Incentive Compatible Influence Maximization in Social Networks

Y. Narahari [*]
Indian Institute of Science
Bangalore, India
hari@csa.iisc.ernet.in

Mayur Mohite
Indian Institute of Science
Bangalore, India
mayur@csa.iisc.ernet.in

ABSTRACT

Information diffusion and influence maximization in social networks are well studied problems and various models and algorithms have been proposed. The main assumption in these studies is that the influence probabilities are known to the social planner. The influence probabilities, however can vary significantly with the type of the information and the time at which the information is propagating. The most accurate sources to obtain influence probabilities are the agents in the social network. In this work, we formulate game theoretic models of the information diffusion process so as to elicit influence probabilities truthfully from the agents. For these models we design several mechanisms to truthfully extract the influence probabilities from the users. We consider two different game theoretic models namely (1) influencer model and (2) influencer-influencee model. In the context of the influencer model, we design a Vickrey-Clarke-Groves based mechanism. In the influencer-influencee model, we design a scoring rule based direct mechanism. We analyze the incentive compatibility of all these mechanisms.

Categories and Subject Descriptors

H.4 [**Algorithms and Theory**]: Social Networks, Scoring Rules, Mechanism Design

General Terms

Algorithms

Keywords

Social Networks, Game Theory, Information Diffusion

1. INTRODUCTION

A social network is a graph in which a person is represented as a node and there is an edge between two people if they are associated with each other. Examples of online social networks include orkut, facebook, twitter etc. Social

---

[*]Fellow IEEE

networks are widespread and also, they are the most effective medium to propagate information and to market and advertise products in a short time span.

Consider a situation in which a company has designed a new gaming console which it wants to market on a social network. The company can select a small set of users to whom it will give the product for free. If these users like the product, then they will recommend it to their friends. These friends may get influenced by the users and will perhaps buy the product with some probability. Among these friends, whoever gets influenced will in turn recommend the product to her friends and so on. This is known as the information diffusion process. Here the choice of an initial set of users is critical because they will decide the expected number of users that would get influenced. The problem now is, given a social network graph and influence probabilities on each edge, how do we select a small set of initial users so as to maximize the product sale. This is the influence maximization problem in social networks. An online social network is an effective medium for launching such a marketing campaign because it has much information about the users as well as the relationship graph of users. In this paper we address the influence maximization problem in an incomplete information setting in which, influence probabilities are the private information. We do this by formulating a game theoretic model of information diffusion process using which we are able to find a small set of influential nodes, by making the nodes in the social network reveal the influence probabilities truthfully.

The problem of influence maximization in a social network has been studied extensively. Various algorithms have been proposed to find the set of highly influential nodes in a social network efficiently [8, 11]. All these algorithms make an important assumption that the influence probabilities in the relationship graph are available to the algorithm. However, there do not exist any robust techniques to extract these influence probabilities. Indeed the users of the social network are the most reliable sources that can provide the influence probabilities on the edges that are present in their neighborhood. In this work we formulate the problem of extracting the influence probabilities from the user as a mechanism design problem.

1.1 Motivation

The current models for the information diffusion process assume that the social planner knows the influence probabilities accurately whatever the information that is getting propagated on the social network. However, in practice this need not be the case.

The influence probabilities for product p1

The influence probabilities for product p2
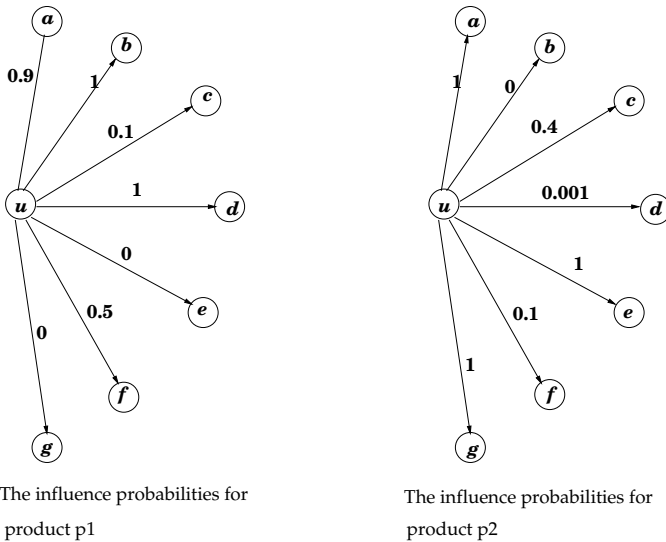
Figure 1: The variation of influence probabilities with product

Consider an example, say a seller wants to sell a newly released book by some author and another seller wants to sell a tennis racket. Both the sellers decide to use viral marketing on a popular online social network to market these products. To do this the seller will have to assume one model, say the independent cascade model, find the values of influence probabilities on each edge and find the influential set of nodes to initiate the cascade process.

But consider a user u and her set of friends S on the social network. The influence probability of u on each member of set S for a product say book need not be the same for another product say tennis racket. This is perhaps because u is a novice tennis player but is good at English literature and her friends know this, and this particular information is not available on this social networking site. Hence her friends will be influenced more by her recommendation of a novel to them than a tennis racket. This is illustrated by figure1

Thus, the influence probabilities can vary drastically with the product being marketed, the time at which the recommendation is made, and perhaps some other factors. Building robust and accurate models for estimating the influence probabilities from the currently available data in the online social networks will be a difficult task. On the other hand, to get the values of the influence probabilities accurately, the best sources are the users of the social network themselves. However, the users need not reveal the true influence probabilities to the seller as they might be better off by lying. For example, a user might pose as having a heavy influence on her friends and get to be a part of the initially active set.

