2011 31st International Conference on Distributed Computing Systems Workshops

Approximation and Inapproximation for The Influence Maximization Problem in Social Networks under Deterministic Linear Threshold Model

Zaixin Lu*, Wei Zhang†, Weili Wu*, Bin Fu‡ and Dingzhu Du*
*Computer Science Department
University of Texas at Dallas, Richardson, USA
Email: zaixinlu,weiliwu,dzdu@utdallas.edu
†Math Department
Xi'an Jiaotong University, Xi'an, PRC
Email: dc.zhangwei@gmail.com
‡Computer Science Department
University of Texas-Pan American, Edinburg, USA
Email: binfu@cs.panam.edu

***Abstract*—Influence Maximization is the problem of finding a certain amount of people in a social network such that their aggregation influence through the network is maximized. In the past this problem has been widely studied under a number of different models. In 2003, Kempe *et al.* gave a $(1-\frac{1}{e})$-approximation algorithm for the *linear threshold model* and the *independent cascade model*, which are the two main models in the social network analysis. In addition, Chen *et al.* proved that the problem of exactly computing the influence given a seed set in the two models is #P-hard. Both the *linear threshold model* and the *independent cascade model* are based on randomized propagation. However such information might be obtained by surveys or data mining techniques, which makes great difference on the properties of the problem. In this paper, we study the Influence Maximization problem in the *deterministic linear threshold model*. As a contrast, we show that in the *deterministic linear threshold model*, there is no polynomial time $n^{1-\epsilon}$-approximation unless P=NP even at the simple case that one person needs at most two active neighbors to become active. This inapproximability result is derived with self-contained proofs without using PCP theorem. In the case that a person can be activated when one of its neighbors become active, there is a polynomial time $\frac{e}{e-1}$-approximation, and we prove it is the best possible approximation under a reasonable assumption in the complexity theory, $NP \not\subset DTIME(n^{\log\log n})$. We also show that the exact computation of the final influence given a seed set can be solved in linear time in the *deterministic linear threshold model*. The Least Seed Set problem, which aims to find a seed set with least number of people to activate all the required people in a given social network, is discussed. Using an analysis framework based on Set Cover, we show a $O(\log n)$-approximation in the case that a people become active when one of its neighbors is activated.**

***Keywords*-influence maximization; social network; approximation; deterministic model;**

I. Introduction

A social network is a graph of relationships (edges) and individuals (nodes). One of the issues usually considered by marketing managers in this field is how to maximize the spread of information through a social network. For instance, in order to promote a new product, one can give a few influential people free samples of the product. Probably those people will recommend the product to their friends and many individuals will ultimately try it through such "word-of-mouth" effect. The Influence Maximization (IM) problem is how to select the "few" initial people as seed set, such that the spread of influence can be maximized. In order to study the complexity of the this problem, we first have to get about another problem, that is how to compute the total influence given a seed set. Please see [1], [2], [9] for recent works.

The node selection problem we just mentioned was first proposed by Domingos and Richardson in [4] and [8] respectively. They considered the relations of individuals and proposed a probabilistic propagation model for this problem. Kempe *et al.* in [6], [7] further formulated it into an optimization problem and studied it on two different models: the *independent cascade model* proposed by Goldenberg *et al.* in [5], [12] and the *linear threshold model* proposed by Granovetter and Schelling in [10] and [11] separately. Kempe *et al.* proved the natural greedy algorithm achieves a $(1-\frac{1}{e})$-approximation simply by showing that the influence spreads under both the two models are submodular. Thereafter, Chen *et al.* in [1], [2] showed that the problem of exactly computing the influence given a seed set in both the *independent cascade model* and the *linear threshold model* are #P-hard, which indicates that the greedy algorithm is not a polynomial time approximation for the two models.

