

# Influence Propagation in Adversarial Setting: How to Defeat Competition with Least Amount of Investment

Shahrazad Shirazipourazad, Brian Bogard, Harsh Vachhani, Arunabha Sen  
School of Computing, Informatics and Decision Systems Engineering  
Arizona State University  
Tempe, AZ 85287  
{sshiraz1, bbogard, hvachhan, asen}@asu.edu

Paul Horn  
Department of Mathematics  
Harvard University  
Cambridge, MA 09322  
phorn@math.harvard.edu

## ABSTRACT

It has been observed that individuals' decisions to adopt a product or innovation are often influenced by the recommendations of their friends and acquaintances. Motivated by this observation, the last few years have seen a number of studies on influence maximization in social networks. The primary goal of these studies is identification of  $k$  most influential nodes in a network. A major limitation of these studies is that they focus on a non-adversarial environment, where only one player is engaged in influencing the nodes. However, in a realistic scenario multiple players attempt to influence the nodes in a competitive fashion. The proposed model considers a competitive environment where a node that has not yet adopted an innovation, can adopt only one of the several competing innovations and once it adopts an innovation, it does not switch. The paper studies the scenario where the first player has already chosen a set of  $k$  nodes and the second player, with the knowledge of the choice of the first, attempts to identify a smallest set of nodes (excluding the ones already chosen by the first) so that when the influence propagation process ends, the number of nodes influenced by the second player is larger than the number of nodes influenced by the first.

The paper studies two propagation models and shows that in both the models, the identification of the smallest set of nodes to defeat the adversary is NP-Hard. It provides an approximation algorithm and proves that the performance bound is tight. It also presents the results of extensive experimentation using the collaboration network data. Experimental results show that the second player can easily defeat the first with this algorithm, if the first utilizes the node degree or closeness centrality based algorithms for the selection of influential nodes. The proposed algorithm also

provides better performance if the second player utilizes it instead of the greedy algorithm to maximize its influence.

## Categories and Subject Descriptors

F.2.2 [Analysis of Algorithms and Problem Complexity]: [Non-numerical Algorithms and Problems]

## General Terms

Algorithms, Experimentation, Performance

## Keywords

Social Networks, Influence Maximization, Adversarial Environment

## 1. INTRODUCTION

It has been widely observed in various studies in social sciences and economics that an individuals' decision to adopt a product, behavior or innovation is often influenced by the recommendations of their friends and acquaintances. Motivated by this observation, the last few years have seen a number of studies on influence maximization problem in social networks [2, 3, 4, 6, 7, 11, 13]. One major goal of several of these studies is identification of *k most influential nodes* in a network. A product manufacturer may want to identify the  $k$  most influential nodes in the network, as she may want to incentivize these nodes to buy the new product by providing free samples to them, on the expectation that once these nodes are convinced about the quality of the product, they will recommend it to their friends on the social network and encourage them to buy the product. This set of  $k$  nodes, being the most influential on the network, will have the largest impact on convincing the rest of the nodes about the quality of the product. Since the manufacturer has a fixed budget for advertising, she can provide free samples only to a limited number of nodes in the network. The size of the advertising budget determines the value of the parameter  $k$ .

It may be noted that most of the studies on influence propagation are geared toward a *non-adversarial environment*, where only one manufacturer (player) is attempting

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.  
Copyright 2012 ACM 978-1-4503-1156-4/12/10 ...\$15.00.

to influence the nodes of a social network to buy her product. However, in a realistic market scenario, most often there exists multiple players, each attempting to sell their competing products or innovations. For example, just as Coke attempts to convince customers in an emerging market about the quality of their beverage, its main competitor, Pepsi, also does the same. Both the competitors have only a finite advertisement budget and both of them want to derive the greatest benefit out of their advertising campaign. The goal of both the players often is to capture a share of this emerging market that is larger than its competition.

The non-adversarial influence propagation models consider scenarios where a user (a node  $u$  in a social network graph  $G = (V, E)$ ) adopts (or does not adopt) an innovation based on how her acquaintances have adopted the innovation. In these models each node  $u$  in the social network graph is in one of the following two states: (i)  $u$  has adopted innovation  $A$ , and (ii)  $u$  has not adopted innovation  $A$  but  $u$  is open to the idea of adoption. One can visualize such a scenario by coloring the nodes of the social network graph with *red* if they have adopted the innovation  $A$  and with *white* if they have not adopted  $A$  yet, but are open to the idea of adopting  $A$  in the future. As the diffusion process progresses with time, by observing changing color of the nodes of the graph one can infer if innovation  $A$  is being adopted by the members of the social network. Although, this paper focus on influence propagation in social networks, conceptually, the scenario is identical for spread of any *contagion* through a network - be it *spread of diseases* through a *human contact network* or *spread of worms* through the Internet.

The influence (contagion) propagation models can be divided into three distinct classes:

- Class I: Non-adversarial
- Class II: Adversarial with passive adversary
- Class III Adversarial with active adversary

The problems in classes I and II can be stated as follows:

- Class I: How to identify a set of  $k$  initial (seed) nodes, so that once they are *influenced/infected*, they will infect the largest number of uninfected nodes in the network?
- Class II: Given that a subset of the nodes is already *influenced/infected*, how to identify a set of  $k$  uninfected nodes, so that when they are *immunized*, they will have the largest impact in preventing the uninfected nodes from being infected.

In most of the influence propagation models, influence propagates in a step-by-step fashion and as such there is a notion of time step (or propagation step) involved. The expected number of nodes influenced at the end of time step  $D$  is at most the expected number of nodes influenced at the end of time step  $D + 1$ . In other words, expected number of nodes influenced at the end of time step  $D$  is a *non-decreasing* function of  $D$ .

The Class I influence propagation problem considered in [11] may be viewed to have three dimensions, (i) the number of seed nodes activated at the beginning (*budget or cost of influence*), (ii) the expected number of activated nodes at the end of propagation (*impact or coverage of initial seed nodes*), and (iii) *time steps* for propagation. The objective of the influence maximization problem considered in [11], is

to *maximize the coverage subject to a budget constraint but without any constraint on the number of time steps*.

The Class I problem considered in [11] can be stated in the following way: “Which  $k$  white nodes should be colored red initially, so that the largest number of white nodes turn to red at the end of propagation process?”. The Class II problems can be stated in the following way: “Given that some nodes are already colored red, which  $k$  white nodes should be colored blue, so that this set of nodes will have the largest impact in preventing the white nodes from turning red.

In Class I, there is no notion of an *adversary*. The red nodes are trying to convert all the white nodes into red nodes and there is no agent that is actively trying to prevent this conversion. The Class II, although it has a notion of an adversary (i.e., the blue nodes) which is trying to slow down (or stop) white-to-red conversion, at best this agent can be viewed as a *passive adversary*, because its goal is to prevent white-to-red conversion, and it is not engaged in white-to-blue conversion. This gives rise to Class III, a truly adversarial scenario, where the red agent is trying to convert all the white nodes into red, while the blue agent is trying to convert all the white nodes into blue. In this case, the blue agent can be viewed as an *active adversary* of the red agent.