The success of a viral marketing strategy is critically dependent on the initially chosen target set. This in turn depends on the accuracy of the influence probabilities.

1.2 Relevant Work

Kempe, Kleinberg, and Tardos in [8] considered the algorithmic problem of influence maximization proposed by Domingos and Richardson in [6]. In this paper they proved that this problem is NP-hard even for simple models of infor-

mation diffusion and for some of the more complex models, it is not even constant factor approximable. They gave a constant factor approximation algorithm for the basic independent cascade model by proving the sub-modularity of the influence function. But, the greedy algorithm they propose assumes that the influence probabilities are available to the algorithm.

Ramasuri and Narahari in [11] propose a heuristic based algorithm for the influence maximization problem. Given an integer k, they use Shapley value to choose k nodes from the social network that have highest influence. In their approach they consider the influence function as the characteristic function of the coalitional form game. This approach works even if the influence function is not sub-modular. They also show experimentally, that this algorithm works efficiently on real world social networks. But, their algorithm also assumes that the influence probabilities are available to the algorithm.

Alon, Fischer, Procaccia, and Tennenholtz in [1] proposed a game theoretic model for truthfully choosing the agents that maximize the sum of indegrees in a directed graph. In the model they propose, they consider the outdegree of an agent as private information. The objective of each agent is to be among the set of nodes chosen by the algorithm. This objective is in some sense orthogonal to that of the social planner. In [1], they propose various deterministic and randomized strategy-proof algorithms to achieve the objective of maximizing the sum of indegrees. In our problem, for the case of 0-1 cascade process and the influencer model, the objective of the social planner is to choose a set of agents that have maximum reachability. Whereas, the objective of each agent is to maximize the number of neighbors that she is able to activate in the 0-1 cascade process.

A mechanism design based framework to extract the information from the agents have been proposed for ranking systems [2], and query incentive networks [5].

Dixit and Narahari in [5] proposed quality conscious query incentive networks in which the nodes, along with the answer are aware of the quality of the answer. They proposed a game theoretic model of query incentive networks in which the quality of answer is the private information. They also designed a scoring rule based mechanism to truthfully extract the quality of the answers from the agents along with the actual answer. In our work, we design the influencer-influencee model which is similar to the game theoretic model presented in [5] for the quality conscious query incentive networks. In [5] the rewards to the agents depends on the truthfulness of the quality of answers they report. In our problem the payments to the agents depends on the truthfulness of the influence probabilities they report.

The work that is closest to our work is that by Goyal, Bonchi, and Lakshmanan in [7]. In [7], they use a machine learning based approach to build the models for predicting the influence probabilities in social networks. Intuitively, the approach they consider is that if a person x takes a total of n actions and out of that if say $m \leq n$ actions were performed by its neighbor y before x, then there is a probability of $\frac{m}{n}$ that person x will be influenced by y in future. Here the "action" is the act of joining a community or group in a social network which does not involve any effort or monetary transfer. They validate the models they build on a real world data set.

To the best our knowledge, the model presented in this pa-

per is the first one which captures the strategic behavior of agents in the information diffusion process. Using this model, our aim is to elicit the true influence probabilities from the agents in order to accurately compute the set of highly influential nodes.

1.3 Contributions and Outline

In this work, we design mechanisms to extract the influence probabilities from the users of the social network. We do this by designing game theoretic models of information diffusion. Given the model of utilities and payments we ask each user to reveal the influence probabilities. If the mechanism is incentive compatible then every user will reveal the influence probabilities truthfully.

The Influencer based Model

First, we develop a game theoretic model for information diffusion process in which we ask only the influencer to reveal the influence probability on an edge. A detailed description of this model is given in Section 2. For this model we obtain following result:

- We design a VCG based payment scheme[9, 10] for the influencer based model for the ideal influence maximizing algorithm. We show that, without using money or any payment scheme the ideal influence maximizing algorithm is manipulable. We show this by constructing an example in which an agent is able to manipulate the algorithm. We can construct similar examples for other current algorithms for influence maximization including the greedy algorithm [8]. This payment scheme is presented in section 2.1.

The Influencer-Influencee Model

Next, we develop a more general game theoretic model in which given an edge in the social network, we ask the influencer as well as the influencee to reveal the influence probability on the edge. This model is more realistic since, both the persons involved in the connection will have information about the influence probability. This model is presented section 3.1. For this model we obtain following result:

- For the influencer-influencee model, we design a direct payment scheme in which we use scoring rules to design the payment scheme. We show that, it is a Nash equilibrium to report true influence probability in this mechanism. We also design the reverse weighted scoring rule derived from the weighted scoring rule [5] which has several desirable properties which, the standard scoring rules like the quadratic and spherical scoring rules do not have. This payment scheme is presented in section 3.2.

In the appendix, we provide preliminaries in which we describe the scoring rules, the independent cascade model of information diffusion process and define the influence maximization problem formally.

2. THE INFLUENCER MODEL

We propose a game theoretic framework for the information diffusion process in order to truthfully elicit the influence probabilities from the users. In this model we ask only the influencer to report the probability values. The model is as follows:

- The social planner has the entire graph structure of the social network. The players in the game $V = \{1, 2, 3..., n\}$ are the users of the social network.

