Mathematical Modeling of Advertisement and Influence Spread in Social Networks

Bridge. Zhao, Y.K. Li, John C.S. Lui
Department of Computer Science & Engineering
The Chinese University of Hong Kong
{qzhao,ykli,cslui}@cse.cuhk.edu.hk

Dah-Ming Chiu
Information Engineering Department
The Chinese University of Hong Kong
dmchiu@ie.cuhk.edu.hk

ABSTRACT

We consider the following advertisement problem in social networks. Given a fixed advertisement investment (e.g., provide free samples to small number of users), a company needs to determine the probability that users of a social network will eventually purchase the product. In this paper, we model social networks as scale-free graphs (with or without high clustering coefficient). We then characterize and model various influence mechanisms that govern the influence spreading in large scale social networks. We use the local mean field (LMF) technique to analyze these social networks wherein states of nodes can be changed by various influence mechanisms. Extensive simulations are carried out to validate the accuracy of our model. These results can provide insight in designing efficient advertising strategies in social networks.

1. INTRODUCTION

In recent years, advertising has become a major commercial activity in the Internet. Traditionally, advertisements are usually broadcast oriented, e.g., via TV or radio stations so as to reach as many people as possible. With the development of the Internet, new advertisement models emerge and blossom. For example, Google provides the *targeted* advertisements: when a user searches for information, related advertisements are returned together with the search results. Such targeted advertisement can enhance the success rate of selling a product. In recent years, social networks offer another new venue of performing advertisement. In social networks, users are logically grouped together by one or more specific types of interdependency such as friendship, values, interests, ideas, . . ., etc. Since the dependency is quite strong, if one user decides to purchase a product, he/she can influence his/her friends, and thereby increase the possibility of sales. With the success of online social networks like Facebook and Myspace, advertising in social networks is receiving more attention.

To advertise on social networks, a company first applies advertising strategies, either traditional or Internet-based, targeted or non-targeted, so as to attract a small fraction of the social network users to purchase the product. Based on this initial fraction of buyers, a cascade of word-of-mouth influence by users is triggered, and eventually large fraction of users in the social networks may decide to purchase the product.

Predicting the final portion of buyers is important for companies since they can design efficient advertising strategies so as to maximize their revenue. However, this is not an easy task since the influence depends on various factors that are difficult to characterize. The first important factor is the topology of the network. Intuitively, a well connected network may allow the influence to spread to more users. But to what extent the connectivity helps is not clear. Moreover, how other topological properties of the network, say, randomness, degree distribution, may affect the spreading of influence is unknown. The second important factor is the mechanism that determines whether a user will purchase the product. In general, the better comment his/her friends give to the product, the more likely the user will purchase the product. But how to characterize such mechanisms and how much they impact the influence spreading is unknown. Thirdly, realistic social networks are usually large in size (e.g., with over ten million nodes) and the analysis of these large graphs is very often complicated.

The contributions of this paper are:

- To be best of our knowledge, we are the first to propose mathematical models to predict influence spreading in social networks.
- We show how to use the *local mean field* (LMF) technique to analyze the influence of nodes in large graphs. Using the local mean field, one can concentrate on the correlation structure of local neighborhoods only, so that one can easily derive the statistical properties of the underlying graphs.
- We formally characterize various influence mechanisms and propose a framework to find the final fraction of buyers under a given mechanism for large random networks. Using this framework, we analyze several influence mechanisms and evaluate their performances via simulations.

The outline of our paper is as follows. In Section 2, we present the model of social networks and the problem statement. Then we introduce the concept of local mean field to analyze the networks. We also present several influence models to illustrate how users can affect others in social networks. In section 3, we extend the models for scale-free graphs with high clustering coefficient. In Section 4, we validate our analysis in Section 2 and 3 via simulation and reveal various factors that affect the influence spreading. Related work is given in Section 5 and Section 6 concludes.

2. BASIC MODEL

In this section, we present the models of advertising in social networks. The problem can be informally stated as follows. A company provides free samples to a small fraction of users in a social network so as to entice them to purchase the product. In a social network, users that have bought the product can also influence their neighbors or friends. The issue is, how does this influence spread, and at the equilibrium, what is the fraction of users that will purchase the product? It is important to point out that this outcome depends heavily on how users influence each other. In the following subsections, we will present various influence mechanisms on social networks and derive the expected fraction of users that will eventually purchase the product.