In the *independent cascade model*, the propagation procedure is based on a probabilistic way; individuals can successfully activate their neighbors with certain probabilities. In the *linear threshold model*, the propagation procedure is in a threshold manner; the influence from an individual v_i to another individual v_j is presented by a weight $w_{i,j}$ and an individual can be activated when the sum of influences it receives exceeds a randomly determined threshold. It is

1545-0678/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDCSW.2011.33

worthy to mention that the thresholds in the *linear threshold model* are randomly updated during the spread process. Therefore, it can be seen that both the *linear threshold model* and the *independent cascade model* are based on randomized propagation. However the thresholds might be estimated by surveys and data mining techniques. If an individual can be activated when the sum of influences exceeds a predetermined threshold, we say the propagation procedure is based on the *deterministic linear threshold model*. In this paper, we focus on studying the approximation and inapproximation for the Influence Maximization problem in the *deterministic linear threshold model*. The main contribution of this paper includes:

1) We show that in the *deterministic linear threshold model*, there is no polynomial time $n^{1-\epsilon}$-approximation for the IM problem unless P=NP even in the simple case that a person needs at most two active neighbors to become active. 2) We also show that the problem of exactly computing the influence given a seed set in this model can be solved in linear time. 3) In the case that a person can be activated after one of its neighbors becomes active, there is a polynomial time $\frac{e}{e-1}$-approximation. 4) The Least Seed Set (LSS) problem, which is a variation of the IM problem, is discussed. It aims at finding a seed set with least number of people such that all the people of interest in the social network can be finally activated. We give a $O(\log n)$-approximation for the case that a node can be activated by anyone of its neighbors.

The rest of this paper is organized as follows. In section II, we present a linear time exact algorithm to compute the influence spread for a seed set. In section III, we study the IM problem in the case that a people can be activated by anyone of its neighbors. An inapproximation result for the case that a person needs two active neighbors to become active is provided in section IV. In section V, we show the approximation and inapproximation for a special case of the LSS problem. In section VI, we conclude our paper and discuss the future work.

II. Computing the Influence Spread in the Deterministic Linear Threshold Model

For the *linear threshold model* and the *independent cascaded model*, the problem of computing the influence spread given a seed set was left as an open problem in [6]. Chen *et al.* closed this open problem by showing its #P-hardness in [1], [2].

Definition 1. *A **social network** is a directed graph $G(V, E)$, each node v_i in V with a threshold t_i represents a person in the social network, each directed edge (v_i, v_j) has weight $w_{i,j}$ that denotes how much the node v_j is influenced by the node v_i.*

In the *deterministic linear threshold model*, the propagation process has the following provisions:

1) Let $x_j = 1$ denote v_j is active, and $x_j = 0$ denote v_j is not.

2) At any time, v_i becomes active if and only if $\sum_{v_j \in neighour(v_i)}(x_j \cdot w_{j,i}) \geq t_i$.

3) The diffusion is a step by step process: in step t, all nodes that were active in step $t-1$ remain active, and any node v satisfies the active condition will be activated.

Definition 2. *Given a social network $G(V, E)$ and a seed set A of initially active nodes, the **Influence Computation** problem is to find all the nodes that will be activated directly or indirectly by the nodes in A.*

In the next, we show that this problem in the *deterministic linear threshold model* can be solved in linear time.

A. A Linear Time Algorithm

Theorem 1. *Given a social network $G(V, E)$ and a seed set A, the problem of exactly computing the influence spread can be solved in linear time in the* deterministic linear threshold model.

Algorithm 1 Influence Computing

Input: A directed graph $G(V, E)$, a threshold t_i for each node $v_i \in V$, a weight $w_{i,j}$ for each edge $(v_i, v_j) \in E$ and a seed set $A \subseteq V$.
Output: the set of all nodes that will be activated in the network.
For each node v_i in V, let $H_i \leftarrow 0$; (H_i holds the sum of influences for v_i.)
Mark all the nodes in A as newly activated nodes;
repeat
  for Each newly activated node v_j **do**
    for Each its non active neighbor v_k **do**
      Let $H_k \leftarrow H_k + w_{j,k}$;
      if $H_k \geq t_k$ **then**
        Mark v_k as newly activated;
      end if
      Mark v_j as active node; (Note that active node $\neq$ newly activated node.)
    end for
  end for
until There is no newly activated nodes

Proof: The time complexity easily follow from the Alg. 1. It terminates when there are no newly activated nodes. Each edge in E has only one chance to be used to adjust its neighbors. Assume the input directed graph G does not have isolated nodes, which means that $|E| \geq |V|$. Hence the Alg. 1 has time complexity $O(|E|)$. ■

