A Mutual Independent Cascade Model for Customer Behavior Propagation

Ali Nazemian
School of Electrical & Computer Engineering
University of Tehran
Tehran, Iran
nazemian@ut.ac.ir

Fattaneh Taghiyareh
School of Electrical & Computer Engineering
University of Tehran
Tehran, Iran
ftaghiyar@ut.ac.ir

Abstract—**Model of spreading behaviors, influences, new trends and innovations through social networks has been studied in a number of domains. These may include the diffusion of medical and technological innovations, the sudden and widespread adoption of strategies in game-theoretic settings, and the effects of word of mouth in the promotion of new products. One of the most important facts that is neglected in previous spread models is "considering cascading negative opinions". This important fact shows that negative opinions may originate and propagate in populations as much as positive opinions and even they are stronger and more dominant. In this paper we propose a new model of influence cascade called Independent Cascade with Positive and Negative WOM (ICPN). ICPN models some important facts that people may encounter in a social environment. These facts include negativity bias, the asymmetric behavior of negative and positive WOM, and different types of consumer complaints behaviors. Moreover, the influence maximization problem is formulated in this model and also, we show that ICPN maintains submodularity in this problem. This fact allows a simple greedy approximation algorithm for maximizing the positive influence within a ratio of $(1 - \frac{1}{e})$ approximation.**

Keywords-component; independent cascade model; influence maximization; customer behavior propagation; negative word of mouth; positive word of mouth ; consumer complaining behavior

I. INTRODUCTION

Online social networks are increasingly being recognized as an important source of information influencing the adoption and use of products and services. Viral marketing - the tactic of creating a process where interested people can market to each other- is therefore emerging as an important means to spread the word and stimulate the trial, adoption, and use of products and services [1]. In fact viral marketing is based on using consumer-to-consumer (or peer-to-peer) communications -as opposed to company-to-consumer communications- to disseminate information about a product or service, thereby leading to more rapid and cost effective adoption by the market [2, 3]. Viral marketing is more powerful than third-party advertising because it conveys an implied endorsement from a friend[1].

A social network – the graph of relationships and interactions within a group of individuals- plays a fundamental role as a medium for the spread of information, ideas and influence among its members [4]. An idea or innovation appears - for example the use of telecommunications service providers or the adoption of a new cell phone in the market, or the rise of a political movement in an unstable society- can either die out quickly or make significant inroads into the social network. If we want to understand the extent to which such ideas are adopted, it can be important to understand how the dynamics of adoption are likely to unfold within the underlying social network: the extent to which people are likely to be affected by decisions

of their friends and colleagues, or the extent to which "word-of-mouth" effects will take hold.

In [5, 6], motivated by applications of marketing, Domingos and Richardson posed a fundamental algorithmic problem for mentioned systems. However, after wards Kempe et al. [4] formulated the problem of maximizing the influence in social networks as a discrete optimization problem and showed that this problem is NP-hard. Motivated by their research, many studies have considered the influence maximization problem in social networks from the algorithmic point of view [7-19.] Most of these studies are based on independent cascade model and linear threshold model, which is defined by Kempe et al. in [4] and their extensions.

According to [18] influence maximization problem for a social network modeled as a graph, starting from a small initial set of vertices in that graph (called seeds or initial adopters). A model of cascading behavior specifies how influence is propagated from these initial seeds to their neighbors and neighbors of neighbors, and so on, until the process ends and a portion of the network is influenced and thereby activated. Thus the influence maximization problem is to find an optimal seed set of size at most k such that the expected number of vertices that are activated from these seed sets maximized at the end of the process; That is referred to as its influence spread is the largest.

Almost all of the previous works ignore an important fact that we often experience in the real world. This fact is that not only positive opinions on products/services that someone receives may propagate through the network, but also negative opinions can establish and propagate the same. Even negative opinions are often more dominant and stronger in affecting the decisions of other people [20-23]. Another important fact that is ignored by previous studies is consumer-complaining behavior, which is directly related to customer's behavior and influence that is propagated in population. For example if you hear from one of your friends that a specific cell phone has a problem with taking photos, in spite of the total satisfaction of your friend on this cell phone, you might decide not to buy that. Even you may tell your colleague not to buy this cell phone, neither.

Propagation of negative opinions is studied extensively in marketing and social science literature, but this subject has been rarely touched in computer science literature from the algorithmic point of view. To the best of our knowledge, the only related paper that discusses diffusion of negative opinions is [18]. However the presented model by Wei Chen et al. ignores the literature related to the negative and positive word of mouth. The only hypothesis that is considered in their model is negativity bias. They ignore some important facts such as customer complaining behavior. Their proposed model has lots of simplifications and is far from the spreading of influence in the real world.

The remaining of this paper is arranged as follows: In section II we review the literature related to the negative and positive word of mouth and also consumer complaining behavior and also some of the related works are presented in this section. Then in the next section, we propose a model of influence cascade in social networks called ICPN with consideration of both positive and negative word of mouth. In section IV we focus on influence maximization and the power of our model for solving this problem is considered. For influence maximization, since it is directly related to the revenue generated by the viral marketing effort, we focus on maximizing the expect number of positive nodes in the network after the cascade ends. The last section concludes this paper and points out some future research for this study.

II. BACKGROUND AND RELATED WORK

A. Influence Maximization

Suppose that we are given a social network with the estimates of mutual influence between individuals in the network. Also suppose that we want to push a new product in the market for the promotion. The problem of influence maximization is considered as follows: given such a network with influence estimates, how the set of initial users should be selected so that they eventually influence the largest number of users on the social network? This problem has received much attention in the algorithms and theoretical computer science communities in the last decade [46].

The first studies that consider the propagation of influence and the problem of identification of influential users from an algorithmic perspective are [5, 6].This problem is modeled by means of Markov random fields and heuristics are given for choosing the target users. In particular, the marketing objective function to maximize is the global expected lift in profit; that is intuitively, the difference between the expected profit obtained by employing a marketing strategy and the expected profit obtained without using any marketing strategy. A Markov random field is an undirected graphical model representing the joint distribution over a set of random variables whose nodes are variables, and edges represent dependencies between them. It is adopted in the context of influence propagation by modeling only the final state of the network at convergence as one large global set of interdependent random variables.