2.1 Modeling Social Networks as Scale-free Random Graphs

For simplicity of presentation, let us first model the underlying social network as an *infinite scale-free* [3] sparse "random" graph $G(V, E)$. In later section, we also extend the models for graphs with high clustering coefficient. A scale-free graph is a graph whose node degree follows a power law distribution. That is, the fraction of nodes that have k neighbors, denoted by $P_0(k)$, is proportional to $k^{-\gamma}$ for large values of k, or

$$P_0(k) \propto k^{-\gamma}, \tag{1}$$

where γ is a positive constant value[1]. Note that for a realistic social network (e.g., Facebook), the number of users is in the order of 10^6 or larger, thus the infinity assumption is justified.

Each user is represented as a node in $G(V, E)$. Each node can influence its neighbors. For example, if node i decides to purchase a product, it may influence its neighbors to purchase the same product. Obviously, one can have different influence models and we will elaborate on them later. In this paper, we focus on the statistical properties of social networks. For example, if each node in this scale-free graph G has a probability ρ of receiving a free sample of a product, then given a particular influence model, we want to derive the probability that a randomly chosen node will eventually purchase the product.

The tight dependency among nodes makes the analysis of the above system difficult. For example, if nodes a and b have a common neighbor, say node c, then the influence between a and b are *coupled*. In general, dependency may occur even if nodes are multiple links away from each other. This type of multi-nodes interaction is generally difficult to solve exactly because of the combinatorics generated by the interactions when summing over all possible influences.

To overcome this problem, we construct a local mean field (LMF) of an *arbitrary* node in G. In essence, LMF is a transformation of G and it allows us to model the correlative structure on local neighborhoods only. More importantly, the LMF provides an asymptotic behavior as the number of nodes of a sparse random graph goes to infinity with a given asymptotic degree distribution $P_0(k)$ [8].

The construction of LMF of G can be described as follows. We randomly choose a node, say $r \in V$, as the starting point of the local mean field. Since r is randomly chosen, according to the property of a scale-free random graph, r has $\deg(r)$ neighbors, say v_1, v_2, ..., $v_{\deg(r)}$, where $\deg(r)$ follows the power law distribution. When we construct the LMF with the starting node r, we model this random scale free graph as a tree rooted at node r and follows the same degree distribution. We refer readers to [2] for similar results. Given this LMF, we calculate the influence spreading on this new structure. We have the following proposition.

Proposition 1. *Let $\mathcal{G}$ be an infinite random graph with asymptotic degree distribution, then for an arbitrarily chosen node r and the corresponding LMF, the local topology of the graph rooted at r can be modeled as a tree with high probability.*

[1]The typically value of γ in the range of $2 < \gamma < 3$.

Remark: The implication of the above proposition is that we can view the scale-free graph as a tree rooted at node r. Node r can be influenced by nodes in sub-trees rooted at v_1 to $v_{\deg(r)}$, but the influence to node r is *independent* between any two sub-trees. Since there exists a recursive tree structure, we can then easily analyze the overall influence by all nodes to the root node r. Figure 1 illustrates the local mean field of a social network as an infinite-depth random tree with node r being the root. It is important to note that since any node in a given social network G can be chosen as the root of the corresponding tree, the performance measure (e.g., average influence by all nodes to the root node) that we will derive can be applied to any node in the original graph G.

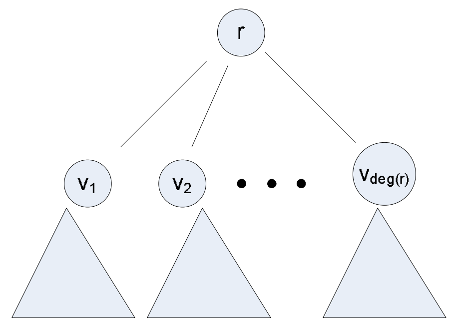


Figure 1: Local topology of node r.

To construct the LMF rooted at node r, we first need to obtain the degree distributions for the root node r and its children nodes. Let $i \in V$ be a node in the scale-free graph G, $\deg(i)$ denote the number of neighbors of i . Then for a tree rooted at r, $\deg(r)$ follows the same power law distribution as Equation (1), or:

$$\text{Prob}[\deg(r) = k] = P_0(k) = Ck^{-\gamma},$$

where C is a constant satisfying $C \sum_{k=1}^{\infty} k^{-\gamma} = 1$. For convenience, let us denote $\zeta(\gamma) = 1/C$. Then we have

$$\text{Prob}[\deg(r) = k] = P_0(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}, \quad k = 1, 2, \ldots \tag{2}$$

We can also derive the degree distribution of any descendant nodes of r. The result is summarized in the following lemma.

