTRODUCTION

Although *Social Network Analysis* has drawn much attention in the past decades, marketing on social networks has just started, and shows great potential to be much more successful than traditional marketing techniques. According to eMarketer[1], advertisement spending on worldwide social-networking sites in 2007 is expected to reach $1.12 billion, up from $445 million in 2006, and will achieve about $2.8 billion in 2010.

Unlike the traditional social network, online social networks have a number of important distinguishing features. Massive quantities of data are available on online social network sites, blogs, knowledge sharing sites, collaborative filtering systems, newsgroups, email systems, etc. Millions of users participate in these social networks, and act as different roles. All of these social networks provide valuable information for decision-making in marketing campaigns, especially in marketing of new products from start-up businesses. Several successful examples, like Hotmail, Google, MySpace, etc. have already shown the powerful abilities of social network marketing. Full understanding of social network marketing and the potential customers that can thus be reached certainly offer lucrative opportunities for prospective sellers.

Research into how information flows in a social network started from a book called "Diffusion of Innovations" by Rogers [21]. Rogers formalized that adopters of any new innovation could be categorized as innovators (2.5%), early adopters (13.5%), early majority (34%), late majority (34%) and laggards (16%). Some similar work in [7, 25] also focuses on developing theories of innovation adoption.

Many researchers started to analyze the diffusion process in terms of "word-of-mouth" marketing [4, 6, 10, 19], since "word-of-mouth" advertising can be much more effective than traditional marketing methods. Although these studies made great contributions to the analysis of innovation diffusions, they were descriptive, rather than predictive — they are built at a very coarse level, typically with only a few global parameters, and are not useful for making actual predictions of the future behavior of the network [8]. In view of the subsequent exponential growth of online social network resources, researchers now have more data to investigate and model the dynamics of viral marketing, the innovation adoption problem and the information diffusion process [9, 12, 13, 18, 20, 22, 23, 24].

In [9, 20], authors attempt to compute customers' network value as well as the intrinsic value of customers. They focused on the following problem: given a potential social network of consumers, if we can try to convince a subset of individuals to adopt a new product or innovation, and the goal is to trigger a large cascade of further adoptions, which set of individuals should we target [13]? Although

[1]http://www.emarketer.com/

CIKM'08, October 26–30, 2008, Napa Valley, California, USA.

simulations showed their work can help companies achieve much more revenue than direct marketing, several remaining problems still need to be studied, which we will introduce in the next section. Song et al. [23] proposed leveraging users' access patterns to model information flow and generate effective personalized recommendations. Their later work [22] proposed an information flow model that leverages diffusion rates for identifying where information should flow to and identifying who will most quickly receive the information.

A social network is a very complex network with all kinds of messages flowing within it. Modeling social network marketing realistically is an extremely difficult problem. Aiming at the limitations of previous work, we propose a social network marketing framework which utilizes the heat diffusion theory from Physics to describe the diffusion of innovations. These heat diffusion models provide our work with the following contributions: (1) Due to the time-dependent property of the heat diffusion process, our model can simulate product adoptions step by step, which helps marketing companies divide their marketing strategies into several phases. (2) Marketing candidates selection algorithms based on the diffusion models can best represent the clustering coefficient property of real-world social networks. The approximation algorithm is proved to be within a good bound of the optimum solution. (3) Every consumer or customer in social networks can not only diffuse positive comments on products but also can influence others with negative comments even if those others have already adopted those products themselves. Our model captures this important feature of social networks, which has not been studied extensively in previous work, and provides suggestion on planing marketing strategies.

The rest of the paper is organized as follows. We review related work in Section 2. Section 3 proposes several heat diffusion models. In Section 4, we provide three marketing candidates selection algorithms. Section 5 gives the complexity analysis of our proposed models and algorithms. In Section 6, we demonstrate the empirical analysis of our models and algorithms. Finally, conclusions and future work are given in Section 7.

2. RELATED WORK

Rogers theorizes in [21] that innovations spread through society in an S curve, as the early adopters select the technology first, followed by the majority, until a technology or innovation is common. A tremendous expansion has occurred in the marketing literature on diffusion since the 1970s. The most important single impetus to this scholarly explosion is a model for forecasting the diffusion of new consumer products proposed by Frank Bass in [4]. The Bass model characterizes the spread of a new product and technology in a market by

$$N(t) = N(t-1)+p(m-N(t-1)+q\frac{N(t-1)}{m}(m-N(t-1)),$$

where $N(t)$ is the cumulative number of adopters by time t; the parameter m is the market potential, indicating the total number of people who will eventually adopt the item; the coefficient p is called the coefficient of innovation, indicating the external influence or advertising effect; and the coefficient q is called the coefficient of imitation, indicating internal influence or word-of-mouth effect. However, the Bass model is an overly simplified representation of a complex reality, and it ignores the network structure, which could significantly influence the diffusion process [22].

Recently, in order to help companies determine which potential customers to market to, Domingos and Richardson [9, 20] proposed a fundamental algorithm to model customers' network value as well as the intrinsic value of the customer. The customer's network value is defined as the expected profit from sales to other customers this customer may influence to buy, the customers those may influence, and so on recursively.

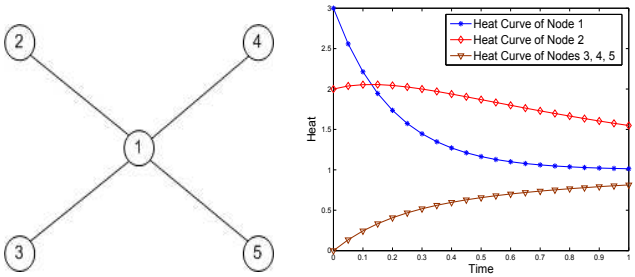
Although simulations show their work can help companies achieve much more revenue than direct marketing, several remaining problems still need to be studied. First, Domingos and Richardson considered making marketing decisions at a specific point in time, whilst in practice, the adoptions of a product by customers happen at different time with different interventions. The time aspect needs to be considered when modeling social network marketing since the social network is evolving from time to time. Second, their work ignores the negative influence of every customer. In practice, one customer will have a negative influence on others (neighbors) if this customer does not like this product. Third, in reality, marketing to individuals with the highest network values is not an appropriate choice since these peoples maybe come from the same community of a social network. The best strategy is to choose marketing candidates from different communities.

In [23], Song et al. proposed to model users' adoption patterns as an information flow network for a recommendation system. An Early Adoption Based Information Flow (EABIF) network model and a Topic-sensitive Early Adoption Based Information Flow (TEABIF) network model were proposed. In [22], Song et al. proposed another information flow model that captures the diffusion rates of information in a network. These two models are well defined, and especially useful for problems in recommendation and ranking. However, these two models need additional data on users' profiles, access patterns, login logs, or purchase logs, which may not always be available.