Kempe et al. in [4] tackle roughly the same problem as a discrete optimization problem. They obtain provable approximation guarantees under various propagation models studied in mathematical sociology, which is described next.

According to [4], a social network can be represented as a directed graph $G = (V, E)$ in which every node is either in one of two states: active or inactive. Here, "active" may correspond to a user buying a product or getting infected and "inactive" corresponds to others. In progressive mode, it is assumed once a node becomes active, it remains active. But in non-progressive mode, a node can switch between active and inactive states during different time steps. Influence is assumed to propagate from nodes to their neighbors according to a propagation model and a node's tendency to become active, increases monotonically as more of its neighbors become active.

In the *independent cascade* (IC) model each active node u has one chance for influencing inactive node v. This node can succeed with probability $p_{u,v}$, the probability which u can influences v. In the *linear threshold* (LT) model, each node u is influenced by each neighbor v according to a weight $b_{v,u}$, such that the sum of incoming weights of in-neighbors of u is not larger than 1. Each node u chooses a threshold θ_u uniformly at random from the interval $[0, 1]$. If at time step t the total weight of the active in-neighbors of u exceeds the threshold θ_u, then u will become active at the next time step $(t + 1)$. In both of the mentioned models, the process repeats until no new node remains active.

For any propagation model, the expected influence spread of a seed set S is the expected number of nodes that finally gets activated by the initial set of activated nodes S. This expected number is denoted by $\sigma_m(S)$, where m stands for the underlying propagation model and S is the set of initially activated nodes. In general the influence maximization problem is defined as follows. Given a directed and edge-weighted social graph $G = (V, E)$, a propagation model m, and a number $k \leq |V|$, find a set $S \subseteq V$, $|S| = k$, such that $\sigma_m(S)$ is maximized.

Under both the IC and LT propagation models, this problem is shown to be NP-hard [4]. However, for both of these propagation models that described, the expected influence spread function $\sigma(\cdot)$ is monotone and submodular. Monotonicity says when the set of activated nodes grows, the likelihood of a node getting activated should not decrease. More precisely, a function f from sets to real numbers is monotone if $f(S) \leq f(T)$ whenever $S \subseteq T$. In addition a function f is submodular if $f(S \cup \{u\}) - f(S) \geq f(T \cup \{u\}) - f(T)$ whenever $\subseteq T$. Submodularity intuitively says the probability of activating some inactive node u, does not increase if more nodes have already attempted to activate u and u is hence more "marketing-saturated". It is also called the law of "diminishing returns".

By the power of submodularity and monotonicity a simple greedy heuristic for influence maximization under IC and LT provides an approximation guarantee by the factor of $\left(1 - \frac{1}{e}\right)$ [44].

B. *Consumer Complaining Behavior*

The consumer complaint behavior, CCB in short, is an area of research, which deals with the identification and analysis of all the aspects involved in the consumer reaction to a product or a service failure and the consequent perceived dissatisfaction [24].

According to [24-29], three distinct dimensions of CCB have been verified: 1) voice complaints (complaining directly to sellers or service providers); 2) private complaints –word of mouth-(refers to the act of telling at least one friend, acquaintance or family about satisfactory or dissatisfactory about product/service experiences [30]); and 3) third party complaints (complaining to independent organizations such as the media, consumer groups, or legal agencies in order to seek redress, e.g. Better Business Bureau).

Most researchers assume some level of dissatisfaction as a starting point for any type of complaint behavior (voice, private, or third party) [30-36]. In addition, some researchers have noted the incidence of complaining even among satisfied consumers. (e.g. [37]) Kowalski argued that some consumers complain not out of dissatisfaction but in an effort to gain fraudulently from retailers or manufacturers [38]. Some satisfied consumers might even complain about minor service problems or product defects due to organizational commitment, loyalty or even to reinforce their earlier buying decisions [39]. Thus we can say that satisfied consumers might complain about certain attribute even though overall satisfaction is relatively high. In this case, the one that heard a complaint does not know that the overall satisfaction of the complainer is high. Therefore it can affect people's decisions in a negative way. In fact neither all complainers are dissatisfied and nor all dissatisfied consumers complain [39].

When dissatisfaction and other factors lead to consumer complaint behavior, the choice of private voice or third party complaint still exists. In fact, some research has found that consumers engage in multiple complaint responses (a "supplemental effort") rather than choosing a single complaint option (a "Substitution effort") [29-31, 39-43]. It means that complaint behaviors such as exit, negative WOM and seller complaints are separate processes that are influenced by different variables or in different ways by the same variables.

In [39] Halstead proposed a framework for consumers complaining behavior and specified four categories of consumers based on satisfaction and complaining behaviors. This framework is shown in Fig. 1.

Based on the framework presented in Fig.1, Halstead showed that negative WOM for all of these four groups can exist. It means that a consumer complaint directly to sellers does not indicate a lack of complaining to friends and family. According to [39] the extent of negative WOM behavior and the probability for a person to participate in negative WOM is as below:

$$Category\ 1 < Category\ 2 < Category\ 3 < Category\ 4$$

Thus the extent of negative WOM for a dissatisfied complainer consumer is the highest one. Also the extent of NWOM for satisfied non-complainer ones is the least one. We will present our model based on this fact in the next section.

	Complainers	Non-Complainers
Satisfied	Category 2	Category 1
Dissatisfied	Category 4	Category 3

Figure 1. consumer satisfaction level and voice complaining framework. [32]

C. Related Work

There are two types of previous studies related to our work. Some of previous researches focused on designing an efficient algorithm for influence maximization under the IC, LT or their extensions, while other ones target the model of spreading the influence in social networks and tried to present a new model that could better fit the spreading influence in the real world.