The Class III models the scenario where a node  $u$  is being actively encouraged by an adversary not only *not to adopt the innovation* but also *to adopt a competing innovation*. In this case, each node  $u$  in the social network graph can be in one of following three states: (i)  $u$  has adopted innovation  $A$ , (ii)  $u$  has adopted innovation  $B$ , and (iii)  $u$  has not adopted any innovation  $A$  or  $B$  but is open to the idea of adopting either one of them. This adversarial scenario can be viewed as a classic case of a strategic conflict game between the proponent(s) and the opponent(s) of adoption of an innovation and a game is won by the proponent(s) if  $u$  decides to adopt the innovation  $A$ .

This paper studies a Class III scenario where two vendors (players) are trying to sell their competing products by influencing the nodes of a social network. The goal of both the players is *to have a market share that is larger than its competition*. It considers the scenario where the first player ( $P_1$ ) has already chosen the  $k$  nodes to have a large influence (coverage) on the social network. The second player is aware of the first player’s choice and the goal of the second player ( $P_2$ ) is to *identify a smallest set of nodes (excluding the ones already chosen by the first player) so that the number of nodes influenced by the second player will be larger than the number of nodes influenced by the first player within  $D$  time steps*. In other words, the objective of the problem is to *minimize the cost subject to the constraint that the coverage of the second player is larger than the coverage of the first player within  $D$  time steps*. Since the goal of the second player is to win the “game” (i.e., to have a larger coverage or market share), with influencing (incentivizing) as few nodes as possible, the problem under study in this paper is referred to as the “Winning with Minimum Investment” (WMI) problem. In [3], the authors study a similar problem belonging to class III. However, the objective of the problem studied in [3] is different from the one being studied in this paper. The goal of the second player in the problem studied in [3] is not to defeat the first player with least amount of investment, but to maximize its own influence.

Using the same two influence propagation models introduced in [3], the contributions of the paper may be listed as follows:

- Introduction of a new influence propagation problem in an adversarial setting where the goal of the second player is to defeat the first within  $D$  time steps and least amount of cost (i.e., number of seed nodes)
- NP-Hardness proof for the problem under both the influence propagation models
- Approximation algorithm for the problem with a tight performance bound.
- Experimental evaluation of the Approximation algorithm with collaboration network data

Experimental results show that utilizing the proposed algorithm, the second player can easily defeat the first, if the first player utilizes the node degree or closeness centrality based algorithms for the selection of the initial (seed) nodes. The proposed algorithm also provides better performance for the second player if she utilizes it instead of the algorithm to maximize influence proposed in [3], in the sense that it requires selection of a fewer number of seed nodes to defeat the first player.

The rest of the paper is organized as follows. The section II summarizes related work on influence propagation. The section III describes the propagation models used in the paper in detail. The sections IV, V and VI discuss the problem statement, computational complexity and approximation algorithm results respectively. The results of experimental evaluation is presented in section VII and section VIII concludes the paper.

## 2. BACKGROUND AND RELATED WORK

The studies on identification of influential nodes in a social network were triggered by a paper authored by Domingos and Richardson [6]. They introduced the notion of “network value” of a node in a social network and using a Markov random field model where a joint distribution over all node behavior is specified, computed the network value of the nodes. Kempe, Kleinberg and Tardos followed up the work in [6] by providing new models derived from mathematical sociology and interacting particle systems [11]. They made a number of important contributions by providing approximation algorithms for maximizing the spread of influence in these models by utilizing the *submodularity* property of the objective functions. In addition to providing algorithms with provable performance guarantee, they also presented experimental results on large collaboration networks. Their experimental results showed that their greedy approximation algorithm significantly out-performed the node selection heuristics based on *degree centrality* and *distance centrality* [18].

The approximation algorithm proposed in [11] is computationally intensive. Accordingly, several researchers approached the issue of scalability from different directions. Chen et. al. in [4] provided improvement of the original greedy algorithm of [11] and proposed a *degree discount* heuristic to improve influence spread. Mathioudakis in [13] introduced the notion of *sparsification of influence networks* and presented an algorithm, SPINE, to compute the “backbone” of the influence network. Utilizing SPINE as a pre-processing step for the

influence maximization problem, they showed that computation on the sparsified model provided significant improvements in terms of speedup without compromising accuracy. Wang et. al. in [17] considered the influential node identification problem in a mobile social network and presented a two step process, where in the first step, communities in the social network are detected and in the second step a subset of communities is selected to identify the influential nodes. Experimental results with data from large real world mobile social network showed that their algorithm performed an order of magnitude faster than the state-of-the-art greedy algorithm for finding the top-k influential nodes. A simulated annealing (SA) based algorithm for finding the top-k influential nodes was presented in [10]. It has been reported in [10], that using data from four real networks, the SA based algorithm performed 2-3 orders of magnitude faster than the state-of-the-art greedy algorithm.

In addition to attempts to address the scalability issue of the greedy algorithm in [11], efforts on variations of the original problem formulation and also the computation model is underway in the research community. In [7] two new problem formulations are provided. In the first formulation, the goal is to minimize the cost, subject to the constraint that coverage exceeds a minimum threshold  $\nu$  without any constraint on the number of time steps. The goal of the second formulation is to minimize the number of time steps, subject to a budget constraint  $k$  and a coverage constraint  $\nu$ . For the first version of the problem, the authors provide a simple greedy algorithm and show that it provides a bicriteria approximation. For the second version, they show that even bicriteria or tricriteria approximations are hard under several conditions. In [1], the authors argue that a user (a node in the social network) may be influenced by positive recommendations from a group of friends (neighbors in the network) but that does not necessarily imply that she will adopt the product herself. However, she may pass on her positive impression about the product to another group of friends. Clearly, such a model departs from the model considered in [11]. The authors in [1] consider an “adoption maximization” problem instead of “influence maximization” problem and present both analytical and experimental results for the new problem. The authors in [12] argue that a limitation of the traditional influence analysis technique is that they only consider positive relations (agreement, trust) and ignore the negative relations (distrust, disagreement). Moreover, the traditional techniques also ignore *conformity* of people, i.e., an individual’s inclination to be influenced. The paper studies the interplay between influence and conformity of each individual and computes the influence and conformity indices of individuals. The authors in [5] suggest an alternate way of measuring the influencing capability of an individual on her peers, through the individuals *reach* within the social network for certain *actions*.

All the references discussed in the last three paragraphs pertain to the class I (non-adversarial) problems as defined in the previous section. Results on study of class II problems (adversarial with passive adversary) is presented in [8]. It focuses on identification of *blockers*, the nodes that are most effective in blocking the spread of a dynamic process through a social network, and reports that simple local measures such as the degree of a node are good indicators of its effectiveness as a blocker. The blocker identification problem has been extensively studied in the public health community, where

the goal is to stop or slow down progress of an infectious disease by immunizing a small set of key individuals in the community.