B. A Lower Bound for Computing the Influence Spread

Theorem 2. *For any $\alpha \in [0, 1]$, there is a class of graphs $G(V, E)$ with $|E| = \Theta(n^{1+\alpha})$ such that every algorithm that*

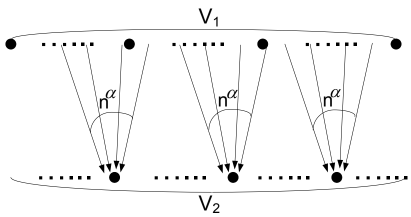

Figure 1. $(1+\alpha)$-Bipartite Graph

exactly computes the final activated nodes given an initial seed set needs at least $\Omega(|E|)$ *running time, where* $n = \frac{|V|}{2}$.

Proof: We can design a graph $G(V_1, V_2, E)$ as shown in Fig. 1, where $|V_1| = |V_2| = n$. Let the in-degree of each node in V_2 be n^α, where $\alpha \in [0, 1]$. Therefore we allow a constant factor degree difference among the degrees of nodes in V_2. It is easy to see the existence of this kind of graphs and we call them $(1+\alpha)$-graph.

Assume that there exists an $o(|E|)$ time algorithm $h(.)$ to find the set of activated nodes from an initial set of active nodes. Let G_1 be a $(1+\alpha)$-graph such that for each edge (v_i, v_j), it has weight $w_{i,j} = \frac{1}{degree(v_j)}$. The threshold $t_j = 1$ for all nodes v_j in G_1. Assume the seed set is the set of all the nodes in V_1. Since $h(.)$ runs in $o(|E|)$ time, there exists an edge (v_i, v_j) that the algorithm does not access. Let G'_1 be the same graph as G_1 except $w'_{i,j} = 0$. Since $h(G'_1, V_1)$ does not access edge (v_i, v_j), we have $h(G_1, V_1) = h(G'_1, V_1) = V_1 \cup V_2$, which is a contradiction. ■

III. Approximation and Inapproximation for the One-Activate-One Model

In this section, we consider the IM problem under the condition that a person can be activated by anyone of its neighbors. It is a special case of the general *deterministic linear threshold model*, in which a person needs multiple active neighbors with weights to be activated. The problem is solved by transforming it into the maximum coverage problem. We derive a constant factor approximation that matches the bound of the inapproximation.

The Influence Maximization Problem under the One-Activate-One Model: Let $G(V, E)$ be a directed graph. Each directed edge (u, v) from node u to node v represents that person u can activate person v. The problem is to initially activate k people so that the largest number of people will be activated eventually.

For an approximation solution, let $\mathrm{Approximate}(G, k)$ represent the number of people who are eventually activated directly or indirectly by the first selected k people and let $\mathrm{Optimal}(G, k)$ represent the optimal solution. The problem has a polynomial time approximation ratio c if and only if $\mathrm{Approximate}(G, k)$ can be computed in polynomial time and $\mathrm{Optimal}(G, k) \le c \times \mathrm{Approximate}(G, k)$ for every input instance (G and k).

The maximum coverage problem is a classical problem in the computational complexity theory. The input is a list of m sets and an integer k. The target is to select k sets from the list to cover the largest number of elements in the ground set. It is well known that the maximum coverage problem, such as Set Cover problem, Vertex Cover problem, and Independent Set problem, has a $\frac{e}{e-1}$-approximation and the approximation ratio is optimal.

Theorem 3. *There is a polynomial time* $\frac{e}{e-1}$*-approximation algorithm for the Influence Maximization problem in the One-Activate-One model.*

Proof: Assume that we have an input instance (G, k) for the One-Activate-One IM problem. As shown in Alg. 2, for each vertex v_i in G, use the depth first search algorithm to find all the vertices that are reachable from v_i. This can be done in linear time. The problem becomes a maximum cover problem. Therefore, it has an polynomial time approximation with $\frac{e}{e-1}$. ■

Algorithm 2 Influence Maximization

Input: A directed graph $G(V, E)$ and an integer k.
Output: a set of k nodes in V.
Let H denote the set of active nodes and P denote the initial set.
for each node $v_i \in V$, which has no incoming edges **do**
  Let S_{v_i} represent the set of nodes reachable from v_i.
end for
for $j \leftarrow 1$ to k **do**
  Select the set S_{v_i} that maximize $|H \cup S_{v_i}|$
  $H \leftarrow H \cup S_{v_i}$ and $P \leftarrow P \cup \{v_i\}$.
end for

Theorem 4. *If the Influence Maximization problem in the One-Activate-One model has an approximation algorithm with approximation ratio d, then the maximum coverage problem has an approximation with ratio* $d + o(1)$.