From the algorithm perspective, which we considered as a first category, a simple greedy approximation was the first algorithm introduced for influence maximization [4, 7]. However, this algorithm is very slow in practice and not scalable to the size of network. In [9], Leskovec et al. propose lazy-forward optimization that significantly speeds up the greedy algorithm, but it still cannot scale well to large networks with hundreds of thousands of nodes and edges. A number of heuristic algorithms are also proposed [8, 12, 11, 10, 56] for the independent cascade model. SPM/SP1M of [8] is based on shortest-path computation, and SPIN of [10] is based on Shapley value computation. Both SPM/SP1M and SPIN have been shown to be not scalable enough [11, 14]. The simulated anneal approach is proposed in [56], which provides reasonable influence coverage and running time.

One of the best heuristic algorithms so far is believed to be the PMIA algorithm proposed by Chen et al. in [11], which provides matching influence spread while running at three orders of magnitude faster than the optimized greedy algorithm. The PMIA computes local tree structures for each node, this procedure speed up the computation of influence spread. However, the disadvantages of PMIA are that it incurs high memory cost for local data structures and it needs to estimate influence spread of many nodes one by one. The IRIE algorithm presented in [17],tries to overcome both of the shortcomings of the PMIA that mentioned and thus can perform even much faster than it. IRIE algorithm integrates influence ranking with influence estimation together with the greedy approach, overcoming the general issue of ignoring overlapping influence coverage suffered by all pure ranking methods.

In the second category primary works belong to Domingos and Richardson [5, 6] and Kempe et al. [4] as we mentioned. However there are many studies that extend IC or LT to better fit to the real world phenomena [15, 18, 47-55]. These studies try to extend previous works from four different perspectives:1) History sensitiveness; 2) Multiple cascade and competitiveness; 3) Time dependent, and 4) Negative influence. Some of these researches are presented in the following.

From the subject of "history sensitiveness" Zhanget al. [50] proposed the History Sensitive Cascade Model (HSCM), a model of information cascade through a network over time. The authors considered the "activation" problem of finding the probability of receiving information by particular node from some nodes that initially informed. In this study it was also proven that selecting a set of k nodes with greatest expected influence is NP-hard, and results

from submodular functions are used to provide a greedy approximation algorithm with a $1-1/e-\epsilon$ lower bound, where ϵ depends polynomially on the precision of the solution to the "activation" problem.

Many research papers have considered the subject of "multiple cascade and competitiveness" [47, 48, 15, 51, 52, 53]. Borodin et al. in [47] introduced a threshold model that could work in competitive situations, which aims to maximize the spread of our technology in the presence of one or more competitors. They proved the NP-hardness of this problem. In addition, they showed the $(1-1/e)$ approximation guarantee for a greedy algorithm to solve this problem. Also Bharathi et al. considered the game of innovation diffusion with multiple competing innovations such as when multiple companies' market competing products using viral marketing. They give an $(1-1/e)$ approximation algorithm for computing the best response to an opponent's strategy and proved that the "price of competition" of this game is 2 at most [15]. Moreover in the study presented by Pathak et al. [48], a generalized version of the linear threshold model that is capable of handling multiple cascades in a non-progressive mode is presented. The corresponding stochastic process is shown to be a rapidly mixing Markov chain. Also the StochColor algorithm is provided for discovering the most likely states of the cascades' spread in a given graph.

For the perspective of "time dependency" some studies can be found [49, 54, and 55]. Saito et al. in [49] presented a model of Continuous-Time Independent Cascade (CTIC), which is the extension of the IC model to allow continuous time delays. They modeled a time-delay by an exponential distribution as a natural extension. Moreover they addressed the problem of estimating the parameters for CTIC from the observed data by rigorously formulating the likelihood of obtaining these data and maximizing the likelihood iteratively with respect to the parameters (time-delay and diffusion).

As far as we know, in the subject of "negative influence" the only presented spread model is ICN that is introduced by Chen et al. [18]. This model is an extension of the independent cascade model that incorporates the emergence and propagation of negative opinions. ICN has an explicit parameter called quality factor to model the natural behavior of people turning negative to a product due to product defects. This model ignores some basic features acknowledged in the social psychology literature. The only characteristic that is considered in ICN is negativity bias (negative opinions usually dominate over positive opinions). This paper proved that ICN maintains some nice properties such as submodularity, which allows a greedy approximation algorithm for maximizing the positive influence within a ratio of $(1-1/e)$. In ICN some important facts about the real life spreading model are ignored. Some of these facts are: consumers switching causes; different types of consumers complaining behavior; and supplemental consumers complaining effort. The only parameter that is considered in their model was quality factor, which has an absolute value. ICN has too much simplification and does not fit well in real life

situations. We try to target these simplifications in our study. As a result, we introduce a model that can better fit into the spreading influence in real life.

III. INDEPENDENT CASCADE MODEL WITH POSITIVE AND NEGATIVE WOM

As we mentioned a social network is modeled as a directed graph $G = (V, E)$, where V is the set of nodes representing individuals and E is the set of directed edges. E represents relationships between individuals while in ICPN this relationship is associated with a propagation probability, which is formalized by function $p: E \rightarrow [0,1]$. We refer to the tuple $G = (V, E, P)$ as an influence graph. ($P = \{p, p_1, p_2, p_3, p_4\}$) For a node $v \in V$, let $N^{in}(v)$ and $N^{out}(v)$ denote v's in-neighbors and out-neighbors respectively.

The dynamic of our presented model is as follows. Each node has three states:neutral, positive and negative. Discrete time steps $t = 0, 1, 2$ ...are used to model dynamic changes in network. A node v is called activated at time t if it is positive or negative at time t and neutral at time $t - 1$ (if $t > 0$). This model has a parameter q called *quality ratio*, which represents the probability that a node remains positive after it is activated by its neighbors. Also ICPN has a parameter C called *complaint factor* that is used for specifying the complaint status of each node. Initially at time $t = 0$, all nodes in a pre-determined seed set $S \subseteq V$ are activated. Each node $v \in S$ becomes positivewith probability q, and becomes negative with probability of $1 - q$. Also at same time step, the complaint status of each node in seed set S determines.