Lemma 1. *For an infinite random power law graph, the probability that a descendant node has degree k is:*

$$P_1(k) = \frac{k^{1-\gamma}}{\zeta(\gamma - 1)}, \quad \textit{for } k = 1, 2, \ldots \tag{3}$$

Where $\zeta(x) = \sum_{k=1}^{\infty} k^{-x}$ is the Riemann zeta function.

Proof: A descendant node with degree k is k times as likely to be chosen as one with degree 1, so the distribution of the number of neighbors of a descendant node is for $k \geq 1$

$$P_1(k) = \frac{k \cdot P_0(k)}{\sum_{k=1}^{\infty} k \cdot P_0(k)} = \frac{k^{1-\gamma}}{\zeta(\gamma - 1)}, \qquad k = 1, 2, \ldots \quad \blacksquare$$

Thus, the degree distribution of the descendants of node r all follow a *shifted power-law distribution* $P_1(k)$ of Equation (3). Now the local mean field of G is completely determined. It describes the distribution of the local topology of a randomly chosen node in graph G. It is recursive and free of loops, which makes it convenient to derive statistical properties of the social network. In the following subsections, we will use it to study several influence mechanisms.

2.2 q-Influence Model

Let's say each user is represented by a node in $G(V, E)$. Suppose a company provides free samples as advertisement to $\rho < 1$ fraction of users in this social network. Users receiving the free sample will buy the product by their own will with probability p^+, while users who do not receive the free sample may also buy the product by their own will with probability p^-. We assume $p^+ > p^-$. Users who buy the product can also influence their friends (e.g., neighbors in the social network) to buy the product with probability q. Our goal is to derive the fraction of users in the social network that will eventually purchase the product.

To answer the above question, let us first define the following random variables. Let ϕ_i be the Bernoulli random variable to indicate whether node i decides to purchase the product by his own will (e.g., without the influence of other nodes), then ϕ_i has the parameter μ where

$$\mu = \rho p^+ + (1-\rho)p^-. \quad (4)$$

Let θ_{ij} be the Bernoulli random variable to indicate whether node i can influence his neighbor j to purchase the product. Under the $q-$influence model, it is easy to see that θ_{ij} has parameter q.

Before we derive the fraction of users that will purchase the product in the LMF tree, let us illustrate the intuition on how nodes can influence other nodes via a *deterministic* example. Consider a *finite* tree with a pre-defined root r and all ϕ_i and θ_{ij} for all nodes in the tree are also known, e.g., they are equal to either 0 or 1. Then for node i, if ϕ_i is already 1, i obviously buys the product; if i has a neighbor j such that $\phi_j = \theta_{ji} = 1$, i will also buy the product. If neither of these two conditions hold, i may still buy the product if there is a path $i - i_1 - i_2 \ldots - i_k$ such that $\phi_{i_k} = \theta_{i_k i_{k-1}} = \ldots = \theta_{i_1 i} = 1$. Conversely, if no such path exists and i decides not to purchase, then i will not buy the product. Therefore, to compute the final state of the root node r (e.g., whether node r will purchase the product either due to his own will, or due to the influence of all other nodes in the tree), we can update the states of all other nodes in this tree in a bottom-up manner. That is, we can determine the state ϕ_i of any leaf node i. Given the values of ϕ_i in the leaf nodes, we can determine the state of their parent nodes based on the influence model.

We can now generalize the above intuition to an infinite-depth random tree. Let X indicate whether the root node r finally buys the product, $\text{cld}(a)$ be the set of children of node a, Y_i indicate whether a non-root node i will buy the product only due to the influence of the advertisement and its descendants, then $X = Y_r$. Based on the definition of the q-influence model, a node i does not purchase the product if and only if it does not purchase by its own will, and none of its neighbors who have bought the product can successfully influence it. Thus we have the following relationships:

$$1 - Y_i = (1-\phi_i) \prod_{j \in \text{cld}(i)} (1 - \theta_{ji} Y_j), \quad (5)$$

$$1 - X = (1-\phi_r) \prod_{j \in \text{cld}(r)} (1 - \theta_{jr} Y_j). \quad (6)$$

In effect, Y_i sums up all the influence of all descendant nodes of node i, and X sums up all the influence of the subtrees in $cld(r)$.