- Let $N(i)$ be the set of neighbors of an agent i that is $N(i) = \{v \in V | (i, v) \in E\}$. Then, the agent i has the influence probability vector $\theta_i \in [0, 1]^{|E|}$ as her private information. The jth component of θ_i will give the influence probability of i on node j. Thus for all the non-neighbors of node i, the influence probability will be zero and this is known to the social planner as the social planner has the structure of the graph. However, the social planner does not know anything about the influence probability of node i on its neighbors. Before starting the marketing of the product, the social planner asks each agent to report her influence vector. The agents here can lie about their influence on its neighbors.

- Given the reported influence probabilities, the social planner now computes the target set using some influence maximization algorithm. Let this target set be A. The social planner now starts her advertising campaign from this chosen set of nodes. That is, this chosen set of nodes would get the product for free and the cascade starts with them.

- Let $\theta \in [0, 1]^{|E|}$ be the true influence probability vector, representing the influence probability on each edge of the graph. Then the utility of a player when the social planner chooses a target set A is nothing but the expected number of neighbors activated by that player. Also, without involving payments, the valuation function for each agent is equal to its utility that is,

$$u_i(\theta, A) = v_i(\theta, A)$$

Thus utility is proportional to the expected number of neighbors an agent is able to activate, given the target set. The exact formula for $v_i(\theta, A)$ is given in lemma 1.

Here the valuation function represents the preferences of the agents over the target set chosen by the social planner.

2.1 A VCG Based Mechanism

Consider the exact influence maximization algorithm which optimizes the influence function. It has been proven in [8] that this problem is NP-hard. Given our game theoretic model and the algorithm to choose the target set is the exact influence maximizing algorithm, then we can easily come up with an example in which the agents can lie about their preferences and still be better off. One such example is given in the next subsection. We can make this algorithm incentive compatible introducing appropriate incentives to the agents. Now the utility of the agents with payments or discount will be:

$$u_i(\theta, A) = v_i(A) + t_i(v_1, v_2, ..., v_n)$$

where t_i is the discount offered to agent i by the social planner. We will use Vickrey-Clarke-Groves payments (see for example [9]) to make this mechanism incentive compatible, but to use it we need to prove that the social choice function is allocatively efficient. In our framework the social choice function is nothing but the algorithm being used to choose the target set.

We will first prove the following useful lemma which will immediately imply that the exact influence maximization algorithm is allocatively efficient. To prove the lemma, we will first describe an equivalent view of the independent cascade model given by Kempe, Kleinberg and Tardos in [8].

The influence probability of node i on j denoted by θ_{ij} gives the probability that node i will activate j given that node j will be inactive at the instant when node i becomes active. This event can be viewed as the flip of a biased coin. In the process, say we flip the coins on all the edges before the start of the cascade process and check the result of the coin flip only when a node becomes active and its neighboring node is inactive. This change will not affect the final result and it is equivalent to the original cascade process. We call the edges on which the coin flip resulted in heads as live edges and the remaining edges as blocked. Given this equivalent view, if we fix the outcomes of all coin flips and initially active set of nodes A, then we will get a graph in which some edges are live and some are blocked depending on the outcome. Clearly in this graph, if we run the cascade process, then the number of nodes that are active will be the number of nodes that are reachable from set A on a path that consists of only live edges.

Thus we will consider a sample space S in which each sample point corresponds to one possible outcome of all the coin flips. If X denotes one such fixed outcome of coin flips, we define $\sigma_X(A)$ to be the number of active nodes at the end of the cascade process for the fixed outcome X and target set A. Then $\sigma(A)$ is given by:

$$\sigma(A) = \sum_{X \in S} P[X] \sigma_X(A)$$

Given this formula for $\sigma(A)$ we can now prove the following lemma.

LEMMA 1. *Given a target set A, then*

$$\sigma(A) = \sum_{i=1}^{n} v_i(\theta, A) + |A|$$

where v_i is the valuation function of agent i which is equal to the expected number of neighbors activated by that agent.

PROOF. Fix one sample point X from the sample space S of all possible coin flips on the edges. Consider some arbitrary node i in the graph. Let $v_{i_X}(A)$ be the number of neighbors activated by node i for outcome X. More concretely, we define $d(A, v)$ where $A \subseteq V$ and $v \in V$ as the shortest path distance between A and a node v. Thus, $d(A, v) = 0$ if $v \in A$. Let, $R_{Av} = \{u \in V | (u, v) \in E_X \wedge d(A, u) \text{ is finite} \wedge d(A, u) + 1 = d(A, v)\}$ where, E_X is the edge set that is active for the outcome X. R_{Av} is the set of nodes that are lying on the shortest path from set A to node v. Thus, we define
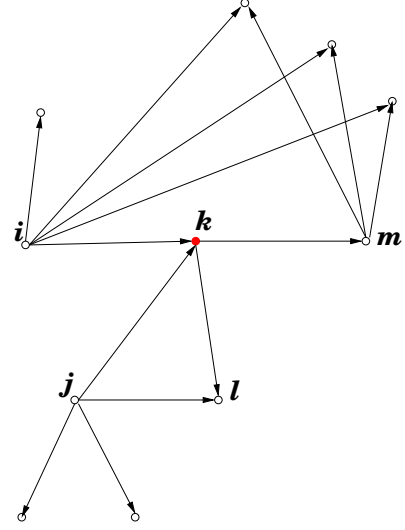
$$v_{i_X}(A) = |\{v \in N(i) | i \in R_{Av} \wedge i \succeq j \, \forall j \in R_{Av}\}|$$

where, the ordering $\succeq$ is lexicographic. Thus we are breaking the ties in favor of the node with the highest lexicographic order. Also lexicographic ordering ensures that a node is activated deterministically by exactly one node for a fixed outcome X.