In time $t > 0$ for any neutral node v, let $A_t(v) \subseteq N^{in}(v)$ be the set of in-neighbors of v that were activated at time $t - 1$. Every node $u \in A_t(v)$ based on its category tries to activate v positively with an independent probability of $p^*(u, v). p(u, v)$ and tries to activate it negatively with an independent probability of $(1 - p^*(u, v)). p(u, v)$ which p^*represents the probability related to each category. If one of them becomes successful, v is activated at step t. If v is activated positively it remains positive with probability of q and its status will change to negative with probability of $1 - q$. To determine which node activates v, we randomly permute all nodes in $A_t(v)$, and let each node in $A_t(v)$ try to activate v following the permutation order until we find the first node u that activates v successfully. Once v is activated and fixed its state (positive or negative), it will not change its state any more. The activation process stops when there is no new activated node in a time step. The dynamic process of ICPN is given in Fig. 2 and 3.

A. Conceptual justification of the ICPN model

The ICPN model reflects several phenomena of negative and positive influence that match our daily experiences as well as the studies in both social psychology and marketing. First, negative opinions originate from imperfect or defected products/services. In addition there are two major reasons for a customer

to switch to another service provider or product producer: 1) Dissatisfaction; and 2) looking for better deals [30-34]. In ICPN model when a node v is positively activated by a positive node u it means that v is positively influenced by u and subsequently buys a product/service. However, due to either defects of the product/service or/and better deals of other producers/service providers, v may dislike the product/service and generate negative opinions about it. The *quality ratio q* reflects the quality of product/service rather than other similar products/services and represents satisfaction that is related to product quality and better deals since it is in form of ratio that can represent products comparison. Thus, the *quality ratio* is the property of product/service, not the property of the network. Therefore it is reasonable to use same quality ratio across the network. Typically before a product is put into the market, the producer performs quality control by testing and/or focus group studies, and comparing the product/service with similar ones from other competitors based on product's features. Thus it is reasonable to assume that an estimate of q is available. In ICPN model we consider the *quality ratio* as a standardized value between 0 and 1.

Second, negative and positive influences are asymmetric and negative influence is more dominant, which is reflected in the ICPN model from two aspects. The first aspect is that, when a node v is negatively activated by using the product / service it will stay negative even if it later sees other neighbors turning to positive. The second aspect is when node v is negatively activated without experiencing product/service, it will also affects its neighbors negatively. This is the manifestation of negativity dominance in the domain of contagion.

Third, we considered 4 categories of consumers based on CCB in determining the probability of influence activation.

Forth, according to the literature neither all complains come from dissatisfaction, nor all dissatisfied consumers complain. This fact is reflected in the ICPN model from two aspects: First, when node v is activated positively by experiencing product/service, it can activate its neighbors negatively with a little probability. Second aspect is that when node v is activated negatively by experiencing product/service, it can activate its neighbors positively with a little probability.

Fifth, for indicating the complaining status of each person we used *complaining factor*. This factor represents the percent of consumers that had complaining behavior to the total number of consumers. Therefore this value can be obtained easily.

Finally, we use positive influence spread as our objective since it is directly related to the expected revenue that the seller would gain from the viral marketing effort.

```
Algorithm 1 ICPN Model: G= (V, E, p, p₁, p₂,
p₃, p₄) seed set S, quality ratio q,
complaint factor C.
/* initially for all v ∈ V, state (v) =
neutral */
i= 0, S₀ = S
forall v ∈ S₀ do
   state (v)= positive with probability q
otherwise
   state (v)= negative
   setCategory(v)
end for
while Sᵢ ≠ ∅ do
   Sᵢ₊₁ = ∅
   for all v ∈ ∪ᵤ∈ₛ Nᵒᵘᵗ(u) and state (v)
=neutral do
      Order set Sᵢ ∩ Nⁱⁿ(v) uniformly at
random into sequence ρ
      for all u ∈ ρ , according to the order
of ρ do
         setStatus(u,v)
      end for
   end for
   i=i+1
end while

procedure setCategory(v)
   if state (v)= positive then
      category (v)=2 with probability C
otherwise
      category (v)=1
   else
      category (v)=4 with probability C
otherwise
      category (v)=3
   end if
end procedure
```

Figure 1.   The ICPN model pseudo code

B. ICPN in details

As mentioned before, each customer should be categorized in 4 groups based on customer satisfaction/dissatisfaction and complaint status. These customers are persons that adopted a product/ service.

In ICPN model, when a person buys a product or uses a service (called adopted product/service), at the beginning of the cascade process the state of that person is positive. After the product/service is used by the mentioned person, assumes in time step $t = t'$, his state will change. The new state of this person is defined by the *quality ratio* that was introduced. In this phase the person will choose positive state with the probability of q and negative state will be chosen with the probability of $1 - q$. For every person (node in influence graph), one of the following conditions can be identified:

```
Algorithm 1 ICPN Model: G= (V, E, p, p₁, p₂,
p₃, p₄) seed set S, quality ratio q,
complaint factor C.

procedure setStatus(u,v)

   if state (u)= positive then
      if category(u)=1 then
         state (v) =positive and
Sᵢ₊₁ = Sᵢ₊₁ ∪ {v} with probability of
p₁(u,v).p(u,v)otherwise
         state (v)= negative and
Sᵢ₊₁ = Sᵢ₊₁ ∪ {v} with probability of (1-
p₁(u,v)).p(u,v)
      end if
      if category(u) =2 then
         state (v) =positive and
Sᵢ₊₁ = Sᵢ₊₁ ∪ {v} with probability of
p₂(u,v).p(u,v)otherwise
         state (v)= negative and
Sᵢ₊₁ = Sᵢ₊₁ ∪ {v} with probability of (1-
p₂(u,v)).p(u,v)
      end if
   else
      if category (u)= 3 then
         state (v) =positive and
Sᵢ₊₁ = Sᵢ₊₁ ∪ {v} with probability of
p₃(u,v).p(u,v)otherwise
         state (v)= negative and
Sᵢ₊₁ = Sᵢ₊₁ ∪ {v} with probability of (1-
p₃(u,v)).p(u,v)
      end if
      if category (u)= 4 then
         state (v) =positive and
Sᵢ₊₁ = Sᵢ₊₁ ∪ {v} with probability of
p₄(u,v).p(u,v)otherwise
         state (v)= negative and
Sᵢ₊₁ = Sᵢ₊₁ ∪ {v} with probability of (1-
p₄(u,v)).p(u,v)
      end if
   end if
   if state(v)= positive then
      state (v)= negative with
probability of 1-q
   end if

end procedure
```