We can now consider an infinite tree with the root node at r and apply LMF analysis on Eq. (5)-(6). That is, now consider X, Y_i as Bernoulli random variables with mean $E[X]$, $E[Y_i]$, then we can prove that Equation (5)-(6) have a unique solution, and $E[X]$ is the fraction of buyers in the social network.

Theorem 1. *For the infinite local mean field tree, all $Y_i, i \neq r$ are identically distributed. If Y_i and Y_j are at the same depth, then they are also independent of each other. Moreover, Eq. (5)-(6) have a unique solution.*

Proof: we refer readers to [1]. ∎

By Theorem 1, we can let $Y_j \sim Y$ for all $j \neq r$. To solve Eq. (5)-(6), we take expectation on both sides of the equation. Since ϕ_i, θ_{ij} and Y_j that share the same parent are all independent of each other, we have:

$$1 - E[Y_i] = (1-\mu) E[\prod_{j \in \text{cld}(i)} (1 - \theta_{ji} Y_j)],$$

$$1 - E[X] = (1-\mu) E[\prod_{j \in \text{cld}(r)} (1 - \theta_{jr} Y_j)].$$

To derive the expectation term on the right hand side, note that the influence from children is independent with the parent node's degree, we can condition on the node degree:

$$E[\prod_{j \in \text{cld}(i)} (1 - \theta_{ji} Y_j)] = \sum_{k=0}^{\infty} P_1(k+1) \prod_{j=i_1}^{i_k} E[1 - \theta_{ji} Y_j] = \sum_{k=0}^{\infty} P_1(k+1)(1 - qE[Y])^k, \quad (7)$$

$$E[\prod_{j \in \text{cld}(r)} (1 - \theta_{jr} Y_j)] = \sum_{k=1}^{\infty} P_0(k) \prod_{j=r_1}^{r_k} E[1 - \theta_{jr} Y_j] = \sum_{k=1}^{\infty} P_0(k)(1 - qE[Y])^k. \quad (8)$$

Here i_j is the j^{th} child of node i and we use theorem 1 in equation (7) and (8). $P_1(k)$ is the probability that a descendant node has degree k and $P_0(k)$ is the probability that the root node has degree k. We finally obtain we called the **recursive distributional equation (RDE)** for the q-influence model:

$$1 - E[Y] = (1-\mu) \sum_{k=0}^{\infty} P_1(k+1)(1 - qE[Y])^k, \quad (9)$$

$$1 - E[X] = (1-\mu) \sum_{k=1}^{\infty} P_0(k)(1 - qE[Y])^k. \quad (10)$$

The performance measure, $E[X]$, is the fraction of users that will eventually purchase the product. Lastly, the above equations can be easily solved using standard numeric methods.

2.3 m-threshold Influence Model

In the $m-$threshold influence model, a user will buy the product either by his own will, or when at least m of his friends (or neighbors) have purchased the product. To illustrate, consider a deterministic example on a finite tree in Figure 2. As before, let the random variable $\phi_i = 1$ if node i decides to purchase by its own will and $\phi_i = 0$ otherwise. In this deterministic example, the value of ϕ_i is shown and labeled in the figure. Suppose we set the threshold $m = 2$, then node v_1 will buy the product under the influence of node v_4 and v_5. Also, the root node r will buy the product under

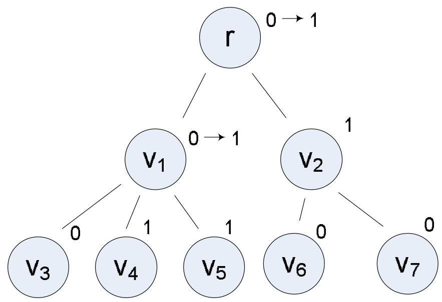


Figure 2: Deterministic example for $m-$threshold model, $m = 2$.

the influence of node v_1 and v_2. In general, to compute the state of the root node under the $m-$threshold influence model, we can apply the same bottom-up updating algorithm as before.