Proof: Assume $S_1, \cdots, S_m$ is a list of sets for the maximum coverage problem. We are going to construct a directed graph $G(V, E)$. Assume $S_1 \cup S_2 \cup \cdots \cup S_m = \{a_1, \cdots, a_n\}$. For each set S_i, create a vertex v_i in V. For each a_j, create hk^2 vertices $u_{j,1}, \cdots, u_{j,t}$ in V, where $t = hk^2$ and h is a large constant. If $a_j \in S_i$, add directed edges from v_i to all $u_{j,1}, \cdots, u_{j,t}$ in E.

We claim that an optimal solution for the maximum coverage problem can cover x_o elements by selecting k sets if and only if the One-Activate-One IM problem can activate $x_o t + k$ people by initially activating k people.

Let a d-approximation solution for the IM problem to activate $yt+k$ people. We have $x_o t + k \le d(yt+k)$. We can assume $y \ge 1$; otherwise, all sets $S_1, \cdots, S_m$ are empty. We also assume $k \ge 1$; otherwise, the problem is trivial. Thus,

$$
\begin{aligned}
x_o &\le \frac{dyt + (d-1)k}{t}, &(1)\\
&\le dy + \frac{(d-1)k}{t}, &(2)\\
&\le (d + \frac{(d-1)k}{ty})y, &(3)\\
&\le (d + \frac{(d-1)}{h})y, &(4)\\
&\le (d + o(1))y. \quad (h \text{ is large}) &(5)
\end{aligned}
$$

Hence, if the One-Activate-One IM problem has d-approximation, we can also have an $d - o(1)$ approximation for the maximum coverage problem. ■

Corollary 1. *There is no polynomial time* $(\frac{e}{e-1} - o(1))$*-approximation for the One-Activate-One Influence Maximization problem unless* $NP \subseteq DTIME(n^{\log\log n})$.

Proof: It simply follows from the Thm. 4 and Feige *et al.*'s paper [3]. ■

IV. Inapproximation for the Two-Activate-One Model

In this section, we study a more general IM problem that one person can be activated by one or two people. We derive a strong inapproximation result for this problem even in bounded degree graphs. To show the inapproximation, we reduce the Set Cover problem to the IM problem in polynomial time. The input of a Set Cover problem is an integer k, and many sets $S_1, \cdots, S_m$ and S, where $S_1, \cdots, S_m$ are m subsets of set S. The target is to find k subsets $S_{i_1}, \cdots, S_{i_k}$ such that $S_{i_1} \cup \cdots \cup S_{i_k} = S$. It is well known that the Set Cover problem is NP-hard.

A. Bounded Degree Graphs

Definition 3. *A graph G is a directed bounded (d_1, d_2)-graph if every node in G has at most d_1 incoming edges and at most d_2 outgoing edges. The IM problem over such a directed graph is called (d_1, d_2)-Influence Maximization ((d_1, d_2)-IM) problem.*

Lemma 1. *Assume that $w_1, \cdots, w_n$ are nodes with at most one incoming edge, and x is an isolated node. We can create $O(n)$ new nodes to form a $(2, 2)$ graph such that x is activated if and only if one of the nodes in $\{w_1, \cdots, w_n\}$ is activated. Furthermore, x has at most one incoming edge and no outgoing edge.*

Proof: We can achieve the goal by finishing n phases. In the first phase, we create a new node b_1 and add an edge from w_1 to b_1, and another edge from w_2 to b_1. If w_1 or w_2 is activated, then b_1 will be activated. In the phase $i+1$, for any $1 < i < n-1$, we create a new node b_{i+1}, add an edge from b_i to b_{i+1} and another edge from w_{i+2} to b_{i+1} such that b_{i+1} will be activated when b_i or w_{i+2} is activated.