As indicated in the previous section, the WMI problem studied in this paper belongs to Class III (adversarial with active adversary). Unfortunately, there exists only a handful of studies on problems belonging to Class III. Bharathi *et al.* were one of the earliest to study a Class III problem [2]. They proposed a mathematical model for diffusion of multiple innovations in a network, an approximation algorithm with a  $(1 - 1/e)$  performance guarantee for computing the best response to an opponent’s strategy. In addition they prove that the “price of competition” of the game is at most 2. While game theoretic framework was utilized for deriving the results in [2], Carnes *et al.* used an algorithmic framework to study a Class III problem [3]. Their research primarily extends the problem studied in [11] from the Class I domain to the Class III domain. They study the follower’s perspective (i.e., the player who entered the market after the first player) and investigate how a follower can maximize her influence in the network with a limited budget, given that the first player has already entered the market and influenced a certain number of key individuals (nodes in the network). They prove that the influence maximization problem for the second player is NP-complete and provide an approximation algorithm that is guaranteed to produce a solution within 63% of the optimal. Adversarial models in evolutionary game dynamics was studied by Istrate *et al.* in [9].

In all the problems discussed in [2, 3] once a node adopts an innovation (i.e., changes its color from white to red or white to blue), it is not allowed to change its color, i.e., the model precludes the possibility of an individual changing her mind. However, the model considered by Nowak *et al.* in [16] there are only red and blue nodes (no white nodes) and the model allows a node to change its color from red to blue and vice-versa. Although this model was developed to capture a biological phenomenon involving viruses and cells, this model can be equally effective in capturing the phenomenon of the spread of ideas and behaviors in human population. Using evolutionary game theoretic and evolutionary graph theoretic techniques, the authors establish fundamental laws that govern choices of competing players regarding strategies.

### 3. INFLUENCE PROPAGATION MODELS

A number of influence propagation models for the non-adversarial scenario have been proposed in the literature [11]. Among these, the *Linear Threshold Model (LTM)* and the *Independent Cascade Model (ICM)* have drawn most attention in the research community. As indicated earlier, the literature on influence propagation in adversarial scenario with active adversaries is very sparse [2, 3]. Bharati *et al.* in [2] and Carnes *et al.* in [3] have studied influence propagation in adversarial scenario with active adversaries, and have proposed two different models for it. Both of these two models are generalizations of the Independent Cascade Model. The model proposed in [2] is suitable for a multi-player scenario, whereas the model proposed in [3] is for two competing players. Bharati *et al.* in [2] study the problem from a game-theoretic perspective and focus on finding best response strategies for the players. Carnes *et al.* on the other hand study the problem from an algorithmic per-

spective. Since this paper studies the problem with only two competing players, the models proposed in [3] are more relevant for this study than the one proposed in [2]. Accordingly, the influence propagation models of [3] are used here. Since these models, *Distance-based Model (DBM)* and *Wave-propagation Model (WPM)*, are generalization of the ICM, the paper first discusses ICM and then DBM and WPM.

#### 3.1 Independent Cascade Model

The social network is modeled as a graph  $G = (V, E)$ , where each node represents an individual. Each individual may either be *active* (i.e., has adopted innovation) or *inactive*. A node can switch from an inactive state to an active state but cannot switch back in the other direction. The propagation process from the perspective of an inactive node  $v \in V$  can be described in the following way: With passage of time, more and more of  $v$ ’s neighbors become active and this may cause  $v$  to become active at some time step. The activation of  $v$  in turn may trigger activation of some of  $v$ ’s inactive neighbors. In the ICM model there exists a set of nodes  $V' \subset V$  that are active (seed nodes) initially and the rest of the nodes are inactive. Influence propagation unfolds in discrete steps following a randomized process. When a node  $v$  first becomes active in time step  $d$ , it has a single chance to activate each of its inactive neighbors  $w$  with probability  $p_{v,w}$  at time step  $d + 1$ . If  $v$  succeeds,  $w$  become active at  $d + 1$ . However, if  $v$  fails, it doesn’t get another chance to turn  $w$  active. The process of conversion of nodes from the inactive to the active state continues, till no further activation is possible. Since  $v$  influences  $w$  with probability  $p_{v,w}$ , the  $v - w$  edge is considered active with probability  $p_{v,w}$ . The set of active edges is denoted by  $E_a$ .

#### 3.2 Generalized ICM for Adversarial Scenario

The ICM can be adapted to handle adversarial scenario by allowing the nodes to be in one of the following three states - (i) active by adopting innovation  $A$ , (ii) active by adopting innovation  $B$ , and (iii) inactive. We use the notation  $I_A$  and  $I_B$  to indicate the initial adopters (seed nodes) of technologies  $A$  and  $B$  respectively. The nodes in the set  $V - (I_A \cup I_B)$  are the nodes that are inactive initially. The sets  $I_A$  and  $I_B$  are disjoint, i.e.,  $I_A \cap I_B = \emptyset$ . Just as in ICM, an active node  $v$  may influence each one of its inactive neighbors  $w$  with probability  $p_{v,w}$ . However, in an adversarial scenario, an inactive node  $w$ , may be in a situation where one of its active neighbor  $v$  attempts to influence  $w$  with innovation  $A$ , whereas another active neighbor  $u$  attempts to influence  $w$  with innovation  $B$ . In order to deal with this situation, the authors in [3] proposed two new models - (i) Distance-based Model, and (ii) Wave-propagation Model. The models specify the probability with which the node  $w$  will be influenced, when its active neighbors attempt to influence  $w$  with two competing technologies. The GICM operates on a random subgraph of the social network graph  $G = (V, E)$ , where each edge is included independently with probability  $p_{v,w}$ . The details of these two models are described in the following two subsections.

#### 3.3 Distance-based Model

Suppose that  $d_u(I, E_a)$  denotes the shortest path distance from the node  $u$  to the node set  $I$  where  $I = I_A \cup I_B$  along the active edges in the edge set  $E_a$ . If  $u$  is not connected to any node of  $I$  using only the active edges  $E_a$ , then  $d_u(I, E_a) =$

$\infty$ . Let  $\nu_u(I_A, d_u(I, E_a))$  and  $\nu_u(I_B, d_u(I, E_a))$  be the number of nodes in  $I_A$  and  $I_B$  respectively, at distance  $d_u(I, E_a)$  from  $u$  along edges in  $E_a$ . The probability that node  $u$  adopts innovation  $i \in \{A, B\}$  when maximum number of propagation steps is  $D$  is denoted by  $P_i(u|I_A, I_B, E_a, D)$  and is computed in the following way:

if  $d_u(I, E_a) \leq D$ ,

$$P_i(u|I_A, I_B, E_a, D) = \frac{\nu_u(I_i, d_u(I, E_a))}{\nu_u(I_A, d_u(I, E_a)) + \nu_u(I_B, d_u(I, E_a))};$$

otherwise, it is zero.