As before, let X indicate whether the root node r finally buys the product, $\text{cld}(a)$ be the set of children of node a, Y_i indicate whether a non-root node i will buy the product only due to the influence of the advertisement and its descendants, and $Y_i \sim Y$ for all $i \neq r$. By the definition of the m-threshold influence model, a node does not purchase the product if and only if it does not purchase by its own will and the total number of its neighbors that have bought the product is less than m. Therefore we have the following relationships:

$$1 - Y_i = (1 - \phi_i) \cdot \mathbf{1}[\sum_{j \in \text{cld}(i)} Y_j < m], \tag{11}$$

$$1 - X = (1 - \phi_r) \cdot \mathbf{1}[\sum_{j \in \text{cld}(r)} Y_j < m]. \tag{12}$$

Here the Bernoulli random variable $\mathbf{1}[\sum_{j \in \text{cld}(i)} Y_j < m]$ indicates whether less than m friends of node i have contributed influence to i. Local mean field method can also be applied to Eq. (11)-(12) so as to compute the state distribution of the randomly chosen root node. Taking expectation on both sides of Eq. (11)-(12), we have:

$$\begin{aligned} 1 - E[Y_i] &= (1 - \mu)\text{Prob}[\sum_{j \in \text{cld}(i)} Y_j < m], \\ 1 - E[X] &= (1 - \mu)\text{Prob}[\sum_{j \in \text{cld}(r)} Y_j < m]. \end{aligned}$$

To derive the probability term on the right side of the above equations, we can condition on the number of children nodes:

$$\text{Prob}[\sum_{j \in \text{cld}(i)} Y_j < m] = \sum_{k=0}^{\infty} P_1(k+1) \sum_{j=0}^{\min\{m-1,k\}} C_k^j E[Y]^j (1 - E[Y])^{k-j},$$

$$\text{Prob}[\sum_{j \in \text{cld}(r)} Y_j < m] = \sum_{k=1}^{\infty} P_0(k) \sum_{j=0}^{\min\{m-1,k\}} C_k^j E[Y]^j (1 - E[Y])^{k-j}.$$

So the final *recursive distributional equation* (RDE) for the $m-$threshold mechanism is:

$$1 - E[Y] = (1 - \mu) \sum_{k=0}^{\infty} \sum_{j=0}^{\min\{m-1,k\}} P_1(k+1) C_k^j E[Y]^j (1 - E[Y])^{k-j}, \tag{13}$$

$$1 - E[X] = (1 - \mu) \sum_{k=1}^{\infty} \sum_{j=0}^{\min\{m-1,k\}} P_0(k) C_k^j E[Y]^j (1 - E[Y])^{k-j}. \tag{14}$$

In other words, $E[X]$ is the fraction of users in the social network that will eventually purchase the product. Again, the above equations can be easily solved using standard numeric methods.

2.4 Majority Rule Influence Model

In the majority rule influence model, a user will buy the product either by his own will, or if over η fraction of his friends have bought the product. Figure 3 shows a deterministic example of a finite tree under the majority rule model. As before, ϕ_i indicates whether node i decides to purchase by its own will. In this deterministic example, the values of ϕ_i are all known and labeled in the figure. If we define the majority as 50%, according to the initial condition, v_1 will be influenced to purchase the product since half of his friends, v_4 and v_5, have bought the product. Also, the root node will be influenced to purchase the product as half of its friends (e.g., v_1) have purchased the product.

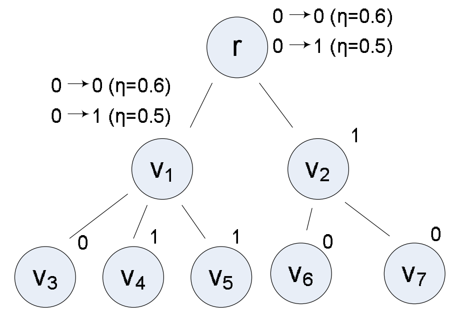


Figure 3: Deterministic example for majority rule model.