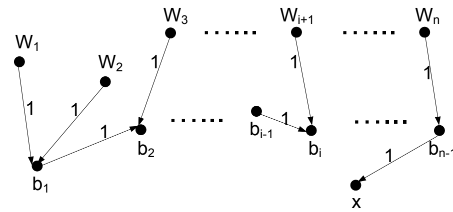

Figure 2. One-to-One Graph

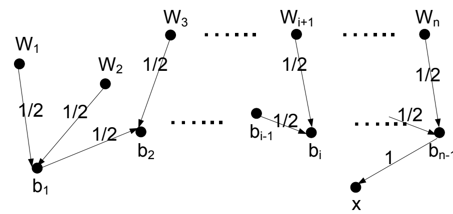

Figure 3. All-to-One Graph

After the phase $n-1$, we add an edge from b_{n-1} to x so that b_{n-1} can activate x. The total number of new nodes is $n-1$. Each node b_i has two incoming edges and one outgoing edge. Node x has one incoming edge and no outgoing edge. The final $(2, 2)$ graph is shown in Fig. 2, each node in the graph has a determined threshold 1. ■

Lemma 2. *Assume that $w_1, \cdots, w_n$ are nodes with at most one incoming edge, and x is an isolated node. We can create $O(n)$ new nodes to form a $(2, 2)$ graph such that x is activated if and only if $w_1, \cdots, w_n$ are all active. Furthermore, x has at most one incoming edges and no outgoing edge.*

Proof: The overall argument is similar to the proof of Thm. 1. We still use n phases to construct the $(2, 2)$ graph. In the first phase, create b_1 and add an edge from w_1 to b_1 with weight $\frac{1}{2}$, and another edge from w_2 to b_1 with weight $\frac{1}{2}$. If both w_1 and w_2 are activated, then b_1 will be activated. In the phase $i+1$, we create node b_{i+1}, add an edge from b_i to b_{i+1} and an edge from w_{i+2} to b_{i+1}. If both b_i and w_{i+2} are active, then b_{i+1} will be activated.

After phase $n-1$, we add an edge from b_{n-1} to x so that b_{n-1} itself can activate x. The total number of new nodes is $n-1$. Each node b_i has two incoming edges and one outgoing edge. Node x has one incoming edge and no outgoing edge. The final $(2, 2)$ graph is shown in Fig. 3, each node in the graph has a determined threshold 1. ■

B. An Inapproximation Result in the Bounded Degree Graphs

Theorem 5. *For any constant $\epsilon \in (0, 1)$, there is no polynomial time $p^{1-\epsilon}$-approximation for the $(2, 2)$-IM problem* *unless P=NP, where p is the number of nodes in the directed graph of social network.*

Proof: We give a polynomial time reduction from the Set Cover problem to a directed $(2,2)$ graph. Let $S_1, \cdots, S_m$ be the input for the Set Cover problem and $S_1 \cup S_2 \cup \cdots S_m = \{a_1, \cdots, a_n\}$. Without loss of generality, assume $\epsilon < \frac{1}{100}$. Let p be the number of nodes in the graph, we first define the following parameters:

$$\begin{aligned}
p &= (n+m)^{\frac{20}{\epsilon}}, & (6)\\
g(p) &= p^{\epsilon}, & (7)\\
m_5 &= p^{1-8\epsilon}, & (8)\\
p^{8\epsilon} \geq m_8 &\geq \frac{3}{4}p^{8\epsilon}, & (9)\\
n &\leq p^{\frac{\epsilon}{10}}, & (10)\\
k \leq m &\leq p^{\frac{\epsilon}{10}}. & (11)
\end{aligned}$$

The inequalities (10) and (11) follow from equality (6). We construct the $(2,2)$ graph as follows:

Phase 1: For each set S_i, create a vertex u_i.

Phase 2: For each set S_i with $a_j \in S_i$, create a vertex $x_{i,j}$ and an edge from u_i to $x_{i,j}$ such that node u_i activates node $x_{i,j}$. If there are more than two elements in S_i, create a binary tree with root u_i such that u_i activates all of them.

Phase 3: Let H_j be the set of all nodes $x_{i,j}$. For each group H_j, create a vertex y_j and add some additional vertices such that y_j will be activated if one of the nodes in H_j is active. By Lemma 1, this part can be done in polynomial time.