Then we have,

$$\sigma_X(A) = \sum_{i=1}^{n} v_{i_X}(A) + |A|$$

Also we can see that

$$\sum_{i=1}^{n} v_i(\theta, A) + |A| = \sum_{X \in S} P[X] \{ \sum_{i=1}^{n} v_{i_X}(A) + |A| \}$$

Since $\sum_{X \in S} P[X] = 1$, we have

$$\implies \sum_{i=1}^{n} v_i(\theta, A) + |A| = \sum_{X \in S} P[X] \sigma_X(A)$$

$$\implies \sum_{i=1}^{n} v_i(\theta, A) + |A| = \sigma(A)$$

$\square$

Thus, by using Lemma 1, we can say that the exact influence maximization algorithm is allocatively efficient and hence VCG payments will give us strategy-proof mechanism for this algorithm. Thus with VCG payments the utility for each agent will be :

$$u_i(\theta, A) = v_i(A) + \sum_{j \neq i} v_j(A) - h(v_1, ... v_{i-1}, v_{i+1}, ... v_n)$$

Here h is some function independent of v_i. Thus we have the following result:

THEOREM 1. *The exact influence maximization algorithm is allocatively efficient and hence is dominant strategy incentive compatible under VCG payments, but the exact algorithm is not incentive compatible without involving payments.*

2.2 An Example

We now give one simple example to illustrate how the proposed model functions. Consider a simple social network graph as shown in Figure 2.

Assume that the true influence probabilities are all 1 in the graph and the algorithm for target set selection is the exact influence maximizing algorithm. If we want to choose only one node as the target set, then clearly node j will be

Figure 2: An Example Social Network

chosen, because $\sigma(\{j\}) = 8$, which is the maximum influence among all nodes present, as can be seen from the figure. Now assume that, all nodes except node k report their true influence. Consider node k, if this node reports its true influence, then node j will be chosen as target set and its utility will be 1. Because she will only be able to influence node m. This is because when cascade reaches node k at $t = 1$, then till that time its neighbor node l would have already been influenced by node j at time $t = 0$. Now if node k lies about its influence on node m as 0, then node i will be chosen as the influence maximizing target set. For this target set, the utility of agent k will be 2 but now the influence will be $\sigma(\{i\}) = 7$. Thus agent k is better off by lying rather than telling the truth. Hence exact influence maximization algorithm is not incentive compatible without payments.

Now assume that we include the Vickrey-Clarke-Groves payment scheme in this scenario and $h(v_1, ..., v_{i-1}, v_{i+1}, ... v_n) = 0$. If node j is chosen as the target set then agent k will be able to influence agent m and she will get the monetary payment of 7 units, thus total utility will be 8 and if any other node is chosen as the target set, then node $k's$ payoff will be lower.

For the value of $k = 1$ the greedy algorithm proposed by Kempe, Kleinberg and Tardos in [8] is same as the exact influence maximization algorithm. Thus the above example also shows that greedy algorithm is not incentive compatible. We can construct examples for other heuristic based algorithms like the high degree heuristics [8], the shapley value based algorithm [11], degree discount heuristics [4] etc. All these algorithms use the information reported by the agents directly to select the target set. Thus, none of these algorithms are strategy-proof. However the randomized algorithm in which we select the nodes in the target set uniformly randomly is strategy-proof. But the expected influence of this algorithm is low as shown by the experiments done in [8].

3. INFLUENCER-INFLUENCEE MODEL

In a real world social network, given a social connection between two persons, both the persons will have information about various properties of the connection. The influencer and the influencee model tries to leverage this fact in designing strategy-proof mechanisms. The advantage of the mechanisms designed for this model is that while deciding the strategies, agents need not know any information beyond its neighborhood.

3.1 The Model

We will now consider the influence and the influencee model of information diffusion and design a payment scheme. The influencer and the influencee model is as follows:

- Given a directed edge (u, v) in the social network, the social planner will ask agent u the influencer about her influence probability θ_{uv} on v and we will also ask agent v the influencee about $u's$ influence on her. Thus we will ask each agent to reveal the probability distribution over each edge which is incident on it and which is emanating from it.

- Consider a directed edge (u, v) with influence probability θ_{uv} on it. Now agent u can activate v with probability θ_{uv} thus we can consider the activation proba-

bility on each edge as probability distribution over set $S = \{active, inactive\}$.

- Given these influence probabilities the social planner will now compute the influence maximizing target set along with the amount of discount to be given to the agents based on their reported probability distribution on edges.

- Consider an agent u. Let $Influencer(u) = \{v|(u, v) \in E\}$ and $influencee(u) = \{v|(v, u) \in E\}$. Thus agent u acts as influencer to nodes in the set $Influencer(u)$ and agent u is the influencee for the nodes in set $Influencee(u)$. In this model an assumption is that agent u knows about the influence probabilities on the edges that are incident on u and that are emanating from u. Thus agent only knows about the influence probabilities in its neighborhood and nothing beyond that.

- Also agent does not know what influence probability is reported by the agents in its neighborhood. The only way an agent can predict the reported probability by its neighbor is by her own assessment of it. Thus we assume that for any given pair of nodes u and v having edge (u, v) between them. We assume the conditional probability distribution function $P(\theta_{uv}^v|\theta_{uv}^u)$ which has all the probability mass concentrated at $\theta_{uv}^v = \theta_{uv}^u$.