3. HEAT DIFFUSION MODELS

Heat diffusion is a physical phenomenon. In a medium, heat always flows from a position with high temperature to a position with low temperature. Recently, heat diffusion-based approaches have been successfully applied in various domains such as classification and dimensionality reduction problems [5, 16, 17]. [17] approximated the heat kernel for a multinomial family in a closed form, from which great improvements were obtained over the use of Gaussian or linear kernels. In [16], Kondor et al. proposed the use of a discrete diffusion kernel for categorical data, and showed that the simple diffusion kernel on the hypercube can result in good performance for such data. Belkin et al. employed a heat kernel to construct the weight of a neighborhood graph, and apply it to a nonlinear dimensionality reduction algorithm in [5]. In [28], Yang et al. proposed a ranking algorithm known as the DiffusionRank using heat diffusion process; simulations showed that it is very robust to Web spamming.

In this paper, we model diffusion of innovations as processes of heat diffusion. Actually, the process of people influencing others is very similar to the heat diffusion phenomenon. In a social network, the innovators and early adopters of a product or innovation act as heat sources, and



(a) A small undirected social network

(b) Curve of heat change with time in a small undirected social network

Figure 1: A Simple Example

have a very high amount of heat. These peoples start to influence others, and diffuse their influence to the early majority, then the late majority. Finally, at a certain time point, heat is diffused to the margin of this social network, and the laggards adopt this product or innovation.

The heat flows throughout a geometric manifold with initial conditions can be described by the following second order differential equation:

$$\begin{cases} \frac{\partial f(x,t)}{\partial t} - \Delta f(x,t) = 0, \\ f(x,0) = f_0(x), \end{cases} \tag{1}$$

where $f(x,t)$ is the temperature at location x at time t, beginning with an initial distribution $f_0(x)$ at time zero, and Δf is the *Laplace-Beltrami operator* on a function f [17].

In the light of the several successful existing applications of the heat kernel, it is natural to investigate the heat equation whose special solution is the heat kernel $K_t(x,y)$. The heat kernel $K_t(x,y)$ describes the heat distribution at time t diffusing from the initial unit heat source at position y, and thus describes the connectivity (which is considered as a kind of similarity) between x and y. However, it is very difficult to represent the social network as a regular geometry with a known dimension. This motivates us to investigate the heat flow on a graph. The graph is considered as an approximation to the underlying manifold, and so the heat flow on the graph is considered as an approximation to the heat flow on the manifold.

In this paper, we model a social network as a graph, and each consumer or customer in the social network is defined as a node on this graph. The relationships between peoples are represented by edges that connect nodes. We propose three different diffusion models to describe different social networks: undirected social networks, directed social networks and directed social networks with prior knowledge of their diffusion probabilities.

3.1 Diffusion on Undirected Social Networks

Consider an undirected social network graph $G = (V, E)$, where V is the vertex set, and $V = \{v_1, v_2, \ldots, v_n\}$. $E = \{(v_i, v_j) \mid \text{there is an edge from } v_i \text{ to } v_j\}$ is the set of all edges. The edge (v_i, v_j) is considered as a pipe that connects nodes v_i and v_j. The value $f_i(t)$ describes the heat at node v_i at time t, beginning from an initial distribution of heat given by $f_i(0)$ at time zero. $\mathbf{f}(t)$ denotes the vector consisting of $f_i(t)$.

We construct our model as follows. Suppose, at time t, each node i receives an amount $M(i,j,t,\Delta t)$ heat from its neighbor j during a period Δt. The heat $M(i,j,t,\Delta t)$ should be proportional to the time period Δt and the heat difference $f_j(t) - f_i(t)$. Moreover, the heat flows from node j to node i through the pipe that connects nodes i and j. Based on this consideration, we assume that $M(i,j,t,\Delta t) = \alpha(f_j(t) - f_i(t))\Delta t$, where α is the thermal conductivity-the heat diffusion coefficient. As a result, the heat difference at node i between time $t + \Delta t$ and time t will be equal to the sum of the heat that it receives from all its neighbors. This is formulated as:

$$\frac{f_i(t+\Delta t) - f_i(t)}{\Delta t} = \alpha \sum_{j:(v_j,v_i)\in E} (f_j(t) - f_i(t)), \tag{2}$$

where E is the set of edges. To find a closed form solution to Eq. (2), we express it in a matrix form:

$$\frac{\mathbf{f}(t+\Delta t) - \mathbf{f}(t)}{\Delta t} = \alpha \mathbf{H}\mathbf{f}(t), \tag{3}$$

where

$$H_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \text{ or } (v_j, v_i) \in E, \\ -d(v_i), & i = j, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

In the limit $\Delta t \to 0$, this becomes

$$\frac{d}{dt}\mathbf{f}(t) = \alpha \mathbf{H}\mathbf{f}(t). \tag{5}$$

Solving this differential equation, we have:

$$\mathbf{f}(t) = e^{\alpha t \mathbf{H}}\mathbf{f}(0), \tag{6}$$

where $d(v)$ denotes the degree of the node v, and $e^{\alpha t \mathbf{H}}$ could be extended as:

$$e^{\alpha t \mathbf{H}} = \mathbf{I} + \alpha t \mathbf{H} + \frac{\alpha^2 t^2}{2!}\mathbf{H}^2 + \frac{\alpha^3 t^3}{3!}\mathbf{H}^3 + \cdots. \tag{7}$$

The matrix $e^{\alpha t \mathbf{H}}$ is called the diffusion kernel in the sense that the heat diffusion process continues infinitely many times from the initial heat diffusion.

In order to interpret Eq. (6) and the heat diffusion process more intuitively, we construct a small undirected social network graph with only five consumers as showed in Figure 1(a).

Initially, at time zero, suppose node 1 is given 3 units of heat, and node 2 is given 2 units of heat; then the vector $\mathbf{f}(0)$ equals $[3,2,0,0,0]^{\mathbf{T}}$. The entries in matrix $\mathbf{H}$ are

$$\mathbf{H} = \begin{pmatrix} -4 & 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Without loss of generality, we set the thermal conductivity $\alpha = 1$, and vary time t from 0 to 1 with a step of 0.05. The curve for the amount of heat at each node with time is shown in Figure 1(b). We can see that, as time passes, the heat sources node 1 and node 2 will diffuse their heat to nodes 3, 4, and 5. The heat of nodes 3, 4, and 5 will increase respectively, and the trends of their heat curves are the same since these three nodes are symmetric in this graph.

Now we can interpret this figure in the aspect of social network marketing. Node 1 and node 2 are two consumers representing innovators or early adopters, who influence their direct neighbors in this social network, and to diffuse innovations to others. Eventually, nodes 3, 4, and 5 are successfully influenced by nodes 1 and 2 step by step as time elapses.