Phase 4: For each element a_j, create a vertex v_j. For each y_j, create an edge from y_j to v_j such that node y_j activates v_j.

Phase 5: For each node v_j, create a binary tree T_j with root v_j such that the tree has m_5 leaves and v_j can activate all of those leaves. We label all the leaves as $l_{1,j}, \cdots, l_{m_5,j}$.

Phase 6: Create m_5 groups $G_1, \cdots, G_{m_5}$ of nodes. Each group G_i contains nodes $w_{i,1}, \cdots, w_{i,n}$. Create an edge from $l_{i,j}$ to $w_{i,j}$ such that node $l_{i,j}$ activates node $w_{i,j}$.

Phase 7: For each group G_i, create a node x_i such that x_i will be activated if and only if all elements $w_{i,1}, \cdots, w_{i,n}$ in G_i are activated. By Lemma 2, this part can be done in linear time for each group G_i.

Phase 8: For each x_i, create a path, which is called Y_i, starting from x_i and has length m_8 such that every nodes in the path Y_i can activate the next one.

Let $A_{1,7}$ be the set of nodes created from Phase 1 to Phase 7. Let A_8 be the set of nodes created in Phase 8. By inequalities (6) to (11), the number of nodes in $A_{1,7}$ is bounded by

$$|A_{1,7}| \leq O((n+m)^2 m_5) = o(p^{1-\epsilon}). \quad (12)$$

We also have the following equation:

$$p = |A_{1,7}| + m_5 m_8. \quad (13)$$

If we can select k nodes among $u_1, \cdots, u_k$ such that the corresponding k subsets cover the entire set S, then $v_1, \cdots, v_n$ can be activated. Thus, all the p nodes in the constructed $(2,2)$ graph will become active.

Assume that there is a $g(p)$-approximation algorithm which activates nodes in a set B. We claim that if $|B| < \frac{p}{g(p)}$, then there is no solution for the Set Cover with k subsets. Assume

$$|B| \geq \frac{p}{g(p)} = p^{1-\epsilon}. \quad (14)$$

Thus, the number of activated nodes in the set A_8 is at least

$$\begin{aligned}
|B| - |A_{1,7}| &\geq |B| - o(p^{1-\epsilon}), \quad \text{(by (12))} & (15)\\
&\geq \frac{1}{2}|B|, \quad \text{(by (14))} & (16)\\
&\geq 50km_8. \quad \text{(by (6) to (11))} & (17)
\end{aligned}$$

We are going to transform the first k selected vertices so that only the nodes in $u_1, \cdots, u_m$ are selected. The transformation follows the following rules:

1) If a selected node is in group G_i, then remove this node and the nodes activated by the nodes from G_i. This loses the number of activated nodes by at most $O(n) + m_8 \leq 2m_8$. The total number of activated nodes lost by this rule is at most $k(O(n)+m_8) \leq 2km_8$ since we select at most k nodes to start the activation.

2) For each selected node, which is activated by a node v_j, replace it by v_j. This does not decrease the number of activated node.

3) For each selected node v_j, replace it by a u_i with $v_j \in S_i$. This does not decrease the number of activated nodes.

Finally, we only have the nodes in $\{u_1, \cdots, u_m\}$ to be selected to start the process of activation. The total number of nodes lost activation is at most $2km_8$. By inequality (17), we have the following equation:

$$|B| - |A_{1,7}| \geq 50km_8 > 2km_8. \quad (18)$$

This implies that some nodes in Phase 8 are activated. Therefore, we have a solution for the Set Cover problem with k subsets. ■

V. Approximation for the Least Seed Set Problem

In this section, we consider the Least Seed Set (LSS) problem in the case that a person can be activated by anyone of its neighbors. This problem was proposed by Ning Chen in [13].

Least Seed Set Problem in the One-Activate-One Model: Let G be a directed graph and T be a given set of nodes need to be activated. The LSS problem is to select the least of nodes as seed set so that all the nodes in T will be activated.