- Here we discretize the continuous interval [0,1] into $1/\epsilon + 1$ equally spaced numbers and agents will have to report the influence probability by quoting one of the $1/\epsilon + 1$ numbers. More concretely, given set $T = \{1, 2, ..., t\}$ we define $z \in \{0, \epsilon, 2\epsilon, ..., 1\}^t$ such that $\sum_{i=1}^{t} z_i = 1$. For the case of our problem, $T = \{active, inactive\}$ thus agents will only have to quote one number $\theta_{uv} \in \{0, \epsilon, 2\epsilon, ..., 1\}$.

Based on this model we will now design scoring rule based payment schemes. The overview of the scoring rules is given in the appendix along with lemma 2 that is used in proving theorem 2.

3.2 Scoring Rule Based Direct Payment Scheme

We now consider a scoring rule based payment scheme in which we incentivize agents for reporting the true probability distribution on each edge.

In this mechanism, the payment to an agent i depends on the truthfulness of the distribution she reveals on edges incident on i as well as on the edges emanating from i. Here we assume that the scoring rule used is quadratic scoring rule. In this mechanism the payment received by an agent i is given by

$$\left(v_i(A, \theta) + \frac{d_i^2}{2\epsilon^2}\right)\left(\sum_{j \in Influencee(i)} V_{ji}^i(\hat{\theta}_{ji}^j|\hat{\theta}_{ji}^i) + \sum_{j \in Influencer(i)} V_{ij}^i(\hat{\theta}_{ij}^j|\hat{\theta}_{ij}^i)\right)$$

where, d_i is the degree of agent i, $V_{ij}^i()$ is the expected score that agent i gets for reporting the distribution $\hat{\theta}_{ij}^i$ on the edge (i, j)

THEOREM 2. Reporting true probability distribution is a Nash equilibrium in the direct payment scheme.

PROOF. We will first consider the strategic behavior of some arbitrary agent i considering only the agents in the set $S = Influencee(i) \cup Influencer(i)$. In the payment scheme the agents belonging to the set $V \setminus S$ can only affect the valuation $v_i(A, \theta)$. The expected payoff an agent i gets is given by

$$\sum_{j \in Influencee(i)} \sum_{\theta_{ji}=0}^{1} P(\theta_{ji}^j | \theta_{ji}^i) v_i(A, \theta) V_{ji}^i(\hat{\theta_{ji}^j} | \hat{\theta_{ji}^i}) +$$

$$\sum_{j \in Influencer(i)} \sum_{\theta_{ij}=0}^{1} P(\theta_{ij}^j | \theta_{ij}^i) v_i(A, \theta) V_{ij}^i(\hat{\theta_{ij}^j} | \hat{\theta_{ij}^i}) +$$

$$\frac{d_i^2}{2\epsilon^2} \left(\sum_{j \in Influencee(i)} \sum_{\theta_{ji}=0}^{1} P(\theta_{ji}^j | \theta_{ji}^i) V_{ji}^i(\hat{\theta_{ji}^j} | \hat{\theta_{ji}^i}) + \right.$$

$$\left. \sum_{j \in Influencer(i)} \sum_{\theta_{ij}=0}^{1} P(\theta_{ij}^j | \theta_{ij}^i) V_{ij}^i(\hat{\theta_{ij}^j} | \hat{\theta_{ij}^i}) \right)$$

Note that the valuation $v_i(A, \theta)$ is dependent on the assessment of the influence probabilities by agent i in its neighborhood.

Every agent will now try to maximize the expected payoff by considering the strategy of agents in its neighborhood. If all the agents in the neighborhood are truthful then we have

$$\sum_{j \in Influencee(i)} \sum_{\theta_{ji}=0}^{1} P(\theta_{ji} | \theta_{ji}^i) v_i(A, \theta) V_{ji}^i(\theta_{ji} | \hat{\theta_{ji}^i}) +$$

$$\sum_{j \in Influencer(i)} \sum_{\theta_{ij}=0}^{1} P(\theta_{ij} | \theta_{ij}^i) v_i(A, \theta) V_{ij}^i(\theta_{ij} | \hat{\theta_{ij}^i}) +$$

$$\frac{d_i^2}{2\epsilon^2} \left(\sum_{j \in Influencee(i)} \sum_{\theta_{ji}=0}^{1} P(\theta_{ji} | \theta_{ji}^i) V_{ji}^i(\theta_{ji} | \hat{\theta_{ji}^i}) + \right.$$

$$\left. \sum_{j \in Influencer(i)} \sum_{\theta_{ij}=0}^{1} P(\theta_{ij} | \theta_{ij}^i) V_{ij}^i(\theta_{ij} | \hat{\theta_{ij}^i}) \right)$$

Now consider the expression

$$\left(v_i(A, \theta) + \frac{d_i^2}{2\epsilon^2} \right) \left(\sum_{j \in Influencee(i)} V_{ji}^i(\theta_{ji} | \hat{\theta_{ji}^i}) + \right.$$

$$\left. \sum_{j \in Influencer(i)} V_{ij}^i(\theta_{ij} | \hat{\theta_{ij}^i}) \right)$$

Now Let $\beta = \sum_{j \in Influencee(i)} V_{ji}^i(\theta_{ji} | \theta_{ji}) + \sum_{j \in Influencer(i)} V_{ij}^i(\theta_{ij} | \theta_{ij})$

Thus β is the expected score agent i gets when she reports the true distribution over all the edges. Also, by probability mass assumption this is the expected score she will receive when she reports truthfully.