Figure 2.   The ICPN model pseudo code

- If in time step $t = t'$ the state of a person was positive (he was satisfied by the product/service), then that person will participate in negative WOM (propagate negative opinion about product/service) with the probability of $1 - p_1$ if he was not complainer, and with the probability of $1 - p_2$ if he was a complainer. We can justify this condition by the fact that the one can be satisfy about a product but also participate in negative WOM. In this situation the person will participate in positive WOM about the product based on his complaining status with probability of p_1 and p_2. Now if this person propagates his negative opinion about the product/service, the state of his neighbors in the social network becomes negative and he will remain negative until the cascade ends.

- If in time step $t = t'$ the state of the person becomes negative, then according to his negative/positive WOM. For a non-complainer person the probability of participating in positive WOM will be p_3 and for a complainer one it will be p_4. Also for a non-complainer person the probability of participating in negative WOM will be $1 - p_3$ and for a complainer one it will be $1 - p_4$.

- When somebody receives a positive idea about product/service he will adopt it and the *quality ratio* will specify his future status.

According to Halstead framework and the extent of NWOM for each category, for the probability of p_1, p_2, p_3, p_4 in ICPN, this result could be derived:
$$1 - p_1 < 1 - p_2 < 1 - p_3 < 1 - p_4$$

, which could be substituted with
$$0 < p_4 < p_3 < p_2 < p_1 < 1$$

So this relation between p_1, p_2, p_3, p_4 should be considered in ICPN model.

IV. APPROXIMATION GUARANTEE IN ICPN

The *positive influence spread* of a seed set S in influence graph G with *quality ratio q* is the expected number of positive nodes activated in the influence graph, and is denoted as $\sigma_G(S, q)$. Given an influence graph $G = (V, E, P)$, a target seed set size k, and a *quality ratio q*, the *influence maximization* problem is to find a seed set S^* of cardinality k such that S^* has the largest positive influence spread in G.

Given an influence graph $G = (V, E, P)$, seed set S and *quality ratio q*, let $pap_G(v, S, q)$ reflects "*positive activation probability*", the probability that node v is positive after the influence cascade from S ends. By the linearity of expectation, it is clear that
$$\sigma_G(S, q) = \sum_{v \in V} pap_G(v, S, q)$$

Definition 1: $d_G(S, v)$ denotes the distance from S to v in graph G. That is the length of the shortest path from any node in S to v in graph G.

If there is no path from any node in S to v then we can say that $d_G(S, v) = +\infty$. As a convention, we considered $q^{+\infty} = 0$, for all $0 \leq q \leq 1$ (even when $q = 1$).

Definition 2: let $a_G(S, i)$ denote the number of nodes that are i steps away from set S in graph G, thus we can say
$$a_G(S, i) = |\{v | d_G(S, v) = i\}|$$

The following lemma shows a basic property of the ICPN model that leads to many results.

LEMMA 1: for influence graph $G = (V, E, P)$ suppose that $P(e) = 1$ for all $e \in E$. Then for all $v \in V$, we have,
$$pap_G(v, S, q) = q^{d_G(S,v)+1}$$

complaining status, he will participate in

and
$$\sigma_G(S, q) = \sum_{i=0}^{n-1} a_G(S, i) q^{i+1}$$

Proof: It is sufficient to show that for $v \in G$ with $d_G(S, v) = i$, the equality $pap_G(v, S, q) = q^{i+1}$ holds. This statement can be proved by induction. For the base case $(i = 0)$ it is obvious. Because of every node in the seed set S is activated and has probability of q for becoming positive.

Now consider a node v with $d_G(S, v) = i \geq 1$. Let $U = N_{in}(v, i - 1)$ be the set of incoming neighbors of v that are at distance of $i - 1$ from S. Clearly, all nodes U are activated at time $i - 1$ because of the assumption $P(e) = 1$, and v will be activated at time i. In ICPN v will be activated by one of the nodes in U which is chosen randomly. By induction, and because of the $P(e) = 1$ assumption, every node in U becomes positive with probability q^i. Therefore, v will be positively activated with a probability of q^{i+1} no matter which node in U activates v at time i. Thus lemma 1 follows by taking summation over all nodes.

For any influence graph $G = (V, E, P)$, after all random events on all edges based on their propagation probability are determined, a sub graph $\acute{G} = (\acute{V}, \acute{E}, \acute{P})$ is obtained, where $V = V$, $\acute{E} \subseteq E$, and $\acute{P}(e) = 1$ for all $e \in \acute{E}$ (there is no difference that the probability for each edge follows $p_1, p_2, \ldots$ in ICPN, we considered the influence edges that are effective in activating). Therefore $\acute{G}$ is obtained with probability $Pr_G(\acute{G}) = \prod_{e \in \acute{E}} P(e) . \prod_{\acute{e} \in E \setminus \acute{E}} (1 - P(\acute{e}))$. Let Ω_G denote the set of all such subgraphs $\acute{G}$. Thus an edge e is activated if e is selected in the random subgraph $\acute{G}$.

An alternative view of the ICPN model is that first we select edges to obtain $\acute{G}$, and then the influence is propagated on $\acute{G}$. In the graph $\acute{G}$ when multiple neighbors of a node v try to activate v in the same step, there is no need to follow the random permutation order on these neighbors because the first neighbor that is selected will always activate v. we use this view of ICPN model for specifying some important results. This alternative view is referred to *edge activation view*. It is clear that in this view we can ignore *the complaint factor* and all parameters that are related to the edge probability.