In this model the expected number of nodes which adopt  $i \in \{A, B\}$  will be computed in the following way:

$$\sigma_j(I_A, I_B, D) = \mathbb{E} \left[ \sum_{u \in V} P_i(u|I_A, I_B, E_a, D) \right]$$

where  $j = 1$  if  $i = A$ ; else  $j = 2$  and the expectation is over the set of active edges.

### 3.4 Wave-propagation Model

In this model, in step  $d < D$  all nodes that are at distance  $d - 1$  from some node in  $I$  have adopted technology  $A$  or  $B$  and all nodes that are farther than  $d - 1$  from  $I$  have not adopted any technology yet (where the distance is measured with respect to active edges). Every node at distance  $d$  from  $I$  chooses one of its neighbors at distance  $d - 1$  from  $I$  independently at random and adopt the same technology as its neighbor. For every node  $u$ ,  $S$  denotes the set of neighbors of  $u$  that are closer to  $I$  than  $u$ ; i.e., their distance from  $I$  is  $d_u(I, E_a) - 1$ . In this model  $P_i(u|I_A, I_B, E_a, D)$ , the probability that node  $u$  adopts innovation  $i \in \{A, B\}$  in at most  $D$  steps, is computed as follows:

If  $d_u(I, E_a) \leq D$ ,

$$P_i(u|I_A, I_B, E_a, D) = \frac{\sum_{v \in S} P_i(v|I_A, I_B, E_a, D)}{|S|};$$

otherwise, it is zero.

In this model the expected number of nodes which adopt  $i \in \{A, B\}$  will be computed in the following way:

$$\sigma_j(I_A, I_B, D) = \mathbb{E} \left[ \sum_{u \in V} P_i(u|I_A, I_B, E_a, D) \right]$$

where  $j = 1$  if  $i = A$ ; else  $j = 2$  and the expectation is over the set of active edges.

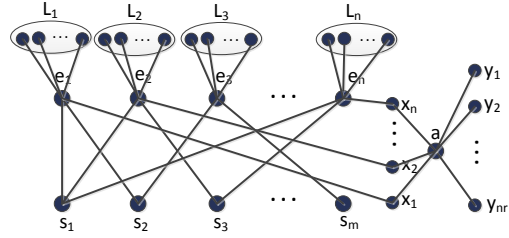
## 4. PROBLEM STATEMENT

The WMI problem can be stated informally as follows: Given a diffusion model and the information that a subset of network nodes  $I_A$  have already adopted innovation  $A$  marketed by player  $P_1$ , what is the fewest number of nodes should player  $P_2$  (marketing innovation  $B$ ) target so that by the end of  $D$  time steps, the number of nodes that adopt innovation  $B$  will exceed the number of nodes that adopt innovation  $A$ ? If  $\sigma_1(I_A, I_B, D)$  and  $\sigma_2(I_A, I_B, D)$  denote the expected number of nodes that adopt innovations  $A$  and  $B$  respectively within  $D$  time steps, the objective of the WMI problem is to

$$\begin{aligned} & \text{minimize } |I_B| \\ & \text{subject to } \sigma_2(I_A, I_B, D) > \sigma_1(I_A, I_B, D) \end{aligned}$$

## 5. COMPUTATIONAL COMPLEXITY

In this section, we prove that WMI problem is NP-hard for both propagation models.



**Figure 1: Graph  $G = (V, E)$  of WMI instance in set cover reduction**

### 5.1 Distance-based Model

Decision version of WMI: Is there a set  $I_B$  where  $|I_B| \leq M$  and  $\sigma_2(I_A, I_B, D) > \sigma_1(I_A, I_B, D)$ ?

**THEOREM 1.** *WMI is NP-hard for the distance-based model.*

*Proof:* In order to prove that WMI is NP-hard when diffusion is based on distance based model, we reduce the NP-complete *Set Cover* problem to WMI. The decision version of the *Set Cover* problem is defined in the following way: A ground set of elements  $S = \{e_1, e_2, \dots, e_n\}$ , a collection of sets  $C = \{s_1, s_2, \dots, s_m\}$  such that  $s_i \subseteq S$  and a positive integer  $K \leq |C|$  are given. The question is whether there exists a collection  $Q \subseteq C$  that covers all the elements in  $S$  and  $|Q| \leq K$ .

Given an instance of set cover problem we construct an instance of WMI. We compute  $G = (V, E)$  in the following way. For every element  $e_i \in S$  we add a node  $e_i$  and for every set  $s_j \in C$  we add a node  $s_j$  to  $V$ . We add an edge  $(e_i, s_j)$  to  $E$  for every  $e_i$  and  $s_j$  if  $e_i \in s_j$ . Also, we add a node  $a$  and nodes  $x_1, \dots, x_n$  to  $V$ . Then, for every  $e_i$  we add edges  $(a, x_i)$  and  $(x_i, e_i)$  to  $E$ . Moreover, for every  $e_i$  we add a set of  $r$  nodes,  $L_i = \{l_{i,j} | 1 \leq j \leq r\}$  to  $V$  and we connect them directly to  $e_i$ . We identify the value of  $r$  later in the proof. Finally, we add  $n \times r$  additional nodes,  $y_1, \dots, y_{n \times r}$ , to  $V$  and edges  $(y_t, a), 1 \leq t \leq n \times r$  (Fig. 1). We consider that all edges are active; i.e.,  $p_{u,v} = 1$  for all edges in  $E$ . We assign  $D = 4$  equal to the diameter of the graph  $G$ ,  $M = K$  and  $I_A = \{a\}$ .

Now, we show that the set cover problem has a solution if and only if there is a set  $I_B \subseteq V - I_A$  such that  $|I_B| \leq M$  and  $\sigma_2(I_A, I_B, D) > \sigma_1(I_A, I_B, D)$ . First we consider that there is a collection  $Q \subseteq C$  that covers  $S$  and  $|Q| \leq K$ . Then  $I_B$  includes all nodes  $s_j$  corresponding to the sets in  $Q$ . In this case, all  $e_i$  will be at distance one from  $I_B$  and two from  $I_A$ . So, all  $e_i$  and the nodes in  $L_i$  will adopt  $I_B$  with probability one. Moreover, the nodes  $s_j \notin I_B$  are two hops away from  $I_B$  while 3 hops away from  $I_A$ . Hence, all nodes  $s_j$  will adopt  $I_B$ . Therefore, we have  $\sigma_2(I_A, I_B, D) = m + n(1 + r)$ ; so,  $\sigma_2(I_A, I_B, D) > \sigma_1(I_A, I_B, D)$ .