Thus, if an agent is truthful then, she will receive a payoff given by

$$\left(v_i(A, \theta) + \frac{d_i^2}{2\epsilon^2} \right) \cdot \beta$$

Let, $v_i(A, \theta)$ be the true valuation on an agent and $v_i'(A, \theta)$ be the valuation when agent lies. That is when agent reports some $\hat{\theta_{ij}^i} \neq \theta_{ij}$ Consider the utility of an agent when she lies on only one of the edges

$$\left(v_i'(A, \theta) + \frac{d_i^2}{2\epsilon^2} \right) \cdot \left(\beta - 2\epsilon^2 \right)$$

When an agent reports probability value that is $\pm \epsilon$ away from true probability value the quadratic scoring rule (lemma 2) ensures that agent gets payoff lower by $2\epsilon^2$ for that edge. Now, we assume the worst case scenario in which agent gains maximum by reporting the false probability value on only single edge and that too minimum possible deviation from the true value, Since we divide the [0,1] probability interval into $1/\epsilon$ numbers, agent has to report the probability value that is at least ϵ away from the true probability value. That is, agent will have to report some probability value $\hat{\theta_{ij}^i} = \theta_{ij} \pm \epsilon$. This gives us $\left(\beta - 2\epsilon^2 \right)$.

An agent cannot get more score for lying as scoring rule is incentive compatible. Thus agent can gain only in the valuation part. Now by reporting false probability agent can get valuation greater than true valuation by say δ that is,

$$v_i'(A, \theta) = v_i(A, \theta) + \delta$$

Where $\delta \in [0, \sum_{j \in influencer(i)} \theta_{ij} - v_i(A, \theta)]$

Thus by reporting false probability value an agent gets payoff of

$$\left(v_i(A, \theta) + \delta + \frac{d_i^2}{2\epsilon^2} \right) \cdot \left(\beta - 2\epsilon^2 \right)$$

Thus for the agent to remain truthful we require that

$$\left(v_i(A, \theta) + \frac{d_i^2}{2\epsilon^2} \right) \cdot \beta \geq \left(v_i(A, \theta) + \delta + \frac{d_i^2}{2\epsilon^2} \right) \cdot \left(\beta - 2\epsilon^2 \right)$$

$$\implies d_i^2 + (v_i(A, \theta) + \delta) \, 2\epsilon^2 \geq \delta \beta$$

Now, we have $\delta \beta \leq d_i^2$ and $(v_i(A, \theta) + \delta) \, 2\epsilon^2 \geq 0$. Thus, it is a best response strategy for an agent to report truthfully when the agent in its neighborhood are truthful.

Now we still have to resolve the case for the agents belonging to set $V \setminus S$. Now these agents only affect the valuation of agent i. In the above analysis the best response strategy for an agent was derived by assuming that with minimum possible deviation from reporting true probability values agent i is able to gain maximum possible valuation. Thus even if agents in $V \setminus S$ report anything the best strategy for agent i is to report truthfully. Thus without knowing the strategy of agents in set $V \setminus S$ the best response for an agent i is to report truthfully. Thus reporting true influence probabilities is a Nash equilibrium in this mechanism. $\square$

Note that the social planner will have to fix the value of ϵ which will decide the accuracy of the probability values extracted from the users. Smaller the value of ϵ, greater the payment the seller will have to make to the users. The main advantage of this mechanism is that the seller can use any

of the algorithms to select the target set. The agents will be truthful regardless of which target set is chosen.

The same payment scheme can be used with other scoring rules namely spherical and weighted scoring rule. For the spherical scoring rule the expression for payoff is given by :

$$\left(v_i(A,\theta) + \frac{2d_i^2}{3\epsilon^2}\right)\left(\sum_{j\in Influencee(i)} V_{ji}^i(\hat{\theta_{ji}^j}|\hat{\theta_{ji}^i})+\right.$$

$$\left.\sum_{j\in Influencer(i)} V_{ij}^i(\hat{\theta_{ij}^j}|\hat{\theta_{ij}^i})\right)$$

For the weighted scoring rule the payoff is given by :

$$\left(v_i(A,\theta) + \frac{d_i^2}{\epsilon^2}\right)\left(\sum_{j\in Influencee(i)} V_{ji}^i(\hat{\theta_{ji}^j}|\hat{\theta_{ji}^i})+\right.$$

$$\left.\sum_{j\in Influencer(i)} V_{ij}^i(\hat{\theta_{ij}^j}|\hat{\theta_{ij}^i})\right)$$

The standard scoring rules like quadratic and spherical are not appropriate for the direct payment scheme. This is because even if the influence probability on an edge is zero, both these scoring rules will give the expected score of one. Thus, even if the social network is the empty graph, in which all the edges are inactive, these standard payment schemes will give maximum possible expected score.

3.2.1   The Reverse Weighted Scoring Rule

We will now propose the reverse weighted scoring rule derived from the weighted scoring rule which has the following desirable properties:

1. Incentive compatibility.

2. The expected score is proportional to the influence probability.

3. If $\theta_{ij} = 0$ then the expected score for the edge (i,j) to both the agents u and v is zero. That is, $V_{ij}^i(\theta_{ij}^j|\theta_{ij}^i) = V_{ij}^j(\theta_{ij}^i|\theta_{ij}^j) = 0$ if $\theta_{ij} = 0$.