LEMMA 2: Given an influence graph $G = (V, E, P)$, a seed set S and *quality ratio q*, we have
$$\sigma_G(S, q) = E_{\acute{G} \leftarrow \Omega_G}[\sigma_{\acute{G}}(S, q)]$$
$$= \sum_{\acute{G} \in \Omega_G} Pr_G(\acute{G}) \, \sigma_{\acute{G}}(S, q)$$
$$= \sum_{\acute{G} \in \Omega_G} Pr_G(\acute{G}) \sum_{i=0}^{n-1} a_{\acute{G}}(S, i) q^{i+1}$$

COROLLARY 1: For any influence graph $G = (V, E, P)$, when fixing a seed set S, function $\sigma_G(S, q)$ on q is *monotonically increasing* and continuous.

A set function f on the vertices of graph $G = (V, E, P)$ is a function $f: 2^V \to \mathbb{R}$. Set function f is *monotone* if

$$f(S) \leq f(T) \text{ for all } S \subseteq T,$$

and it is submodular if

$$f(S \cup \{u\}) - f(S) \geq f(T \cup \{u\}) - f(T) \text{ for all } S \subseteq T \text{ and } u \in V \backslash T.$$

THEOREM 1: for any influence graph $G = (V, E, P)$, when fixing a *quality ratio* q, set function $\sigma_G(S, q)$ on S is *monotone*, *submodular* and $\sigma_G(\emptyset, q) = 0$.

Proof. Notice that

$$\sigma_G(S, q) = \sum_{\acute{G} \in \Omega_G} Pr_G(\acute{G}) \sum_{v \in V} q^{d_{\acute{G}}(S,v)+1}$$

We define $Q_v(S) = q^{d_{\acute{G}}(S,v)+1}$. It is sufficient to show that $Q_v(S)$ is *monotone* and *submodular*. Clearly, $Q_v(S)$ is *monotone* because adding extra elements to the seed set S can only decrease the quantity $d_{\acute{G}}(S, v)$. Thus it is enough to show that the function is also *submodular*.

Let $S \subseteq T \subseteq V$ and $u \in V \backslash T$. Clearly $d_{\acute{G}}(S, v) \geq d_{\acute{G}}(T, v)$. We can determine three different cases.

Case 1: If $d_{\acute{G}}(u, v) \geq d_{\acute{G}}(S, v)$, we have

$$Q_v(S \cup \{u\}) - Q_v(S) = Q_v(T \cup \{u\}) - Q_v(T) = 0.$$

Case 2: If $d_{\acute{G}}(u, v) \leq d_{\acute{G}}(T, v)$, as $Q_v(.)$ is monotonically increasing, we have

$$Q_v(S \cup \{u\}) - Q_v(S) = Q_v(T \cup \{u\}) - Q_v(S) \geq Q_v(T \cup \{u\}) - Q_v(T).$$

Case 3: The only remaining case is $d_{\acute{G}}(T, v) < d_{\acute{G}}(u, v) < d_{\acute{G}}(S, v)$. In this case we have

$$Q_v(S \cup \{u\}) - Q_v(S) > 0 = Q_v(T \cup \{u\}) - Q_v(T)$$ Therefore $Q_v(.)$ is *monotone* and *submodular*.

According to theorem 1, we can apply the mentioned results in [44] to obtain a greedy approximation algorithm that achieves $(1 - \frac{1}{e})$ approximation ratio for *influence maximization* in ICPN model; which shows the same efficiency as an original IC model presented by Kempe et al. Fig.4 shows a simple greedy algorithm from [18]. This algorithm iteratively selects a new seed u that maximizes the incremental change of f into the seed set S until k seeds are selected. In our case f can be replaced by $\sigma_G(S, q)$.

```
Algorithm 2: Greedy approximation for
influence maximization
Initialize S = ∅
for i=1 to k do
    select u = argmax_{u∈V\S}(f(S ∪ {u}) − f(S))
    S = S ∪ {u}
end for
output S
```

Figure 3. The ICPN model pseudo code

V. DISCUSSION

Previous models of spreading influence in social networks have done too many simplifications because of the NP-hardness of influence maximization. These simplifications result to models that cannot fit in real world situations. Hence, they are not usable in simulating the spread of influence is social networks. In this paper, we overcome this simplification by adding some feature of real world situations to the basic model of independent cascade. These features are discussed with details in "Conceptual justification of the ICPN model" section. The summary of these features is shown in table 1. By considering these features, ICPN can fit in real world spreading of the influence better than previous models.

Despite the NP-Hardness of the influence maximization problem, ICPN can be usable for real life simulations since it provides the approximation guarantee of $(1 - \frac{1}{e})$ for this problem.

VI. CONCLUSIONS

Models of cascading influence have an essential role in viral marketing. Some important problems such as influence maximization are defined based on these models. In this paper, we present a new influence cascade model based on positive and negative WOM, called ICPN. In ICPN we modeled several phenomena that might be experienced in real life as well as the studies in social psychology and marketing. Some of these phenomena are: negativity bias and dominance, consumers switching causes, different types of consumers complaining behavior, supplemental consumers complaining effort and etc.

TABLE I. MAPPING BETWEEN ICPN AND REAL WORLD INFLUENCE SPREADING

ICPN	Real World influence spreading
Quality Ratio	Customer switching cause
Dominancy of directly activated negative node/ indirectly activated negative node as influencer	Asymmetry of negative and positive influence and dominancy of NWOM
4 categories of consumers with different probability	Consumer Complaining Behavior
Activating negatively by positive node and vice versa	Supplemental consumers complaining effort
Complaining factor	Complaining status of each person

We showed that our model of influence cascade maintained some good properties such as submodularity, which is effective in finding the most

influencing individuals in the population. Submodularity allows a simple greedy approximation algorithm to maximize the positive influence within a ratio of $1 - 1/e$ approximation.