According to the definition of the majority rule influence model, a node will not buy the product if and only if it does not purchase by its own will, and the fraction of its neighbors that have bought the product is less than the fraction η. Using the same notation as before, the following equations hold for the majority rule influence model:

$$1 - Y_i = (1 - \phi_i) \cdot \mathbf{1}[\sum_{j \in \text{cld}(i)} Y_j < \eta \deg(i)], \tag{15}$$

$$1 - X = (1 - \phi_r) \cdot \mathbf{1}[\sum_{j \in \text{cld}(r)} Y_j < \eta \deg(r)]. \tag{16}$$

Here the Bernoulli random variable $\mathbf{1}[\sum_{j \in \text{cld}(i)} Y_j < \eta \deg(i)]$ indicates whether the fraction of i's friends that have purchased the product is less than the majority line η. In the example of Figure 3, if we raise the majority value η from 50% to 60%, then v_1 and r will not be influenced anymore. To solve X, we take expectation on both sides of Eq. (15)-(16):

$$\begin{aligned} 1 - E[Y_i] &= (1 - \mu)\text{Prob}[\sum_{j \in \text{cld}(i)} Y_j < \eta \deg(i)], \\ 1 - E[X] &= (1 - \mu)\text{Prob}[\sum_{j \in \text{cld}(r)} Y_j < \eta \deg(r)]. \end{aligned}$$

To derive the probability term on the right side, we condition on the number of children:

$$\text{Prob}[\sum_{j \in \text{cld}(i)} Y_j < \eta \deg(i)] = \sum_{k=0}^{\infty} P_1(k+1) \sum_{j=0}^{\lceil \eta(k+1) \rceil - 1} C_k^j E[Y]^j (1 - E[Y])^{k-j},$$

$$\text{Prob}[\sum_{j\in \text{cld}(r)} Y_j < \eta \deg(r)] = \sum_{k=1}^{\infty} P_0(k) \sum_{j=0}^{\lceil \eta k\rceil -1} C_k^j E[Y]^j (1-E[Y])^{k-j}.$$

The final *recursive distributional equation* (RDE) for the majority rule influence model is:

$$1-E[Y] = (1-\mu)\sum_{k=0}^{\infty}\sum_{j=0}^{\lceil \eta(k+1)\rceil -1} P_1(k+1) C_k^j E[Y]^j (1-E[Y])^{k-j}, \quad (17)$$

$$1-E[X] = (1-\mu)\sum_{k=1}^{\infty}\sum_{j=0}^{\lceil \eta k\rceil -1} P_0(k) C_k^j E[Y]^j (1-E[Y])^{k-j}. \quad (18)$$

Again, $E[X]$ is the fraction of users that will eventually purchase the product.

3. SCALE-FREE GRAPH WITH HIGH CLUSTERING COEFFICIENT

In social networks, two common friends of an user are usually friends of each other. This implies that graphs of social network usually exhibit high *clustering coefficient*. In this paper, we use the definition which was first proposed by Watts and Strogatz in [12] to characterize the clustering coefficient c.

Definition 1. $c = \frac{1}{|V|}\sum_{v\in V}\frac{t_v}{k_v(k_v-1)/2}$.

where k_v is the degree of node v, and t_v is the number of edges in the neighborhood of node v.

Obviously, the scale-free graph with high clustering coefficient cannot be modeled as a tree. But we can still employ the LMF model to analyze the influence spreading in the graph. The only thing we need to do is to modify the degree distribution of the descendant nodes in Eq. (3). In essence, for scale-free graphs with high clustering coefficient, the degree distribution of the descendant nodes does not follow shifted power law any more, and we can compute it as following. Consider a descendant node b whose parent is node a, we have:

$$\begin{aligned}
&P(deg(b) = k|cld(a) = m) \\
&= \sum_{j=0}^{m-1} p(b \overset{j}{\sim} a)\cdot p(deg(b) = k|cld(a) = m, b \overset{j}{\sim} a) \\
&= \sum_{j=0}^{m-1} \binom{m-1}{j} c^j (1-c)^{(m-1-j)} \cdot \frac{p_0(k)\binom{k}{j+1}}{\sum_{k=1}^{\infty} p_0(k)\binom{k}{j+1}} \\
&= \sum_{j=0}^{m-1} \binom{m-1}{j} c^j (1-c)^{(m-1-j)} \cdot \frac{k(k-1)...(k-j)k^{-\gamma}}{\sum_{k=1}^{\infty} k(k-1)...(k-j)k^{-\gamma}}
\end{aligned}$$

where $b \overset{j}{\sim} a$ means b connects j edges with the children of node a.

Now we can derive the degree distribution of any descendant node. The result is summarized in the following lemma.