3.2 Diffusion on Directed Social Networks

The above heat diffusion model is designed for undirected social networks, but in many situations, the social network graphs are directed, especially in online recommender systems or knowledge sharing sites. Every user in knowledge sharing sites always has a trust list. The users in the trust list will influence this user deeply. These relationships are directed since user a is in the trust list of user b, but user b might not be in the trust list of user a. Based on this consideration, we modify the heat diffusion model on an undirected social network as follows.

On a directed graph $G(V, E)$, in the pipe (v_i, v_j), heat flows only from v_i to v_j. Suppose at time t, each node v_i receives $RH = RH(i, j, t, \Delta t)$ amount of heat from v_j during a period of Δt. We have three assumptions: (1) RH should be proportional to the time period Δt; (2) RH should be proportional to the heat at node v_j; and (3) RH is zero if there is no link from v_j to v_i. As a result, v_i will receive $\sum_{j:(v_j,v_i)\in E} \sigma_j f_j(t)\Delta t$ amount of heat from all its neighbors that points to it.

At the same time, node v_i diffuses $DH(i, t, \Delta t)$ amount of heat to its subsequent nodes. We assume that: (1) The heat $DH(i, t, \Delta t)$ should be proportional to the time period Δt; (2) The heat $DH(i, t, \Delta t)$ should be proportional to the heat at node v_i; (3) Each node has the same ability to diffuse heat; (4) The heat $DH(i, t, \Delta t)$ should be uniformly distributed to its subsequent nodes. The real situation is more complex than this, in view of the dynamic environment of social networks, but we have to simplify these in order to make our model concise. As a result, node v_i will diffuse $\alpha f_i(t)\Delta t/d_i$ amount of heat to each of its subsequent nodes, and each of its subsequent nodes should receive $\alpha f_i(t)\Delta t/d_i$ amount of heat, where d_i is the outdegree of node i. Therefore $\sigma_j = \alpha/d_j$. In the case that the outdegree of node i equals zero, we assume that this node will not diffuse heat to others. To sum up, the heat difference at node v_i between time $t+\Delta t$ and t will be equal to the sum of the heat that it receives, deducted by what it diffuses. This is formulated as

$$\frac{f_i(t+\Delta t) - f_i(t)}{\Delta t} = \alpha \left(-\tau_i f_i(t) + \sum_{j:(v_j,v_i)\in E} \frac{1}{d_j} f_j(t) \right), \quad (8)$$

where τ_i is a flag to identify whether node i has any outlinks, such that $\tau_i = 0$ if node i does not have any outlinks, otherwise, $\tau_i = 1$. Similarly, solving it, we obtain

$$\mathbf{f}(t) = e^{\alpha t \mathbf{H}} \mathbf{f}(0), H_{ij} = \begin{cases} 1/d_j, & (v_j, v_i) \in E, \\ -\tau_i, & i = j, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

3.3 Diffusion on Directed Social Networks with Prior Knowledge

In Section 3.2, we modeled heat diffusion on directed social networks in which each person will diffuse innovation to all of his (her) directed neighbors with equal probability. In the real case, due to limited time and enthusiasm, each person will not diffuse innovation to everyone in his (her) contact list. For example, if you find Gmail or Hotmail is very useful, you probably will tell your best friends or some of your friends who may need this service, but not all of your friends. This consideration motivates us to propose our third heat diffusion model on directed social networks with prior knowledge of diffusion probabilities.

Consider a directed graph $G = \{V, E, P\}$, where V is the vertex set, and $V = \{v_1, v_2, \ldots, v_n\}$. $P = \{p_{ij} \mid$ where p_{ij} is the probability that edge (v_i, v_j) exists$\}$. $E = \{(v_i, v_j) \mid$ there is an edge from v_i to v_j and $p_{ij} > 0\}$ is the set of all edges. If we consider a more general case, we could also include the parameter that describes each person's personality. Some individuals in the social network are very active and willing to share everything they like or dislike to their friends. On the other hand, some individuals are inactive with regard to diffusing innovation to others. We employ ω to describe the personality factor of each person.

According to the above analysis and the analysis in Section 3.2, the expected heat difference at node i between time $t + \Delta t$ and time t will be equal to the sum of the heat that it receives from all its neighbors. This is formulated as

$$\frac{f_i(t+\Delta t) - f_i(t)}{\Delta t} = \alpha \left(-\frac{\tau_i \omega_i}{d_i} f_i(t) \sum_{k:(v_i,v_k)\in E} p_{ik} + \sum_{j:(v_j,v_i)\in E} \frac{\omega_j p_{ji}}{d_j} f_j(t) \right), \quad (10)$$

where τ_i is a flag to identify whether node i has any outlinks. The parameters diffusion probability p and personality factor ω can be any value in the range $[0, 1]$. Solving it, we obtain $\mathbf{f}(t) = e^{\alpha t \mathbf{H}} \mathbf{f}(0)$, where

$$H_{ij} = \begin{cases} \omega_j p_{ji}/d_j, & (v_j, v_i) \in E, \\ -(\tau_i \omega_i/d_i) \sum_{k:(i,k)\in E} p_{ik}, & i = j, \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

3.4 Discussion on α

Parameter α plays an important role in the innovation diffusion process. α is the thermal conductivity, i.e., the heat diffusion coefficient. If it has a high value, heat will diffuse very quickly. Otherwise, heat will diffuse slowly. In the extreme case, if it is infinitely large, then heat will diffuse from one node to other nodes immediately.

Different social networks have different values of α. Information on hot online social network sites and blogs will transfer information faster than other types of social networks. At the same time, different types of information also have different values of α. For an example, bad news or negative information tends to transfer much faster than good news or positive information in real-world social networks.

4. MARKETING CANDIDATES SELECTION

In most cases, each company has a certain quota of product which is used for marketing candidates selection. These products will be delivered to some pre-selected consumers at a discount or totally free. Supposing we have data on a social network which has N individuals, the problem we need to solve is: given the quota number k, how to choose the initial k "influential" individuals who will be delivered a free sample product, in order to maximize the number of cascade adoptions by which these individuals will influence other individuals on their direct contact list.

In this paper, we model social network marketing using heat diffusion processes. Initially, we choose k individuals as the seeds for heat diffusion, denoted by a set S_k, and give a certain amount of heat h_0 to each individual. At time zero of the heat diffusion process, we set $f_i(0) = h_0$, where $i \in S_k$. As time elapses, the heat will diffuse through the whole social network. If the amount of heat of individual i at time

t is greater than or equal to a threshold θ, this individual i will be considered as having been successfully influenced by others, and will adopt the product. We define the *influence set* of a set of k individuals S_k, denoted as $I_{S_k}(t)$, to be the expected number of individuals who will adopt the product at time t. Now the above problem could be interpreted as: finding the most influential k size set S_k to maximize the size of set $I_{S_k}(t)$ at time t, where $I_{S_k}(t) = \{i \mid f_i(t) \geq \theta, i \leq N\}$. This problem is NP-hard, as already proved in [13].