For future research we can extend our model of influence cascade with considering some situations that may encounter in the real world; for example extending ICPN for allowing each node to have a different quality ratio to model the situation where different individuals have different tendency for product/service; also situation that the process is non-progressive and the one can switch between different products/services continuously is appropriate. As the other work that can be done to continue this research, we can focus on finding the best range for each of mentioned probabilities (p, p_1, p_2, p_3, p_4) to improve the accuracy of the presented model.

REFERENCES

[1] M. R. Subramani and B. Rajagopalan, "Knowledge-Sharing and Influence in Online Social Networks via Viral Marketing", *Communications of the ACM*, vol. 46, no. 12, pp. 300-307, 2003.

[2] A. De Bruyn and G. L. Lilien, "A multi-stage model of word-of-mouth influence through viral marketing," *International Journal of Research in Marketing*, vol. 25, no. 3, pp. 151-163, Sep. 2008.

[3] Krishnamurthy, Sandeep,"Understanding online message dissemination: Analyzing 'Send a message to a friend' data."*First Monday*, 6(5), 2001.

[4] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03*, p. 137, 2003.

[5] P. Domingos, M. Richardson. "Mining the Network Value of Customers."*Seventh International Conference on Knowledge Discovery and Data Mining, 2001.*

[6] M. Richardson, P. Domingos. "Mining Knowledge-Sharing Sites for Viral Marketing."*Eighth Intl. Conf. on Knowledge Discovery and Data Mining, 2002.*

[7] D. Kempe, J. M. Kleinberg, and ´ E. Tardos. "Influential nodes in a diffusion model for social networks."*In Proceedings of the 32nd International Conference on Automata, Languages, and Programming*, pages 1127–1138, 2005.

[8] M. Kimura and K. Saito. "Tractable models for information diffusion in social networks."*In Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 259–271, 2006.

[9] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance. "Cost-effective outbreak detection in networks."*In Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining,* pages 420–429, 2007.

[10] R. Narayanam, "A shapley value-based approach to discover influential nodes in social networks," *IEEE Transactions on Automation*, pp. 1-18, 2010.

[11] W. Chen, C. Wang, and Y. Wang. "Scalable influence maximization for prevalent viral marketing in large scale social networks."*In Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2010.

[12] W. Chen, Y. Wang, and S. Yang. "Efficient influence maximization in social networks."*In Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2009.

[13] Y. Wang, G. Cong, G. Song, and K. Xie. "Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. " *In Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2010.

[14] W. Chen, Y. Yuan, and L. Zhang. "Scalable influence maximization in social networks under the linear threshold model."*In Proceedings of the 10th IEEE International Conference on Data Mining*, 2010.

[15] S. Bharathi , D. Kempe, M. Salek, "Competitive influence maximization in social networks," *Internet and Network Economics*, 2007.

[16] H. Ma, H. Yang, M. R. Lyu, and I. King, "Mining social networks using heat diffusion processes for marketing candidates selection," *Proceeding of the 17th ACM conference on Information and knowledge mining - CIKM '08*, p. 233, 2008.

[17] K. Jung, W. Heo, and W. Chen, "IRIE: A Scalable Influence Maximization Algorithm for Independent Cascade Model and Its Extensions.",*In Proceedings of CoRR*. 2011.

[18] Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincón, Xiaorui Sun, Yajun Wang, Wei Wei, Yifei Yuan: "Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate." *In proceeding of SDM* 2011: 379-390.

[19] A. Shapira, "A note on maximizing the spread of influence in social networks," *Internet and Network Economics*, 2007.

[20] G. Peeters and J. Czapinski. "Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects." *European Review of Social Psychology*, 1:33–60, 1990.

[21] S. E. Taylor. "Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis."*Psychological Bulletin*, 110(1):67–85, 1991.

[22] R. F. Baumeister, E. Bratslavsky, and C. Finkenauer. Bad is stronger than good. Review of General Psychology, 5(4):323–370, 2001.

[23] P. Rozin and E. B. Royzman. "Negativity bias, negativity dominance, and contagion. " *Personality and Social Psychology Review*, 5(4):296–320, 2001.

[24] Singh, Jagdip, "Consumer Complaint Intentions and Behaviors: Definitional and Taxonomical Issues" *Journal of Marketing,* 52, (January), 93-107, 1988..

[25] Blodgett, Granbois, "Toward an Integrated Conceptual Model of Consumer Complaining Behavior," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior,* 5, 93-103, 1992.

[26] Blodgett, Tax, "The Effects of Distributive and Interactional Justice on Complainants' Repatronage Intention and Negative Word-of-Mouth Intentions," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior,* 6,100-110, 1993.

[27] Blodgett, "The Effects of Perceived Justice on Complainants' Repatronage Intentions and Negative Word-of-Mouth Behavior," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior,* 7, 1-14, 1994.

[28] Singh, Jagdip, "Voice, Exit and Negative Word-of-Mouth Behaviors: An Investigation across Three Service Categories," *Journal of the Academy of Marketing Science,* 18 (1), 1-15, 1990.

[29] Singh, Jagdip, "When Consumers Complain: Path Analysis of the Key Antecedents of Consumer Complaint Responses," *Journal of the Academy of Marketing Science,* 24 (fall), 350-365, 1996.

[30] M. L. Richins, "Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study," *Journal of Marketing*, vol. 47, no. 1, p. 68, 1983.

[31] Richins, Marsha, "A Multivariate Analysis of Responses to Dissatisfaction, "*Journal of the Academy of Marketing Science*, 15 (Fall) 24-31, 1987.

[32] R. Laczniak and T. DeCarlo, "Consumers' responses to negative word-of-mouth communication: An attribution theory perspective," *Journal of Consumer*, no. I, pp. 57-73, 2001.

[33] F. V. Wangenheim, "Postswitching Negative Word of Mouth," *Journal of Service Research*, vol. 8, no. 1, pp. 67-78, Aug. 2005.