Lemma 2. *For an infinite random scale free graph with clustering coefficient c, the probability that a descendant node has degree k is:*

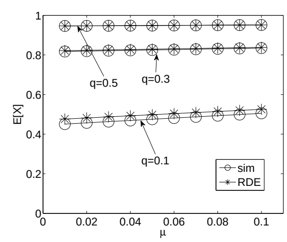
$$P_1(k) = \sum_{m=1}^{\infty} \frac{(m+1)^{-\gamma}}{\zeta(\gamma)} \sum_{j=0}^{m-1} \binom{m-1}{j} c^j (1-c)^{(m-1-j)} \cdot \frac{k(k-1)\cdots(k-j)k^{-\gamma}}{\sum_{k=1}^{\infty} k(k-1)\cdots(k-j)k^{-\gamma}} \quad k = 1, 2, \ldots (19)$$

Now the local mean field of scale-free graph with high clustering coefficient is completely determined. One can use the LMF model defined in subsection 2.2, 2.3 and 2.4 to analyze the influence spreading in scale-free infinite, high clustering graph by using equation (19) to substitute $P_1(k)$.

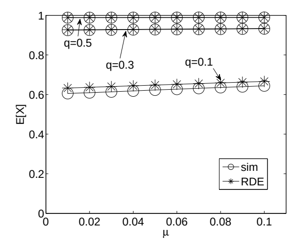
4. PERFORMANCE EVALUATION

In this section, we present the performance evaluation of the extended models presented in Section 3. The local mean field model assumes an infinite scale-free random graph without self-loops and duplicate edges. To evaluate its effectiveness, we first generate a graph with sufficiently large number of nodes and high clustering coefficient by using the GLP model presented in [4]. In our simulation, we generate two types of random scale free graphs both with clustering coefficient c of around 0.5. The first type has a minimum degree of 3 and the second type has a minimum degree of 5. We apply influence models on both of these two graphs. Since ρ, the probability of randomly chosen node will receive the free sample from a company, is usually small, therefore, in our study, we assume that μ is chosen from 0 to 0.1.

4.1 q-Influence Model



(a) minimum degree 3



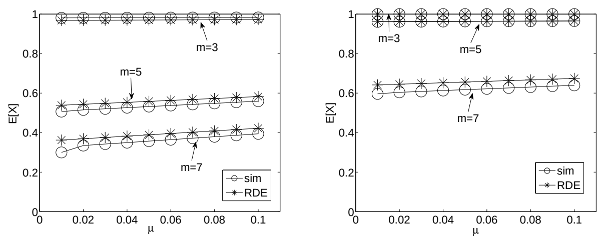
(b) minimum degree 5

Figure 4: Impact of the $q-$influence model

Recall that in the q-influence model, a user that has bought the product will influence its neighbors with probability q. Figure 4 shows the simulation as well as theoretical results by the recursive distributional equation (RDE) Eq. (9)-(10). The horizontal axis μ (see Eq. (4)) is the initial fraction of users that purchase the product, and the vertical axis $E[X]$ is the final fraction of users that will eventually buy the product. In each figure, there are three simulation curves and three theoretical curves corresponding to different q values from 0.1 to 0.5. First of all, we can see that the theoretic results fit well with the simulation results. We also observe that $E[X]$ is much higher in Figure 4b because the underlying graph has a higher average degree. This indicates that high degree networks, which are more connected, are easier for the influence to diffuse. Moreover, we can see that even an infinitesimal advertisement ($\mu = 0.01$) can still lead to high $E[X]$.

4.2 m-threshold Influence Model

In the m-threshold influence model, a user will be influenced by his neighbors if at least m of them have bought the product. Figure 5 shows the simulation as well as theoretical results by the recursive distributional equation (RDE) Eq. (13)-(14). The horizontal axis μ (or from Eq. (4)) is the initial fraction of users that purchase the product, and the vertical axis $E[X]$ is the final fraction of users that will eventually buy the product. In each figure, there are three simulation curves and three theoretical curves corresponding to different m values from 3 to 7. We can see that the theoretic results fit well with the simulation results. For fixed m, the curve is higher in Figure 5b where the power law graph has a higher average degree. Similar with the q influence model, we also observe that small μ can lead to high $E[X]$. Lastly, to compare the m-threshold and the q-influence models in the power law graph with minimum degree five, $m = 7$ curve is closest to the $q = 0.1$ curve.