In this paper, three approximation algorithms are proposed to target the initial k-individual set S_k. All latter algorithms are more realistic than former algorithms, but at the same time, are more computationally expensive than former algorithms.

4.1 Top-k Algorithm

Given a social network composed of N individuals, Algorithm 1 shows the steps in finding the top-k influential individuals. The basic idea of this approximation algorithm is: first calculate the influence set of each individual, and then find the k most influential individuals.

Algorithm 1: Top-k Algorithm

```
Input: Graph of a social network; Parameter θ
Output: Top-k influential individuals
foreach Individual i do
    f(0) = 0; f_i(0) = h_0;
    Execute the heat diffusion process f(t) = e^{αtH} f(0);
    foreach Individual j do
        if f_j(t) ≥ θ then
            Add Individual j into set I_i(t)
        end
    end
end
Sort {I_1(t), I_2(t), ..., I_N(t)} by the set size;
Output top-k individuals;
```

4.2 k-Step Greedy Algorithm

Algorithm 1 is very naive since it ignores the potential overlaps of top-k influential sets. We therefore propose a greedy algorithm to minimize the overlaps between top-k influential sets, as follows: (1) First calculate the influence set of each individual, and set $U = \{I_1(t), I_2(t), \ldots, I_N(t)\}$; (2) set $R = \emptyset$, each time choose the set $I_i(t)$, $i \leq N$ to maximize the size of $\{I_i(t) - R \cap I_i(t)\}$, then $R = R \cup I_i(t)$, $U = U - I_i(t)$, until k sets are selected out. The detail of this greedy approximation algorithm is described in Algorithm 2.

Algorithm 2: k-Step Greedy Algorithm

```
Input: Graph of a social network; Parameter θ
Output: k individuals
foreach Individual i do
    f(0) = 0; f_i(0) = h_0;
    Execute the heat diffusion process f(t) = e^{αtH} f(0);
    foreach Individual j do
        if f_j(t) ≥ θ then
            Add Individual j into set I_i(t)
        end
    end
end
U = {I_1(t), I_2(t), ..., I_N(t)}; R = ∅;
for l = 1 to k do
    Select I_m(t) which maximizes {I_m(t) − R ∩ I_m(t)} ;
    R = R ∪ I_m(t); U = U − I_m(t);
    Output Individual m;
end
```

In each step of the greedy implementation, we wish to find the influence set $I_i(t)$ which provides the maximum improvement in terms of the product adopters. If we let $R^*(k)$ denote an optimal solution, using this k-step greedy approximation algorithm, we will show that this algorithm is a $(1 - \frac{1}{e})$-approximation algorithm, where e is the base of natural logarithm, and

$$\frac{R(k)}{R^*(k)} \geq 1 - \left(1 - \frac{1}{k}\right)^k. \tag{12}$$

THEOREM 4.1. *Algorithm 2 is a $(1 - \frac{1}{e})$-approximation algorithm, and $R(k) \geq (1 - (1 - \frac{1}{k})^k) R^*(k)$.*

PROOF. At least $R^*(k) - R(k-1)$ individuals not covered by $R(k-1)$ are covered by the k subsets of $R^*(k)$. Hence, by the pigeonhole principle, one of the k subsets in the optimal solution must cover at least $\frac{R^*(k) - R(k-1)}{k}$ of these individuals. Let r_l denote the subset or influence set selected by the greedy algorithm at step l. From the above analysis, we have $R(1) = r_1 \geq \frac{R^*(k)}{k}$. Thus,

$$\begin{aligned} R(k) &= R(k-1) + r_k \\ &\geq R(k-1) + \frac{R^*(k) - R(k-1)}{k} \\ &= (1 - \frac{1}{k})^{k-1} R(1) + \sum_{i=0}^{k-2} (1 - \frac{1}{k})^i \frac{R^*(k)}{k} \\ &\geq \left(1 - \left(1 - \frac{1}{k}\right)^k\right) R^*(k). \end{aligned} \tag{13}$$

Since as $k \to \infty$, $(1 - (1 - \frac{1}{k})^k) \to (1 - \frac{1}{e})$, Algorithm 2 provides a $(1 - \frac{1}{e})$-approximate solution. □

As increases k, the accuracy of this greedy algorithm will keep decreasing, but the result will converge to $(1 - \frac{1}{e})$ optimal solution.

4.3 Enhanced k-Step Greedy Algorithm

Although the k-step greedy algorithm above can often generate a very good approximate solution to the problem of marketing candidates selection, it still can not interpret the real-world social network activities very well.

In Algorithm 2, we first compute the influence set of each individual in turn, then choose the k sets with maximum coverage. In reality, at the beginning of the innovation diffusion process, several diffusion sources (innovators or early adopters) in the network diffuse the innovation at the same time, not just one single source. The information one person receives from his (her) social network may come from several diffusion sources. We therefore propose Algorithm 3, an enhanced k-step greedy algorithm, to make our model more realistic, although it is more computationally intensive than the above two algorithms.

In Algorithm 3, before the l-th step starts, we first set marketing candidates who are already selected out in the $(l-1)$-th step as the diffusion sources (or heat sources), then launch the greedy search algorithm to find the l-th marketing candidate. This algorithm best preserves the social network properties among all three algorithms, but increases the algorithm time complexity, which we will analyze in Section 5.

Algorithm 3: Enhanced k-Step Greedy Algorithm

Input: Graph of a social network; Parameter θ
Output: k individuals

```
U = {I_1(t), I_2(t), ..., I_N(t)};
A = ∅; R = ∅;
for l = 1 to k do
    I_1(t) = I_2(t) = ... = I_N(t) = ∅;
    HeatDiffusion(A);
    Select I_m(t) which maximizes {I_m(t) − R ∩ I_m(t)} ;
    R = R ∪ I_m(t); U = U − I_m(t);
    Add Individual m into set A;
end
```

Function `HeatDiffusion(`*Set A*`)`

```
foreach Individual i not in set A do
    f(0) = 0;
    foreach Individual p in set A do
        f_p(0) = h_0;
    end
    f_i(0) = h_0;
    Execute the heat diffusion process f(t) = e^{αtH} f(0);
    foreach Individual j do
        if f_j(t) ≥ θ then
            Add Individual j into set I_i(t)
        end
    end
end
```

A common property of social networks is that cliques form, representing circles of friends or acquaintances in which every member knows every other member [1]. This inherent tendency to cluster is quantified by the clustering coefficient in [27]. Our two greedy algorithms also capture this property by maximizing the improvement of coverage at each step, and finally tend to select out the most influential nodes in each community (cluster) as the marketing candidates.