Next, we show that if there is no collection  $Q$  of size  $K$  that covers all elements then there is no set  $I_B \subseteq V - I_A$  of size  $M$  where  $\sigma_2(I_A, I_B, D) > \sigma_1(I_A, I_B, D)$ . Considering that set cover does not have a solution, there should be at least one  $e_i$  whose distance from  $I_B$  cannot be one; so, there is an  $e_i$  and consequently nodes in  $L_i$  that choose  $A$  with the probability at least  $\frac{1}{K+1}$  and the probability that they choose  $B$  is at most  $\frac{K}{K+1}$ . Also, at most  $K$  nodes from  $x_1, \dots, x_n$  can be at distance less than or equal to 1 from  $I_B$ . Hence  $n - K$  of them will adopt  $A$  with probability one. Therefore, we have

$\sigma_2(I_A, I_B, D) \leq m + (n-1)(1+r) + \frac{K}{K+1}(r+1) + K$  and  $\sigma_1(I_A, I_B, D) \geq 1 + nr + n - K + \frac{1}{K+1}(r+1)$ . We choose  $r$  in our instance large enough such that  $r > \frac{(m+2K-2)(K+1)+K-1}{2}$ . Then we have  $1 + nr + n - K + \frac{1}{K+1}(r+1) > m + (n-1)(1+r) + \frac{K}{K+1}(r+1) + K$ ; so  $\sigma_2(I_A, I_B, D) < \sigma_1(I_A, I_B, D)$ .

## 5.2 Wave Propagation Model

**THEOREM 2.** *WMI is NP-hard for the wave propagation model.*

*Proof:* Similar to Theorem 1, we reduce decision version of *Set Cover* problem to decision version of *WMI* when wave propagation model is used for diffusion. We construct an instance of *WMI* in the same way as in Theorem 1. The only change that should be made to this instance is the value of  $r$  which will be computed later.

We need to show that the set cover problem has a solution if and only if there is a set  $I_B \subseteq V - I_A$  such that  $|I_B| \leq M$  and  $\sigma_2(I_A, I_B, D) > \sigma_1(I_A, I_B, D)$ . First we consider that there is a collection  $Q \subseteq C$  that covers  $S$  and  $|Q| \leq K$ . Then  $I_B$  includes all nodes  $s_j$  corresponding to the sets in  $Q$ . Similar to the proof of Theorem 1 we have  $\sigma_2(I_A, I_B, D) = m + n(1+r)$ ; so,  $\sigma_2(I_A, I_B, D) > \sigma_1(I_A, I_B, D)$ .

Next, we show that if there is no collection  $Q$  of size  $K$  that covers all elements then there is no set  $I_B \subseteq V - I_A$  of size  $M$  where  $\sigma_2(I_A, I_B, D) > \sigma_1(I_A, I_B, D)$ . Considering the construction of  $G$  and the fact that set cover does not have a solution, there should be at least one  $e_i$  whose distance from  $I_B$  cannot be one or smaller. Since the node  $x_i$  connected to this  $e_i$  will have probability 1 to accept  $A$  and the maximum number of nodes in first hop neighborhood of  $e_i$  that are at distance one from  $I_A \cup I_B$  is  $m+1$ , there is an  $e_i$  and consequently nodes in  $L_i$  that choose  $A$  with the probability at least  $\frac{1}{m+1}$  and the probability that they choose  $B$  is at most  $\frac{m}{m+1}$ . Also, at most  $K$  nodes from  $x_1, \dots, x_n$  or  $y_1, \dots, y_{n \times r}$  can be at distance less than or equal to 1 from  $I_B$ . Hence  $n(r+1) - K$  of them will adopt  $A$  with probability one. Therefore, we have  $\sigma_2(I_A, I_B, D) \leq m + (n-1)(1+r) + \frac{m}{m+1}(r+1) + K$  and  $\sigma_1(I_A, I_B, D) \geq 1 + n(r+1) - K + \frac{1}{m+1}(r+1)$ . We choose  $r$  in our instance large enough such that  $r > \frac{m^2}{2} + K(m+1) - \frac{3}{2}$ . Then we have  $1 + n(r+1) - K + \frac{1}{m+1}(r+1) > m + (n-1)(1+r) + \frac{m}{m+1}(r+1) + K$ ; so  $\sigma_2(I_A, I_B, D) < \sigma_1(I_A, I_B, D)$ .

## 6. APPROXIMATION ALGORITHM

Since we proved that finding the optimal solution for *WMI* is hard, in this section we propose a greedy algorithm called *GWMI*. In this algorithm either of the two propagation models discussed before can be used as the diffusion process.

Let  $\omega(I_A, I_B, D)$  be  $(\sigma_2(I_A, I_B, D) - \sigma_1(I_A, I_B, D))$ . We define  $F_i$  to denote the amount of increase in the value of  $\omega$  when node  $i$  is added to  $I_B$ ; i.e.,  $F_i = \omega(I_A, I_B \cup \{i\}, D) - \omega(I_A, I_B, D)$ . Initially  $I_B$  is empty. Hence,  $\omega(I_A, I_B, D) \leq 0$ . The algorithm executes through iterations and in each iteration node  $i \in V - I_A$  with the maximum  $F_i$  is selected. The steps of the algorithm *GWMI* has been shown in Algorithm 1.

---

### Algorithm 1 GWMI

---

Input:  $G = (V, E)$ ,  $I_A, D$

Output:  $I_B$

```

1: while  $\omega(I_A, I_B, D) \leq 0$  do
2:   for every node  $i \in V - (I_A \cup I_B)$  do
3:     Compute  $F_i$ 
4:   end for
5:   Select node  $j$  with maximum  $F_j$ 
6:    $I_B = I_B \cup \{j\}$ 
7: end while
8: return  $I_B$ 

```

---

In [11], it is mentioned that computing the exact value of  $\sigma_1(I_A, \emptyset, D)$  efficiently is an open question. Similarly, there is no known way to compute  $\sigma_1(I_A, I_B, D)$ ,  $\sigma_2(I_A, I_B, D)$  in both propagation models efficiently. However, by sampling the active sets we can get a close approximation with high probability. Given  $I_A, I_B$  and a set of active edges  $E_a$ , computation of  $\sigma_1$  and  $\sigma_2$  in both propagation models has  $O(n^3)$  time complexity since it needs computation of single all-pairs shortest paths. Given  $I_A, I_B$  and input graph  $G$ , using sampling, we can then approximate  $\sigma_1$  and  $\sigma_2$  to within  $(1+\gamma)$  for any  $\gamma > 0$  where the running time depends on  $1/\gamma$  [3].