Theorem 6. *There is a polynomial time $O(\log n)$-approximation algorithm for the LSS problem in the One-Activate-One Model.*

Proof: The proof of Thm. 4 shows that the LSS problem can be converted to a Set Cover problem. Therefore, it has a polynomial time approximation with $O(\log n)$ factor. ■

Theorem 7. *If the One-Activate-One LSS problem has an approximation algorithm with ratio $d(n)$, then the maximum coverage problem has an approximation algorithm with ratio $d(n)$.*

Proof: Assume $S_1, \cdots, S_m$ is the list of sets for the Set Cover problem and $S_1 \cup S_2 \cup \cdots \cup S_m = \{a_1, \cdots, a_n\}$. We can construct a social network as follows.

For each set S_i, create a vertex u_i. For each a_j, create a vertex v_j, add directed edges from u_i to v_j if $a_j \in S_i$. An edge from u_i to v_j means v_j can be activated by u_i. Let T be the set of vertices $\{v_1, \cdots, v_n\}$.

The Set Cover problem is converted into an One-Activate-One LSS problem. It is easy to see the One-Activate-One LSS problem has a $d(n)$-approximation if and only if the Set Cover problem has a $d(n)$-approximation. ■

Corollary 2. *There is no polynomial time $o(\log n)$-approximation for the LSS unless $P = NP$.*

Proof: It follows from the Thm. 7 and the well known inapproximability of the Set Cover problem. ■

VI. Conclusion and Further Research

In this paper, we show that the *deterministic linear threshold model* has no polynomial time $n^{1-\epsilon}$-approximation unless P=NP even in the simple case that one person needs at most two active neighbors to become active. In the case that a person can be activated after one of its neighbors become active, there is a polynomial time $\frac{e}{e-1}$-approximation, and we prove it is the best possible approximation under a reasonable assumption in the complexity theory. We also show that there is a $O(\log n)$-approximation for LSS problem in the One-Activate-One model.

The general IM problem in the *deterministic linear threshold model* looks very hard, but the Influence Computation problem under this model can be solved in linear time. Therefore, we can back up to some simple cases to further study this problem. The following open problem will be considered in our future research:

1) Is there a polynomial time $O(\log n)$-approximation for the LSS problem in degree bounded graphs under the *deterministic linear threshold model*.

2) Is there a polynomial time $O(\log n)$-approximation for the LSS problem in the *linear threshold model* and the *independent cascaded model*.

VII. Acknowledgment

This research work is supported in part by National Science Foundation of USA under grants CNS 1016320 and CCF 0829993.

References

[1] W. Chen, Y. Yuan and L. Zhang: Scalable Influence Maximization in Social Networks under the Linear Threshold Model. *the 2010 International Conference on Data Mining*, 2010.

[2] W. Chen, C. Wang and Y. Wang: Scalable Influence Maximization for Prevalent Viral Marketing in Large-scale Social Networks. *the 2010 ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2010.

[3] U. Feige:A Threshold of ln n for Approximating Set Cover. *Journal of the ACM*, 45: pp. 314–318, 1998.

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

[5] J. Goldenberg, B. Libai and E. Muller: Using Complex Systems Analysis to Advance Marketing Theory Development. *Academy of Marketing Science Review*, 2001.

[6] D. Kempe, J. Kleinberg É. Tardos: Maximizing The Spread of Infuence Through a Social Network. *the 2003 International Conference on Knowledge Discovery and Data Mining*, pp. 137–146, 2003.

[7] D. Kempe, J. Kleinberg and É. Tardos: Influential Nodes in a Diffusion Model for Social Networks. the 2005 International Colloquium on Automata, Languages and Programming, pp. 1127–1138, 2005.

[8] M. Richardson and P. Domingos: Mining Knowledge-Sharing Sites for Viral Marketing. *the 2002 International Conference on Knowledge Discovery and Data Mining*, pp. 61–70, 2002.

[9] F. Zou, J. Willson, Z. Zhang and W. Wu: Fast information propagation in social networks. *Discrete Mathematics, Algorithm and Applications (DMAA)*, 2010.

[10] M. Granovetter: Threshold Models of Collective Behavior. *American Journal of Sociology*, 83(6): pp. 1420–1443, 1978.

[11] T. Schelling: Micromotives and Macrobehavior. *Norton*, 1978.

[12] J. Goldenberg, B. Libai and E. Muller: Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters*, 12(3): pp. 211–223, 2001.

[13] N. Chen: On the Approximability of Influence in Social Networks. *the 2008 annual ACM SIAM symposium on Discrete algorithms*, pp. 1029–1037, 2008.