5. COMPLEXITY ANALYSIS

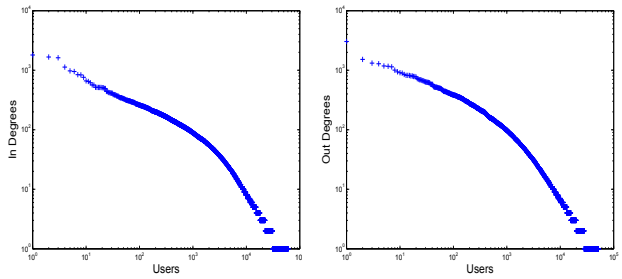
A typical social network always consists of tens of thousands of individuals, and some very large social networks even could reach several millions of individuals. In this section, we will analyze the complexity of our proposed methods, and introduce some very efficient techniques to reduce the complexity, and to ensure our algorithm is scalable for very large social networks.

5.1 Complexity of Heat Diffusion Process

When the graph of a social network is very large, a direct computation of $e^{\alpha t\mathbf{H}}$ is very time-consuming. We adopt its discrete approximation to compute the heat diffusion equation:

$$\mathbf{f}(t) = \left(\mathbf{I} + \frac{\alpha t}{P}\mathbf{H}\right)^P \mathbf{f}(0), \tag{14}$$

where P is a positive integer. In order to reduce the computational complexity, we introduce two techniques: (1) since $\mathbf{f}(0)$ is a vector, we iteratively calculate $(\mathbf{I} + \frac{\alpha t}{P}\mathbf{H})^P\mathbf{f}(0)$ by applying the operator $(\mathbf{I}+\frac{\alpha t}{P}\mathbf{H})$ to $\mathbf{f}(0)$; (2) for matrix $\mathbf{H}$, we employ a data structure which only stores the information of non-zero entries, since it is a very sparse matrix. Thus, supposing a social network is connected by M edges (relationships between individuals), the complexity of executing the heat diffusion process is $O(PM)$, which means the number of iterations P multiplied by the number of edges M in a social network. In most cases, $P = 30$ is enough for approximating the heat diffusion equation. The complexity $O(PM)$ shows that our heat diffusion algorithm has very good performance in scalability since it is linear with respect to the number of edges in social networks.



(a) Indegrees (b) Outdegrees

Figure 2: Degree Distributions of Epinions

5.2 Complexity of Approximation Algorithm

We now consider the three approximation algorithms described in Section 4. Supposing a social network is composed of N individuals and M edges, the complexity for each algorithm is: (1) for Algorithm 1, if the sorting part needs time complexity $N \log N$, the time complexity is $O(N(PM + N + N \log N))$; (2) for Algorithm 2, if the average size of influence set of each individual is d, the time complexity is $O(N(PM + M + kdN))$; (3) for Algorithm 3, the time complexity is $O(kN(PM + N + d))$. We could see that in terms of time complexity, which is the same ranking as for the models' reality, reasonableness and accuracy, Algorithm 3 $>$ Algorithm 2 $>$ Algorithm 1. We will show the detailed comparisons in Section 6.

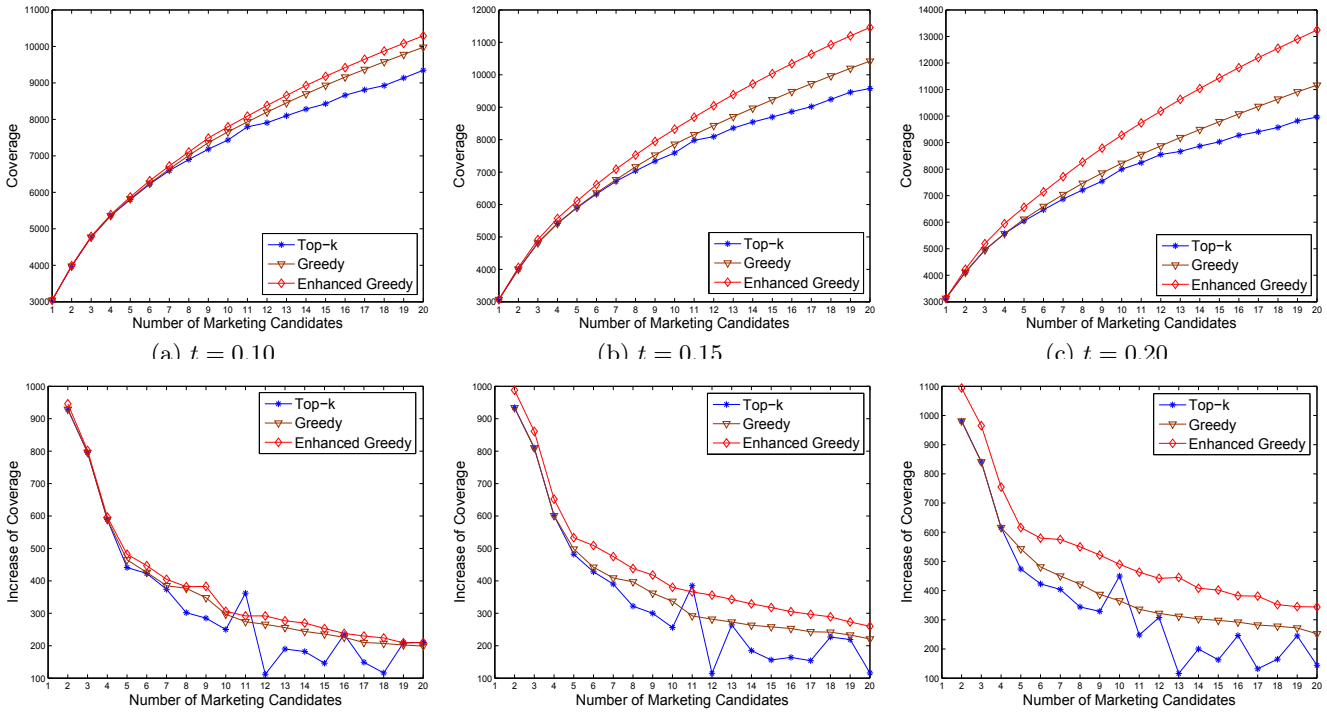
We can employ some techniques to reduce the computation time. In all three algorithms, for each individual, we execute the heat diffusion process, and calculate the influence set. Actually, it is not necessary to compute this for every individual, since the degrees of social networks fit with power-law distribution [2, 3]. A very large number of individuals have very small numbers of neighbors each (outlinks or inlinks). We could assume that selecting these individuals as marketing candidates is not productive. This will greatly decrease the computation time of the algorithms.