## 6.1 Upper Bound Computation

**THEOREM 3.** *GWMI has a log n approximation ratio.*

*Proof.* Let  $I_B^t$  be the set of  $B$ 's initial adopters selected by *GWMI* at step  $t$ . Initially,  $I_B$  is empty and  $\omega(I_A, I_B^0, D) = -\sigma_1(I_A, \emptyset, D)$ . In every iteration  $t$ , the nodes in the optimal set of  $B$ 's initial adopters,  $I_B^{opt}$ , will make  $\omega(I_A, I_B^{t-1} \cup I_B^{opt}, D)$  positive. We denote the size of  $I_B^{opt}$  by  $OPT$  and the size of the solution of *GWMI* by  $H$ . Therefore, There will be at least one node in  $V - \{I_A \cup I_B^{t-1}\}$  that increases  $\omega(I_A, I_B^{t-1}, D)$  at least by  $\frac{|\omega(I_A, I_B^{t-1}, D)|}{OPT}$ . Let,  $v_t$  be the node selected by *GWMI* at iteration  $t$ . Then,  $F_{v_t} \geq \frac{|\omega(I_A, I_B^{t-1}, D)|}{OPT}$ . Therefore, for  $t < H$  we have

$$|\omega(I_A, I_B^t, D)| \leq |\omega(I_A, I_B^{t-1}, D)| - \frac{|\omega(I_A, I_B^{t-1}, D)|}{OPT}$$

$$\leq |\omega(I_A, I_B^0, D)| \left(1 - \frac{1}{OPT}\right)^t$$

Also, we know that  $|\omega(I_A, I_B^0, D)| = \sigma_1(I_A, \emptyset, D) \leq n$ . Hence we have

$$|\omega(I_A, I_B^t, D)| \leq n \left(1 - \frac{1}{OPT}\right)^t \leq ne^{-\frac{t}{OPT}}$$

Since adding a node to  $I_B$  will increase  $\omega(I_A, I_B, D)$  at least by one, we need to find the smallest  $t$  that  $|\omega(I_A, I_B^t, D)| < 1$ . Then adding at most one more node will make  $\omega(I_A, I_B, D)$  positive. Therefore,  $H \leq 1 + OPT \ln n$ . We note that this proof holds for both propagation models.

## 6.2 Lower Bound Computation

We now give a construction giving the lower bound for *GWMI* when distance-based propagation model is used. Let  $X$  and  $Y$  be disjoint sets of  $\frac{n}{2}$  vertices and  $G(n, 3/4)$  be the Erdős-Renyi random graph on  $X \cup Y$  with  $p = 3/4$ .

We take two new vertices  $u$  and  $v$ , connect  $u$  to all vertices of  $X$  and  $v$  to all vertices of  $Y$ . Now, we add a disjoint star  $S$  with  $n+2$  leaves and connect the center of the star to  $u$  and  $v$ . This yields our graph  $G$  (Fig. 2).

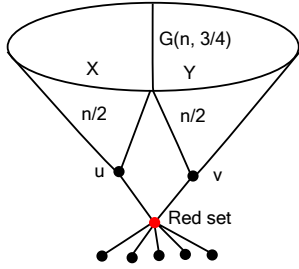


Figure 2: Construction of  $G$ .

We consider that the center of the star is the only initial adopter of  $A$  (red node), and  $p_{u,v}$  is uniform and it is 1 for all the edges of  $G$  and  $D = 3$ . An optimal set of initial adopters of  $B$  (initial blue nodes) includes  $u, v$  and any of the leaves of  $S$ . We claim that the greedy algorithm  $GWMI$  will select  $\Omega(\log n)$  vertices with high probability, assuming  $n$  is large enough.

In order to prove this we first state a technical lemma giving a condition that  $G$  satisfies with high probability. Let  $S \subseteq X \cup Y$ . We say  $S$  is *fair* if

$$1. |X \setminus \partial(S)| = (1/4)^{|S|} \frac{n}{2} + O(n^{3/4}) \text{ and } |Y \setminus \partial(S)| = (1/4)^{|S|} \frac{n}{2} + O(n^{3/4}).$$

where  $\partial(S)$  is the set of one hop neighbors of vertices in  $S$ .

We claim the following lemma, whose proof we defer:

LEMMA 4. *With probability  $1 - o(1)$  every set  $S \subset X \cup Y$  with  $|S| < \frac{1}{100} \ln(n)$  is fair. Furthermore, the induced graph on  $X \cup Y$  has diameter 2, every vertex in  $Y$  is at distance at most 2 from  $u$  and every vertex in  $X$  is at distance at most 2 from  $v$ .*

Assuming Lemma 4 we prove the lower bound. In particular we prove the following: The greedy algorithm selects at least  $\frac{1}{100} \ln n$  vertices from  $X \cup Y$ . We proceed by induction. At the first step, the greedy algorithm has to choose between a vertex in  $X \cup Y$ , one of  $u$  or  $v$ , or one of the vertices in the star. Selecting a vertex in the star will cause the number of blue vertices to increase by one and red vertices to decrease, a net change of two. Selecting  $u$  (or resp.  $v$ ) will increase blue (and decrease red) by a total of  $1 + \frac{n}{2} + \frac{n}{4}$ ; since every vertex in  $X$  will be at distance 1 from a blue vertex and every vertex in  $Y$  will be at distance 2 from both  $u$  and the red vertex if  $u$  is selected. On the other hand, by fairness, if a vertex  $x$  in  $X \cup Y$  is selected; the increase in blue is at least  $\frac{3n}{4} + \frac{n}{8} + O(n^{3/4})$ ; since  $\frac{3n}{4} + O(n^{3/4})$  vertices are at distance 1 from  $x$  and the other  $\frac{n}{4} + O(n^{3/4})$  are at distance 2 from both  $x$  and the red vertex. Therefore the greedy algorithm will select from  $X \cup Y$  at the first time.

Now suppose that the greedy algorithm has selected from  $X \cup Y$  a total of  $k < \frac{1}{100} \ln n$  times. Let  $B$  denote the selected set, and  $X' = X \setminus \partial(B)$  and  $Y' = Y \setminus \partial(B)$ . Every vertex in  $X' \cup Y'$  is at distance two from all  $k$  blue vertices, and hence they are currently blue with probability  $\frac{k}{k+1}$ . Furthermore by fairness  $X'$  and  $Y'$  are both of size  $(1/4)^k \frac{n}{2} + O(n^{3/4})$ .

Again, the greedy algorithm must choose: If  $u$  (or simi-

larly  $v$ ) is chosen, then increase is at most

$$1 + \frac{1}{k+1} |X'| + \frac{1}{(k+1)(k+2)} |Y'| \quad (1)$$

$$= \frac{1}{k+1} (1/4)^k \frac{n}{2} + \frac{1}{(k+1)(k+2)} (1/4)^k \frac{n}{2} + O(n^{3/4}).$$

On the other hand, if a vertex  $x$  in  $X' \cup Y'$  is chosen, the increase is at least

$$\frac{1}{k+1} |\partial(x) \cap X'| + \frac{1}{k+1} |\partial(x) \cap Y'| \quad (2)$$

$$+ \frac{1}{(k+1)(k+2)} |X' \cup Y' \setminus \partial(x)|$$

$$= 2 \cdot \frac{1}{k+1} \cdot \frac{3}{4} (1/4)^k \frac{n}{2} + \frac{1}{(k+1)(k+2)} (1/4)^{k+1} \frac{n}{2} + O(n^{3/4});$$

therefore, (2) - (1) is positive and hence the vertex in  $X' \cup Y'$  will be chosen as desired. We note that this construction is for sufficiently large  $n$  and  $(1/4)^k n \gg n^{3/4}$ .