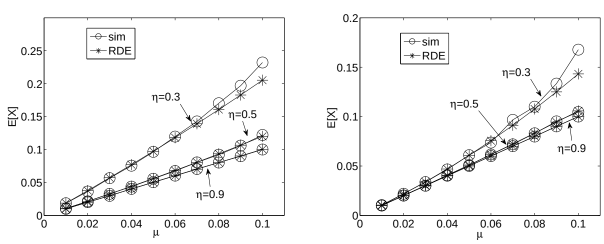


(a) minimum degree 3 (b) minimum degree 5

Figure 5: Impact of the $m-$threshold influence model

4.3 Majority Rule Influence Model

In the majority rule influence model, a user will be influenced by its neighbors if over η fraction of them have bought the product. Figure 6 shows the simulation as well as theoretical results by the recursive distributional equation (RDE) Eq. (17)-(18). The horizontal axis μ (or from Eq. (4)) is the initial fraction of users that purchase the product, and the vertical axis $E[X]$ is the final fraction of users that will eventually buy the product. In each figure, there are three simulation curves and three theoretical curves corresponding to majority factor η of 30%, 50%, 90%. We can see that the theoretic results fit well with the simulation results. In both graphs, when $\eta = 0.9$, $E[X]$ almost equals μ, which means that the mutual influence among users is very weak. Again, we observe low μ can lead to high $E[X]$ when $\eta = 0.3$.



(a) minimum degree 3 (b) minimum degree 5

Figure 6: Impact of the majority rule influence model

In summary, the recursive distributional equation (RDE) framework is a good approximation to analyze influence spreading in social networks. In general, high degree networks are more sensitive to the parameter change of the underlying influence mechanisms.

5. RELATED WORK

Several papers are related to the problem we discuss here. Much work focuses on the epidemic spreading [6,10] via the Susceptible-Infective-Susceptible (SIS) or the Susceptible-Infective-Removed (SIR) model. The underlying graph models are mostly infinite scale-free graphs [3,5] or Erdös-Rényi graph. Results on the speed the epidemic spreads and dies out are obtained. In this work, we consider more general influence models and focus on the final fraction of users that are influenced. Some researchers discuss specific influence models via algorithmic perspective and design approximation algorithms [9], heuristic algorithms [11] for restricted graphs, or prove NP-hardness results of choosing the most influential nodes [7] in social networks. There are a body of literatures on topology generation; see, e.g. [4]. Lastly, in the motivating work of [8], the authors provide local mean field analysis on infinite random graphs and applied this theory for security investment games.

6. CONCLUSION

We propose a general analytical framework to model various influence mechanisms on large scale random networks. We first discuss the probabilistic model (q-influence model), then we present the deterministic threshold models such as the $m-$threshold influence model and the majority rule influence model. Based on these influence models, we compute the expected fraction of users who will eventually purchase the product by applying the local mean field analysis. This fraction is very important for product advertisement because it reveals the maximum number of users who will buy the product or the maximum profit that the company can obtain. It also gives us the insights on how to control or cascade the effect of word-of-mouth so as to maximize the revenue of the company. We validate our theoretic analysis by carrying out extensive simulations on random scale free graphs with power law degree distribution and high clustering coefficient. We show that our models are very accurate when compare with simulations. We observe that even with a small initial investment of free samples (e.g., small value of ρ), one can still include large number of users to purchase the product. Lastly, our framework provides an important building block to design and analyze different product advertisement strategies in social networks.

7. REFERENCES

[1] D. J. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equations. *Ann. Appl. Probab*, 2005.

[2] D. J. Aldous and M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. 2004.

[3] A.-L. Barabasi and R. Albert. Emergence of Scaling in Random Networks. *Science*, 1999.

[4] T. Bu and D. Towsley. On distinguishing between internet power law topology generators. *INFOCOM*, 2002.

[5] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Physics Review Letters*, 2003.

[6] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. *INFOCOM 2005.*, pages 1455–1466 vol. 2, 2005.

[7] D. Kempe, J. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *KDD '03*. ACM, 2003.

[8] M. Lelarge and J. Bolot. Network Externalities and the Deployment of Security Features and Protocols in the Internet. In *ACM Sigmetrics*, 2008.

[9] E. Mossel and S. Roch. On the submodularity of influence in social networks. In *STOC '07*. ACM, 2007.

[10] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physics Review Letters*, 2001.

[11] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *KDD '02*. ACM, 2002.

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