Property 1 is required to truthfully extract the influence probabilities. Property 2 is desirable because the social planner would want to reward the agent which revealed the social connection through which the product can be sold with high probability. Property 3 ensures that an agent does not get anything for revealing a social connection through which the product cannot be sold.

Following is the reverse weighted scoring rule that has all these properties.

$$S_i(z) = 2z_i(t-i) - \sum_{j=1}^{t} z_j^2(t-j)$$

We can show that the above scoring rule is incentive compatible using arguments similar to the weighted scoring rule [5].

Implementation of Mechanisms

Since all the mechanisms presented under the influencer-influencee model involves payments, the influence maximiza-tion process should involve monetary transfers. Viral marketing is one such process and we will now discuss the implementation of the mechanism in the context of viral marketing in an online social network like facebook, orkut, etc.

Consider that a seller wants to market a certain product in the social network. The seller can now ask each user in the social network to reveal her influence on each of the users in their friends list. Users have incentive to participate in this mechanism because each user will get a fixed positive payment based on the influence probabilities they report. The seller can ask each agent to report the influence probability by developing the application on the social network. There are a large number of applications on the social networks like orkut, facebook etc. which users use extensively for playing games and socializing online.

In an online social network like facebook for example, if a user is interested in the product to be sold, then she can grant the access to the application. Now, this application will have full access to the friends list and other profile information of the user that is public. Thus such an application can be easily implemented in an online social network without any privacy issues. The seller will first have to fix the level of accuracy that she needs before starting the information extraction process.

Now, given the influence probabilities, the application will now compute the influence maximizing target set. The application will also compute the payment to be made to the user. This payment can be made in the form of a discount on the product to be marketed. The seller will now give the product for free to the users selected in the target set. The sale of the product will now proceed as in the independent cascade process and users will get appropriate discounts based on the reported influence probabilities. Since the payment scheme ensures incentive compatibility, every user will truthfully reveal the information to the seller via the application.

4.   SUMMARY AND FUTURE DIRECTIONS

In this paper we designed game theoretic models of information diffusion process in a social network in order to extract the influence probabilities from the users. We designed various incentive compatible mechanism which can be deployed in a social network to select the target set. All these mechanisms are individually rational and the payments in these mechanisms come from the seller. To the best of our knowledge this is the first time that the problem of extracting the influence probabilities from the social network has been studied with the perspective of game theory. Indeed, this work opens up a host of interesting problems

- In the influencer model, does there exist an incentive compatible algorithm having a constant factor approximation ratio with or without payments?

- In the influencer model, does there exist a heuristic based incentive compatible algorithm which may not have theoretical guarantees about the approximation but in a practical sense it performs well?

- In the direct payment scheme, the payments depend on ϵ which decides the accuracy of the probability distribution. The higher the accuracy is required, the more is the payment to be made to the user. An interesting

direction of future research would be to design the incentive compatible mechanisms that are independent of this factor.

- The models considered in this paper do not reflect the exact real world situation. Indeed this is the first attempt to design a game theoretic model of the information diffusion process. In this work we only considered the game theoretic version of simple model of independent cascade process. Developing a game theoretic model of a more realistic information diffusion model given in [3] will be an exciting direction of future research.

5. REFERENCES

[1] N. Alon, F. Fischer, A. Procaccia, and M. Tennenholtz. Sum of us: Strategyproof selection from the selectors. *Working paper*, 2010.

[2] A. Altman and M. Tennenholtz. Incentive compatible ranking systems. *Proceedings of the 6th International Conference on Autonomous agents and Multiagent Systems, AAMAS, article no. 5*, 2007.

[3] D. Arthur, R. Motwani, A. Sharma, and Y. Xu. Pricing strategies for viral marketing on social networks. *5th Workshop on internet and network economics, WINE*, pages 101–112, 2009.

[4] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. *Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data mining, KDD*, pages 199–208, 2003.

[5] D. Dixit and Y. Narahari. Quality concious and truthful query incentive networks. *5th Workshop on Internet and Network Economics, WINE*, pages 386–397, 2009.

[6] P. Domingos and M. Richardson. Mining the network value of customers. *Proceedings of the 7th ACM SIGKDD Conference on Knowledge Discovery and Data mining, KDD*, pages 47–56, 2001.

[7] A. Goyal, F. Bonchi, and L. Lakshmanan. Learning influence probabilities in social networks. *Proceedings of The Third ACM International Conference on Web Search and Data Mining, WSDM*, pages 241–250, 2010.

[8] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data mining, KDD*, pages 137–146, 2003.

[9] Y. Narahari, D. Garg, N. Ramasuri, and P. Hastagiri. *Game Theoretic Problems in Network Economics and Mechanism Design Solutions.* Springer Book Series on Advanced Information and Knowledge Processing (AI & KP), Springer, London, 2009.

[10] N. Nisan, E. Tardos, V. Vazirani, and T. Roughgarden, editors. *Algorithmic Game Theory.* Cambridge University Press, 2007.

[11] N. Ramasuri and Y. Narahari. Determining the top-k nodes in social networks using the shapley value. *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems, AAMAS*, 3:1509–1512, 2008.

[12] R. Selten. Axiomatic characterization of the quadratic scoring rule. *Discussion Paper Serie B 390, University of Bonn, Germany*, 1996.

APPENDIX
A. PRELIMINARIES

Various probabilistic models have been proposed to model the spread of information in social networks. In this section, we will give a brief overview of the *Independent Cascade Model.* We will represent any social network by directed graph G. We will say that an individual is active if she is the adopter of the innovation or the behavior and inactive otherwise. We will assume that once an individual becomes active, she cannot switch back to being inactive.