PROOF OF LEMMA 4. Let  $S \subset X \cup Y$ , with  $|S| < \frac{1}{100} \ln n$ . Then

$$\mathbb{E}[|X \setminus \partial(S)|] = (1/4)^{|S|} (|X| - |X \cap S|) = (1/4)^{|S|} \frac{n}{2} + O(\ln n).$$

Let  $X_S = |X \setminus \partial(S)|$ . Chernoff bounds imply that

$$\mathbb{P}(|X_S - \mathbb{E}[X_S]| > n^{3/4}) \leq \exp(-\Omega(\frac{n^{3/2}}{\mathbb{E}[X_S]})) \leq \exp(-\Omega(n^{1/2})).$$

Bounds for  $|Y \setminus \partial(S)|$  follow similarly. On the other hand there are at most

$$\sum_{i=1}^{\frac{1}{100} \ln n} \binom{n}{i} \leq \frac{1}{100} \ln(n) \cdot n^{\ln n},$$

sets  $S$ . Thus union bounds imply every set is fair with probability  $1 - \exp(-\Omega(n^{1/2}))$ .

Note that the expected number of common neighbors between  $x$  and  $y$  in  $X' \cup Y'$  is  $\frac{9n}{16}$ , and Chernoff bounds plus union bounds imply every pair  $x$  and  $y$  is of distance 2 (and in fact has  $(1 - o(1)) \frac{9n}{16}$  common neighbors). Likewise,  $u$  and a vertex in  $Y$  have  $\frac{3n}{8}$  expected neighbors and Chernoff bounds imply that every pair has  $(1 + o(1)) \frac{3n}{8}$  common neighbors. Likewise, for  $v$  and vertices in  $X$ . A union bound over all events completes the proof.  $\square$

## 7. SIMULATION

In this section we evaluate the performance of our approximation algorithm,  $GWMI$ , on a real network data set. It has been suggested in [15] that the co-authorship graphs are representative of typical social networks. As such, we use the real collaboration network data set of the scientists posting preprints on the high-energy theory archive at [www.arxiv.org](http://www.arxiv.org), 1995-1999 [14]. This network has 8361 nodes (authors) and 15751 edges. The largest connected component has 5835 number of nodes (authors) and maximum distance between the nodes in a connected component is 19.

Our experiments were conducted on a high performance computer which is a 5K processor Dell Linux Cluster. The program is parallelized with OpenMP, optimized with Intel compiler and was executed on an 8 core compute node. The cores in the node have equal access to a common pool of shared memory. Each node is comprised of 2.66/2.83 GHz



processors, 8MB cache, 16GB memory and 8 cores. Since our experiments required execution of the algorithm on a large number of instantiation of a social network (the graphs were different as their set of active edges were different), we used OpenMP for parallelization of the graph instances for the simulation with one data set.

In the first set of experiments we evaluate the performance of GWMI algorithm against the results obtained from the heuristics based on node degree and closeness centrality. These heuristics are most often used in social networks to identify most influential nodes [11]. We also compare performance of GWMI with the greedy algorithm proposed in [3] for selection of seed nodes for the second player  $P_2$ . In our model the first player  $P_1$  is trying to market product  $A$  and the second player  $P_2$  is trying to market product  $B$ . Since WMI problem is NP-hard and the input data set is large, computation of the optimal solution within a reasonable amount of time is unlikely. It may be noted that there is no known way of computing the exact value of  $\sigma_1(I_A, I_B, D)$  and  $\sigma_2(I_A, I_B, D)$  efficiently [11]. Accordingly, we use sampling of the active edge sets to obtain close approximation of  $\sigma_1(I_A, I_B, D)$ ,  $\sigma_2(I_A, I_B, D)$  with high probability. As in the experiments reported in papers [11, 3], we assign the edge probabilities to be 0.1. In all the experiments we use WPM as the diffusion model.

The node degree based heuristic selects the nodes in the decreasing order of their degrees and the closeness centrality based heuristic selects the nodes in the increasing order of their average distance to other nodes. The distance between two nodes that are not in the same connected component is taken to be  $n$ , where  $n$  is number of nodes in the network. In the greedy algorithm proposed in [3], in every iteration the node that increases  $\sigma_2(I_A, I_B, D)$  the most is selected. We refer to this algorithm as *Second Player Influence Maximization* (SPIM) algorithm. In these experiments, maximum number of propagation steps is taken to be 10, i.e.,  $D = 10$ . In the experiments, the player  $P_1$  used node degree based heuristic to select its  $k$  initial adopters. In our experiments, the size of initial adopters of  $A$  (20, 40, 60, 80, 100). The results of this set of experiments using the WPM is shown in Fig. 3. The Fig. 3 shows that all five sizes of the initial adopters of  $A$  (20, 40, 60, 80, 100), the GWMI algorithm required the fewest number of initial adopters of  $B$  necessary to defeat  $A$ 's influence at the end of time step 10. The legend Degree-Degree in Fig. 3 denotes that both the players are using the node degree based heuristics to select the initial adopters. Similarly, the legend Degree-GWMI denotes that while  $P_1$  is using the node degree based heuristics to select the initial adopters,  $P_2$  is using the GWMI algorithm to do the same.

The Figs. 4 and 5 show the coverage (i.e., the number of nodes influenced at the end of 10 time steps) for players  $P_1$  and  $P_2$  respectively. Although the GWMI algorithm does not make an effort to minimize the coverage of  $P_1$ , it may be observed from the Fig. 4, the coverage of  $P_1$  is less if  $P_2$  uses GWMI instead of SPIM. Thus  $P_2$  is better off using GWMI instead of SPIM, if in addition to be able to defeat  $P_1$  with least investment (i.e., initial adopters),  $P_2$  wants to have a smaller market share for  $P_1$ . The Fig. 5 shows the coverage of  $P_2$  at the end of ten time steps. It may be observed from the Fig. 5, that at all five data points the coverage for  $P_2$  is highest when she uses the SPIM algorithm. This is not surprising as the stated goal of SPIM is to maxi-

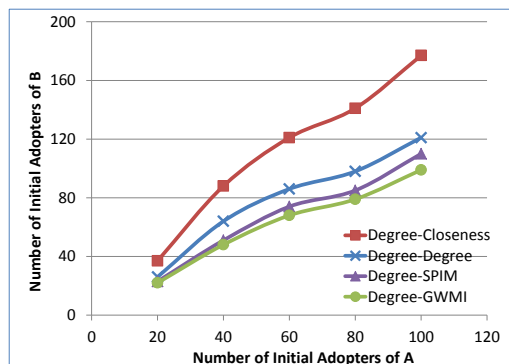


Figure 3: Number of initial adopters of  $B$  for different values of  $|I_A|$

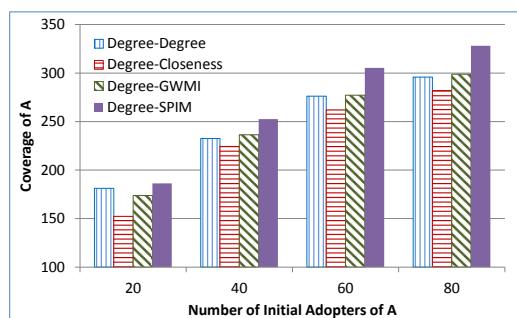


Figure 4: Expected number of nodes adopting  $A$  after 10 propagation steps

mize  $P_2$ 's coverage (influence). However, this figure may be somewhat misleading because it does not provide the information pertaining to the number of initial adopters required by the SPIM algorithm to achieve the higher coverage. By its stated objective, the number of initial adopters required by GWMI to defeat  $P_1$  cannot be higher than the the number of initial adopters required by SPIM. Once this is factored in, and we compute the coverage per initial adopter, we find that the coverage per initial adopter of the SPIM algorithm is very close to that of the GWMI algorithm. This is shown in Fig. 6.