6. EMPIRICAL ANALYSIS

We conduct several experiments to measure the performance of our proposed models and algorithms, but due to the space limitation here, we present only the experimental results of the three marketing candidates selection algorithms on the directed social network. For the undirected social network and directed social network with prior knowledge of diffusion probabilities, the results and trends are similar to this. The experiments address the following questions: (1) What is the performance of our marketing candidates selection algorithms? (2) How does influence diffuse in the social network? (3) How should sequential marketing actions be planned? and (4) How to defend against diffusion of negative information? In the following, Section 6.2 to 6.5 answer these questions respectively.

6.1 Dataset

A tremendous amount of data has been produced on the Internet every day over the past decade. Millions of people



(a) $t = 0.10$ (b) $t = 0.15$ (c) $t = 0.20$

(d) $t = 0.10$ (e) $t = 0.15$ (f) $t = 0.20$

Figure 3: Performance of Three Marketing Candidates Selection Algorithms

influence each other implicitly or explicitly through online social network services, such as MySpace, Facebook, Orkut, etc. As a result, there are many online opportunities to mine social networks for the purposes of viral marketing [20].

We choose Epinions[2] as the data source for our experiments on social network marketing. Epinions.com is a well-known knowledge sharing site and review site that was established in 1999. In order to add reviews, users (contributors) need to register for free and begin submitting their own personal opinions on products, companies or movies, etc. These reviews will influence future customers when they are deciding whether a product is worth buying or a movie is worth watching. Every member of Epinions maintains a "trust" list which presents a network of trust relationships between users, and a "block (distrust)" list which presents a network of distrust relationships. This network is called the "web of trust", and is used by Epinions to re-order the product reviews such that a user first sees reviews by users that they trust. Epinions is thus an ideal source for experiments on social networks and viral marketing.

We construct the graph of the Epinions social network by the following rules: (1) We consider only the "trust" relationship between members of Epinions. (2) If a member v trusts another member w, we create a directed link from node w to node v to interpret the trust relationship between these two members. The intuition behind this consideration is that if v trusts w, then w has a very high probability to influence v by word-of-mouth. Thus, we need to scan the trust list of every member in Epinions to build a social network graph. The dataset we employed is one of the 25 product categories, "Kids & Family", as it had the most reviews per product (10.2 on average) and reviews per person who submitted at least one review in the category (5.8, on average). This social network is composed of 75,888 users from Epinions, and 508,960 edges are created between these users. The indegrees and outdegrees of this social network fits with power-law distribution [2, 3] which has been found in many social networks [26]. The degree distributions of Epinions social network are shown in Figure 2(a) and 2(b).

6.2 Experiments on Marketing Candidates Selection

We describe the simulations on evaluating our three marketing candidates selection algorithms first. As stated above, all the experimental results presented here use the heat diffusion model on a directed social network (Section 3.2). Three parameters need to be specified before starting the candidates selection algorithms. The first parameter is the initial heat vector $\mathbf{f}(0)$, and this vector determines how much heat the heat sources need. For the purpose to let heat sources have enough heat to diffuse, we need to allocate a relatively large heat value to heat sources. We choose $\frac{N}{k}$ as the amount of heat for each heat source, where N is the number of consumers in this Epinions dataset, and k is the number of marketing candidates (heat sources). This consideration indicates that the average amount of heat of every consumer is 1. The second parameter is the thermal conductivity value α, which controls the heat diffusion rate of our model. We set $\alpha = 1$ in all of our experiments. The third parameter is the adoption threshold θ. If at time t, one consumer's heat value is greater or equal to θ, we consider this consumer to adopt this product. Actually, in real life, consumers may have different adoption threshold values, but in this paper, in order to simplify the model, we set $\theta = 0.6$.

We employ the term "coverage", which denotes how many consumers adopt the products, to evaluate our three proposed algorithms. Figure 3 shows the simulation results at $t = 0.10$, $t = 0.15$ and $t = 0.20$ under the following

[2]http://www.epinions.com/.

Table 1: IDs of Marketing Candidates Selected at $t = 0.10$, $t = 0.15$ and $t = 0.20$

$t = 0.10$	Steps	1	2	3	4	5	6	7	8	9	10
	Top-k	18	143	737	790	136	1179	1719	118	4416	780
	Greedy	18	143	737	790	27	136	1179	4415	1719	2239
	Enhanced Greedy	18	143	737	790	27	136	1719	1179	4415	2239
	Steps	11	12	13	14	15	16	17	18	19	20
	Top-k	27	128	1516	34	40	791	1	28	1619	1621
	Greedy	118	4416	1753	791	4969	1619	725	18955	125	763
	Enhanced Greedy	118	4416	1753	791	1619	4969	125	725	849	763
$t = 0.15$	Steps	1	2	3	4	5	6	7	8	9	10
	Top-k	18	143	737	790	1179	136	1719	118	4416	780
	Greedy	18	143	737	790	27	136	1179	4415	1719	2239
	Enhanced Greedy	18	143	737	790	27	136	1719	4415	1179	2239
	Steps	11	12	13	14	15	16	17	18	19	20
	Top-k	27	128	791	1516	40	34	1	1619	1621	28
	Greedy	118	1753	4416	791	763	4969	18955	17991	1619	776
	Enhanced Greedy	4416	1753	118	791	1619	1621	4969	763	1749	18955
$t = 0.20$	Steps	1	2	3	4	5	6	7	8	9	10
	Top-k	18	143	737	790	1179	136	1719	118	4416	27
	Greedy	18	143	737	790	27	4415	2239	12642	136	1719
	Enhanced Greedy	18	143	737	790	136	27	4415	2239	1753	1719
	Steps	11	12	13	14	15	16	17	18	19	20
	Top-k	780	791	128	1516	40	1619	28	1	1621	34
	Greedy	1179	1753	791	22381	763	4969	118	18955	5313	776
	Enhanced Greedy	1179	791	4416	1621	1619	4969	118	4282	1749	5144

scenario: if we are given 1 to 20 product samples ($k = 20$), who we should choose as the marketing candidates and what is the performance (measured by the value of coverage) if we choose these candidates. Figure 3(a) and Figure 3(d) show the simulation results at time $t = 0.10$. The star, down triangle and diamond in solid line represent the Top-k algorithm, k-step Greedy algorithm and Enhanced k-step Greedy algorithm, respectively. In Figure 3(a), 3(b) and 3(c), the x-axes denote the number of marketing candidates we are given, and the y-axes present the resulting coverage. We can easily draw the conclusion from these results that the Enhanced k-step Greedy algorithms has the best performance among the three algorithms. Figure 3(d), 3(e) and 3(f) show the increase of coverage of each selection step compared with the previous selection step. Since the Top-k algorithm does not consider the overlap between each influence set, we can observe that its curve fluctuates sharply. This indicates that the Top-k algorithm is a very naive algorithm and it does not performs well in practice. Generally, the curves of k-step Greedy algorithm and Enhanced k-step Greedy algorithm are much smoother than that of the Top-k algorithm; also the Enhanced k-step Greedy algorithm causes more consumers to adopt products than the k-step Greedy algorithm.