A.1 Independent Cascade Model

In this model we activate some set of nodes A_0 initially. The information diffusion process unfolds in discrete time steps as follows.

Each node that becomes active at time step t will try to activate each of its neighbors. A node will get the chance to activate its neighbors only once, that is at the time instant in which it became active. A node i will successfully activate its neighbor j with probability p_{ij}. Thus p_{ij} is the probability that node i will activate node j conditioned on the event that node j will be inactive when node i got activated. The successfully activated nodes at time instant t will now activate their neighbors at time instant $t+1$. This process ends when no more nodes get activated. Clearly this process will end in at most $n-1$ time steps.

A.2 Influence Maximization Problem

We now introduce the influence maximization problem in social networks. Before that, we define the notion of influence function:

Definition 1. Given an initially active set A and influence probabilities, the influence function denoted by $\sigma(A)$ is the expected number of active nodes at the end of the diffusion process. Thus for all $A \subseteq V$ and influence probabilities on each edge, we have

$$\sigma : 2^V \longrightarrow [0, n]$$

The influence maximization problem is, given a parameter k, a social network graph G, and a model of information diffusion, find a set of k nodes in G to be activated initially (also known as target set) such that, it will maximize the influence function $\sigma(A)$.

B. SCORING RULES

Scoring rules [12] are used to compare the observed probability distribution with a predicted one. Scoring rules consist of a sequence of scoring functions $S_1, S_2, ..., S_t$ where each S_i assigns score $S_i(p)$ to each $p \in \triangle(T)$ where $T = \{1, 2, ..., t\}$. For our problem, we will consider only real valued scoring functions.

Expected Score

Let us assume that $w \in \triangle(T)$ be the true or observed distribution and $z \in \triangle(T)$ be the predicted distribution then the expected score of reporting z for observed distribution

of w is given by

$$V(z|w) = \sum_{i=1}^{t} w_i S_i(z)$$

The expected score loss $L(z|w)$ of reporting z for observed distribution w is given by

$$L(z|w) = V(w|w) - V(z|w)$$

We now define proper scoring rule or incentive compatible scoring rule as:

Definition 2. Proper Scoring rule is defined as sequence of scoring functions $S_1, S_2, ..., S_n$ such that for any $z, w \in \triangle(T)$ and $z \neq w$ we have, $L(z|w) > 0$

Thus, if the scoring rule is proper, or incentive compatible then, it is the best response for each agent to report its true probability distribution.

There are several scoring rules which are incentive compatible, some of them are:

1) Quadratic Scoring rule

$$S_i(z) = 2z_i - \sum_{j=1}^{t} z_j^2$$

2) Weighted scoring rule

$$S_i(z) = \frac{2i \cdot z_i - \sum_{j=1}^{t} z_j^2 \cdot j}{t}$$

3) Spherical scoring rule

$$S_i(z) = \frac{z_i}{\sqrt{\sum_{j=1}^{t} z_j^2}}$$

The following lemma quantifies the amount of loss that an agent suffers by deviating from the true value.

LEMMA 2. *If $w, z \in \{0, \epsilon, 2\epsilon, ..., 1\}^t$, $0 < \epsilon \leq 1$ such that $\sum_{i=1}^{t} w_i = 1$ and $\sum_{i=1}^{t} z_i = 1$ and $z_i = w_i \pm \epsilon$ for some integer $1 \leq i \leq t$, then*

- *For quadratic scoring rule*

$$V(z|w) = V(w|w) - 2\epsilon^2$$

- *For weighted and reverse weighted scoring rule*

$$V(z|w) = V(w|w) - \epsilon^2$$

- *For the spherical scoring rule*

$$V(z|w) = V(w|w) - 1.5\epsilon^2$$

Table 1: Notations used in the paper

Symbol	Meaning
$V = \{1, 2, ..., n\}$	Set of agents in a social network
E	Set of edges in the social network
$N(i)$	Set of neighbors of a node i in the given graph
d_i	Degree of agent i
$d(x, y)$	Shortest path distance between vertices x and y in a graph
A	Target set
$\sigma_X(A)$	Number of active nodes given a fixed outcome X and a target set
$\sigma(A)$	Expected number of nodes influenced by target set A
θ_{ij}	Probability that node i will influence node j
$\hat{\theta_{ij}^i}$	The reported probability of influence on edge $(i, j) \in E$ by agent i
θ_{ij}^i	The probability of influence on edge $(i, j) \in E$ as perceived by agent i
θ_i	The influence vector reported by agent i to the social planner
θ	The true influence vector for the entire graph
$u_i(\theta, A)$	Utility of an agent i
$v_i(\theta, A)$	Valuation function of agent i
$t_i(v_1, v_2, ..., v_n)$	The payment function for agent i
$S_1, S_2, S_3, ..., S_n$	Set of scoring functions that define the scoring rule
$\triangle(T)$	Set of all possible probability distributions over set
$V(z|w)$	The expected score agent gets for reporting distribution z given the true distribution w
$L(z|w)$	The expected loss agent suffers for reporting distribution z given the true distribution w
ϵ	The quality of approximation of the [0,1] interval
$V_{ij}^i(\hat{\theta_{ij}^i}|\theta_{ij}^j)$	The expected score agent i gets for reporting $\hat{\theta_{ij}^i}$ given the estimate θ_{ij}^j by agent j
$P(\theta_{ji}^j|\theta_{ji}^i)$	The belief probability of i given her type θ_{ji}^i