[34] I. Wetzer and M. Zeelenberg, "'Never Eat In That Restaurant, I Did !': Exploring Why People Engage In Negative Word-

Of-Mouth Communication," *Psychology and Marketing*, vol. 24, no. 2007, pp. 661-680, 2007.

[35] G. T. Lau and S. Ng, "Individual and Situational Factors Influencing Negative Word-of-Mouth Behaviour," *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, vol. 18, no. 3, pp. 163-178, Apr. 2009.

[36] J. Palmer, V. Eidson, and C. Haliemun, "Predictors of Positive and Negative Word of Mouth of University Students: Strategic Implications for Institutions of Higher Education," *ijbssnet.com*, vol. 2, no. 7, pp. 59-63, 2011.

[37] J. Jacoby, J. Jaccard, "The Sources Meaning, and Validity of Consumer Complaint Behavior: A Psychological Analysis," *Journal of Retailing,* 57 (fall), 4-24, 1981.

[38] R. M. Kowalski, "Complaints and Complaining Function, Antecedents and Consequences." *Psychological Bulletin,* 119 (2), 179-196, 1996.

[39] D. Halstead, "Negative word of mouth: substitute for or supplement to consumer complaints?," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 2002.

[40] R. Day, "Modeling Choices Among Alternative Responses to Dissatisfaction," *in Advances in Consumer Research,* 11, 496-499, 1984.

[41] C. M. Voorhees, "A Voice From the Silent Masses: An Exploratory and Comparative Analysis of Noncomplainers," *Journal of the Academy of Marketing Science*, vol. 34, no. 4, pp. 514-527, Oct. 2006.

[42] R. Ladhari, "The Effect of Consumption Emotions on Satisfaction and Word-of-Mouth," *Psychology*, vol. 24, no. 2007, pp. 1085-1108, 2007.

[43] I. Phau and M. Baird, "Complainers versus non-complainers retaliatory responses towards service dissatisfactions," *Marketing Intelligence & Planning*, vol. 26, no. 6, pp. 587-604, 2008.

[44] G. Nemhauser, L. Wolsey, and M. Fisher. "An analysis of the approximations for maximizing submodular set functions."*Mathematical Programming*, 14:265–294, 1978.

[45] S. Jurvetson, "What exactly is viral marketing?," Red Herring Communications, pp. 110–111, 2000.

[46] A. Goyal and F. Bonchi, "Approximation analysis of influence spread in social networks," *arXiv preprint arXiv: ...*, 2010.

[47] A.Borodin, Y. Filmus, J. Oren, "Threshold models for competitive influence in social networks," *Proceedings of the 6th international conference on Internet and network economics*, pp. 539-550, 2010.

[48] N. Pathak, A. Banerjee, and J. Srivastava, "A Generalized Linear Threshold Model for Multiple Cascades," *2010 IEEE International Conference on Data Mining*, pp. 965–970, Dec. 2010.

[49] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Learning continuous-time information diffusion model for social behavioral data analysis," *Advances in Machine Learning*, pp. 1–16, 2009.

[50] Y. Zhang, M. Tsikhanovich, and G. Smilyanov, "History Sensitive Cascade Model," *International Journal of Agent Technologies and Systems (IJATS)*, vol. 3, no. 2, pp. 1–14, 2011.

[51] T. Carnes, C. Nagarajan, S. M. Wild, and A. V. Zuylen, "Maximizing Influence in a Competitive Social Network : A Follower ' s Perspective," *Proceedings of the ninth international conference on Electronic commerce*, pp. 351–360, 2007.

[52] Q. Yan, S. Guo, and D. Yang, "Influence Maximizing and Local Influenced Community Detection Based on Multiple Spread Model," *Proceedings of the 7th international conference on Advanced Data Mining and Applications*, pp. 82–95, 2011.

[53] S. Datta, A. Majumder, and N. Shrivastava, "Viral Marketing for Multiple Products," *2010 IEEE 10th International Conference on Data Mining (ICDM)*, pp. 118–127, 2010.

[54] W. Chen, W. Lu, and N. Zhang, "Time-Critical Influence Maximization in Social Networks with Time-Delayed Diffusion Process," *arXiv preprint arXiv:1204.3074*, no. Aaai, pp. 1–23, 2012.

[55] M. Rodriguez and B. Schölkopf, "Influence Maximization in Continuous Time Diffusion Networks," *arXiv preprint arXiv:1205.1682*, 2012.

[56] Q. Jiang, G. Song, G. Cong, Y. Wang, W. Si, and K. Xie, "Simulated Annealing Based Influence Maximization in Social Networks," *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pp. 127–132, 2011.

Ali Nazemian is currently a M.Sc. student of Information Technology Engineering at the School of Electrical and Computer Engineering of the University of Tehran, Tehran, Iran. He received his B.Sc. degree from the Shiraz University of Technology, Shiraz, Iran, in 2008. His research interests include Social Networks, Computer Algorithms, Customer Relationship Management and Data Mining.

Fattaneh Taghiyareh is assistant Professor of Computer Engineering, Software and Information Technology, at the School of Electrical and Computer Engineering, University of Tehran, where she has served since 2001. She received her Ph.D. degree in Computer Engineering from the Tokyo Institute of Technology in 2000, and her M.Sc. and B.Sc. degree from Sharif University of Technology in 1994, and 1990, respectively. Currently she is serving both Software engineering and Information Technology Department as a member and she is supervising the Multi-Agent System Lab founded in 2005 as well as eLearning Lab founded in 2012 by her. Her research interests are in collaborative environments for eLearning and Multi-Agent Systems. Her current research involves the intersection of Web-Services, Personalization, and adaptive LMSs as well as applying ontology to semantic web. She was ex-manager of the Department of Information Technology at the University of Tehran and former Director of Technology Incubator, University of Tehran. She is a member of the Editorial Board of International Journal of Information & Communication Technology. Moreover she is supervising IT Foundation lab which has been founded by her in 2006. She is the author of more than 80 publications including Conference and Journal Papers.