We also list the IDs of corresponding consumers who are selected as the marketing candidates by our algorithms in Table 1. The ID numbers range from 0 to 75,887 since we have 75,888 users in Epinions social network totally. We observe that the first four selected candidates (IDs are 18, 143, 737 and 790) are the same for all of these algorithms. This phenomenon matches the real life situation that a social network is a network composed of several different communities or clusters, and every community has some very influential persons. These four persons are very likely to be belong to four different communities and fulfil important roles in these communities, like the authoritative entities in a social network which can be identified by the HITS algorithm [14, 15]. Beyond the first four marketing candidates, however, the remaining candidates chosen by each of our three algorithms are different. The above analysis indicates that our proposed framework can best represent the clustering coefficient property of real-world social networks.

6.3 Diffusion of Influence

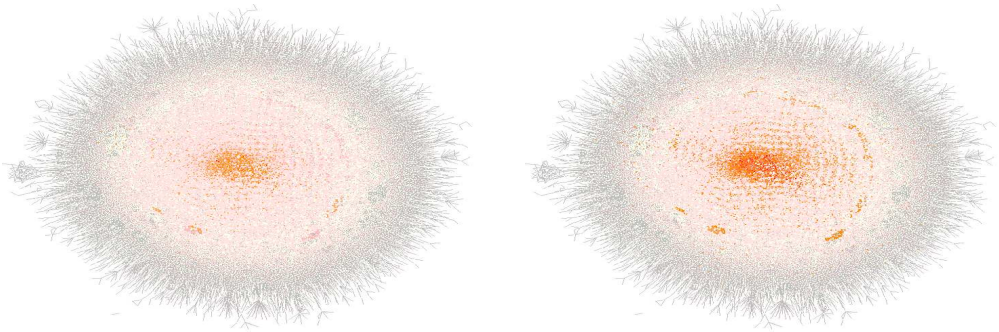
In the preceding section, we select 20 marketing candidates out of 75,888 consumers using each of our three selection algorithms. These 20 consumers will be treated as the diffusion sources, and start to influence others through word-of-mouth process. More and more consumers will adopt this product. Due to the large number of consumers, it is impractical to list here all the detailed data on how influence diffuses in this social network. Hence, in this section, we illustrate this process in a simple, clear and visual way.

First we plot out the Epinions social network graph. Every relationship in this social network is represented by a gray line. At time $t = 0$, we color the 20 marketing candidates which are selected by Enhanced k-step Greedy algorithm in dark red using RGB value. When the time reaches $t = 0.2$, as Figure 4(a) shows, more and more consumers are influenced by their neighbors(consumers colored dark red have higher probabilities to adopt this product than consumers colored light red; the degree of darkness is computed using our heat diffusion models). In Figure 4(b), when $t = 0.5$, we observe that more consumers tend to adopt this product than when $t = 0.2$. This phenomenon coincides with the intuition that the number of adopters will grow rapidly through a word-of-mouth process in real-world.

6.4 Sequential Marketing Actions

So far, our algorithms plan the marketing strategies only once; actually most companies hope to plan a marketing strategy step by step to satisfy the market demand. Suppose we have k samples for marketing a product, and we wish to deliver a quota $k_0, k_1, \ldots, k_{n-1}$ of samples to preselected candidates at specific time point $t_0, t_1, \ldots, t_{n-1}$, where $k_0 + k_1 + \ldots + k_{n-1} = k$. The problem is how to design the marketing strategy and how to choose the marketing candidates to maximize the adoptions of this product.

Since our diffusion models have the advantage of being time-dependent, we can easily design marketing strategies



(a) $t = 0.20$ (b) $t = 0.50$

Figure 4: Graph Demo on Diffusion of Influence

at different time points. Suppose at time $t_0 = 0$, we deliver k_0 samples to k_0 consumers selected by our algorithms, denoted by a set C_{t0}, and from time t_0 to time t_1, several other consumers adopt this product, denoted by a set A_{t0-t1}. We could observe that a total of $\{U - (C_{t0} + A_{t0-t1})\}$ consumers have not adopted this product at time t_1, where U is the set of all consumers. We then start our next step marketing action by selecting C_{t1} candidates from these $\{U - (C_{t0} + A_{t0-t1})\}$ consumers, and marketing products to them. Thus, at time point t_2, A_{t1-t2} consumers are successfully influenced by others. We repeat this process until all the actions are executed. In real life, n is often less than 3.

We have also conducted several experiments to simulate the sequential marketing actions at different time points; we omit the details here due to space limitations. In general, we observe that the above heuristic algorithm ensures that our marketing strategies cover the greatest number and most diverse possible range of consumers.

6.5 Diffusion of Negative Information

Social networks are very dynamic and complex networks. All kinds of information flows on social networks; we can simply classify this information as positive and negative. The presence of negative information makes modeling social network marketing extremely difficult; no previous models have considered negative information. Although in [11], a propagation model of trust and distrust is proposed, this model tries only to answer the question of why people trust and distrust others, and does not address the question of how positive and negative comments diffuse.

Previous diffusion models all assume that people will always positively recommend a product to their friends if they adopt it. Actually, in real life, one has a high probability to tell one's friends not to buy a product if he or she feels that this product is not good enough after he or she used it. Unfortunately, no previous work has taken account of this serious problem.

In this paper, our proposed diffusion models can naturally simulate product adoptions in the presence of both positive and negative comments. At the beginning of the marketing candidates selection algorithms, based on some prior knowledge, if we find a consumer u does not like the product that will be marketed to him(her), this consumer will be allocated a negative value of heat $f_u(0)$ by our algorithms. Thus, this consumer will diffuse negative comments to his (her) friends or neighbors in his (her) social network. At time point t, a consumer v in this social network will receive different types of comments from their neighbors. Some are positive and others are negative. Consumer v will decide whether to adopt this product or not depending on the heat diffusion equation $\mathbf{f}(t) = e^{\alpha t \mathbf{H}}\mathbf{f}(0)$. If $f_v(t) \geq \theta$, consumer v will adopt this product, and if $f_v(t) < \theta$, consumer v will not adopt this product. Another possibility is $f_v(t) < 0$, which means consumer v is persuaded by the negative comments, and will diffuse negative comments to his (her) neighbors too.

How can the influence of negative comments be minimized? We utilize a heuristic algorithm to try to alleviate the impact of negative comments as much as possible. Before describing this algorithm, we make the following assumptions: (1) Determining whether a consumer likes or dislikes a specific product needs some prior knowledge on this consumer's previous profiles, such as purchase record, reviews and feedback. Due to the lack of such data, we preassign some consumers to diffuse negative comments. (2) In order to simplify the model, we assume that only the first selected marketing candidates could diffuse negative comments, which means, at time t, if a consumer adopts this product, in the remaining time, this consumer will only diffuse positive comments to others.

Now the heuristic defense algorithm on how to choose marketing candidates to defend against negative information can be described as follows: (1) First select k marketing candidates using the k-step Greedy algorithm or the Enhanced k-step Greedy algorithm, and suppose some of these candidates dislike this product, and will be allocated negative initial heat values. (2) For each candidate who dislikes the product, select an additional k_a candidates who have the most common influence sets with this candidate as the complementary diffusion sources. These complementary candidates can alleviate the damage that the negative comments cause, or even eliminate it.

Suppose we are given 10 product samples to market to consumers, and we select 10 candidates based on the Enhanced k-step Greedy algorithm. If no-one among these candidates diffuses negative comments on this product, the coverage curve or adoption curve is shown in Figure 5 as the star curve in a solid line. If the first candidate dislikes this product, he(she) will diffuse negative information to his (her) friends. Since the first candidate is the most influential node in this social network, the adoption curve will drop substantially, as shown in the down triangle curve in Figure 5. We then execute the defense algorithm described above, and suppose $k_a = 2$, which means we select two defense candidates to alleviate the negative impact as much as possible. From the diamond curve in Figure 5, we see that this defense algorithm works well and alleviates the damage

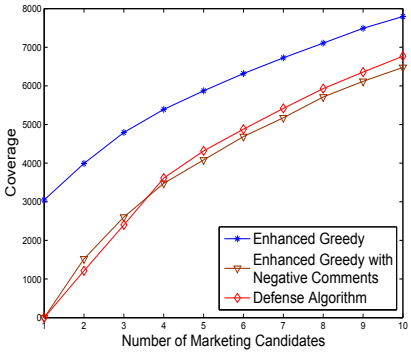


Figure 5: Defense Against Negative Information

caused by the negative comments. The reason why the diamond curve will drop a little at the beginning is that the two defense candidates are employed to alleviate the damage from the first candidate; however, eventually, they will increase the number of product adoptions. Since the social network is really a very complicated network, we have illustrated only a simple example of how to defend against negative information in this paper. A more detailed and extensive analysis of diffusion of negative information will be included in the future work.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a social network marketing framework which includes three diffusion models and three marketing candidates selection algorithms. The purpose of our work is to model social network marketing as realistically as possible. Our proposed methods have several advantages compared with previous work, including how to defend against diffusion of negative information, which has not been explored previously. The complexity analysis shows our framework is scalable for very large social networks, and an empirical study using the Epinions dataset confirms the promise of our proposed framework.

Although we have developed some models and algorithms for social network marketing, several remaining issues need to be studied in the future. We coin the concept of "negative information diffusion" in this paper, and introduce how to defend against this kind of information by employing a simple heuristic algorithm. The understanding of how negative information is diffused is still at a crude level, and more theoretical analysis will be conducted in the future.

As we mentioned in this paper, a social network is a very complicated network, and that is why we always try to model social network marketing as accurately as possible within a simple model. An important property of any social network is evolution. Every social network is evolving all the time. So far, our work considers social network as a static network only, and ignores newcomers, new relationships between existing members and the growth of the network's size. In the future, we plan to consider the evolution property of social networks, and permit our social network to grow at a certain rate. The power-law distribution property of social networks and some link prediction algorithms may be helpful in modeling dynamic social networks.

8. ACKNOWLEDGMENTS

The work described in this paper was fully supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK4150/07E and GRF #412507).

9. REFERENCES

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[2] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world wide web. *Nature*, 401:130–131, 1999.

[3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[4] F. Bass. A new product growth model for consumer durables. *Management Science*, 15:215–227, 1969.

[5] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[6] J. Brown and P. Reinegen. Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14(3):350–362, 1987.

[7] J. Coleman, H. Menzel, and E. Katz. *Medical Innovations: A Diffusion Study*. Bobbs Merrill, 1966.

[8] P. Domingos. Mining social networks for viral marketing. *IEEE Intelligent Systems*, 20(1):80–82, 2005.

[9] P. Domingos and M. Richardson. Mining the network value of customers. In *Proc. of the ACM SIGKDD Conf.*, pages 57–66, 2001.

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

[11] R. V. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proc. of the ACM WWW Conf.*, pages 403–412, 2004.

[12] J. D. Hartline, V. S. Mirrokni, and M. Sundararajan Optimal marketing strategies over social networks. In *Proc. of the ACM WWW Conf.*, pages 189–198, 2008.

[13] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. of the ACM SIGKDD Conf.*, pages 137–146, 2003.

[14] J. Kleinberg. Authoritative sources in a hyperlinked environment. In *Proc. of the ACM-SIAM SODA Conf.*, pages 668–677, 1998.

[15] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of ACM*, 46(5):604–632, 1999.

[16] R. I. Kondor and J. D. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *Proc. of ICML Conf.*, pages 315–322, 2002.

[17] J. D. Lafferty and G. Lebanon. Diffusion kernels on statistical manifolds. *Journal of Machine Learning Research*, 6:129–163, 2005.

[18] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, 1(1), 2007.

[19] V. Mahajan, E. Muller, and F. Bass. New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54(1):1–26, 1999.

[20] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proc. of the ACM SIGKDD Conf.*, pages 61–70, 2002.

[21] E. M. Rogers. *Diffusion of Innovations (5th ed.)*. Free Press, New York, 2003.

[22] X. Song, Y. Chi, K. Hino, and B. L. Tseng. Information flow modeling based on diffusion rate for prediction and ranking. In *Proc. of the ACM WWW Conf.*, pages 191–200, 2007.

[23] X. Song, B. L. Tseng, C.-Y. Lin, and M.-T. Sun. Personalized recommendation driven by information flow. In *Proc. of the ACM SIGIR Conf.*, pages 509–516, 2006.

[24] M. R. Subramani and B. Rajagopalan. Knowledge-sharing and influence in online social networks via viral marketing. *Communications of the ACM*, 46(12):300–307, 2003.

[25] T. Valente. *Network Models of the Diffusion of Innovations*. Hampto Press, 1995.

[26] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK, 1994.

[27] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[28] H. Yang, I. King, and M. R. Lyu. DiffusionRank: a possible penicillin for Web spamming. In *Proc. of the ACM SIGIR Conf.*, pages 431–438, 2007.