From Fig. 3 it is clear that the node degree and centrality based heuristics and the SPIM algorithm require a larger number of initial adopters of  $B$  to beat  $A$  than is needed by the *GWMI* algorithm. While this is a negative aspect of SPIM (cost), it also has a positive aspect in the sense that at the end of ten time steps, it also secures a larger coverage for  $B$  (benefit). We compute the additional benefit provided by the additional initial adopters. Let  $I_{B(X)}$  be the smallest set of initial adopters of  $B$  that is required by algorithm  $X$  to defeat  $A$  and  $\sigma_{2(X)}$  be the expected number of nodes that adopt  $B$  after  $D$  propagation steps. Here  $X$  can be node-degree or centrality based heuristic or the SPIM algorithm. In the case,  $(\sigma_{2(X)} - \sigma_{2(GWMI)})$  indicates the additional benefit and  $(|I_{B(X)}| - |I_{B(GWMI)}|)$  indicates the additional cost. In this case,  $(\sigma_{2(X)} - \sigma_{2(GWMI)}) / (|I_{B(X)}| - |I_{B(GWMI)}|)$  indicates the average market share gain of  $B$  with each additional initial adopter when using algorithm



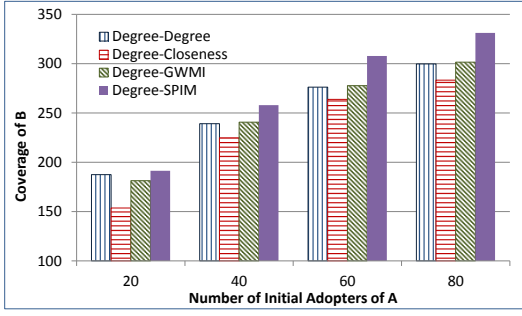


Figure 5: Expected number of nodes adopting  $B$  after 10 propagation steps

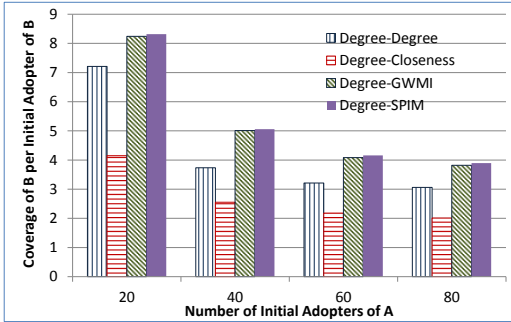


Figure 6: Expected number of nodes adopting  $B$  per initial adopter of  $B$  after 10 propagation steps

$X$ . The Fig. 7 depicts the results for the heuristics and SPIM. The negative gains are not shown. It may be observed from Fig. 7 that the average market share gain of  $B$  with each additional initial adopter diminishes with increase of the number of initial adopters of  $A$ , when it uses the SPIM algorithm.

While the stated objective of  $P_2$  is to have a larger market share than  $P_1$  with the fewest number of initial adopters, it may also have two other unstated objectives - (i) to have a large  $\sigma_2(X)$  and (ii) a small  $\sigma_1(X)$  for all  $X$  ( $\sigma_1(X)$  be the expected number of nodes that adopt  $A$  after  $D$  time steps). Therefore while considering the benefit of the additional initial adopters, we can consider not only  $(\sigma_2(X) - \sigma_2(GWMI))$

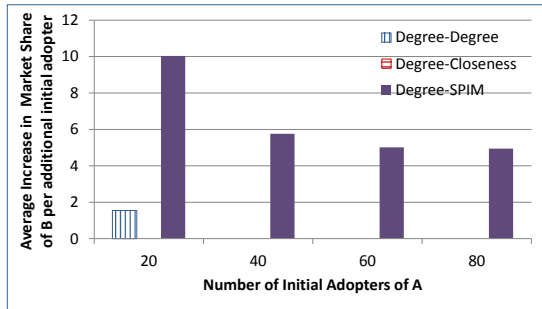


Figure 7: Average market share increase that innovation  $B$  can capture per additional initial adopter with respect to  $GWMI$

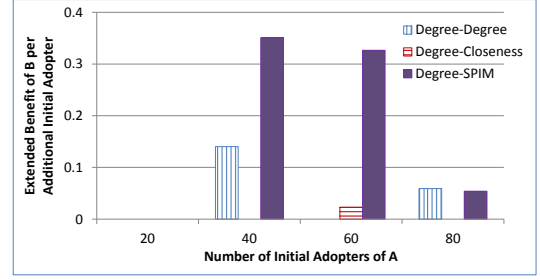


Figure 8: Extended benefit that  $B$  can capture per additional initial adopter with respect to  $GWMI$

but also  $(\sigma_1(GWMI) - \sigma_1(X))$ . It introduces a notion of *extended benefit* by combining these two factors in the following way:  $(\sigma_2(X) - \sigma_2(GWMI)) - (\sigma_1(GWMI) - \sigma_1(X))$ . With this notion of *extended benefit*,

$$\frac{((\sigma_2(X) - \sigma_2(GWMI)) + (\sigma_1(GWMI) - \sigma_1(X)))}{|I_{B(X)}| - |I_{B(GWMI)}|}$$

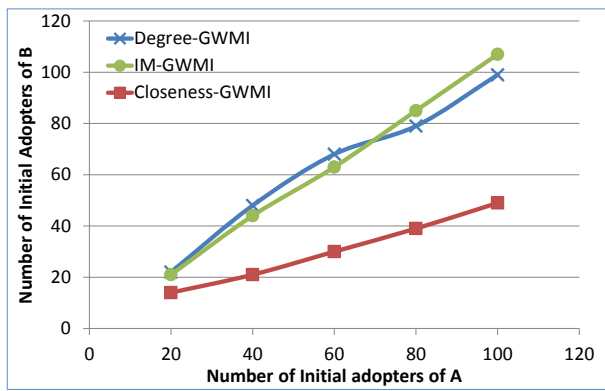
indicates the average market share gain of  $B$  with each additional initial adopter when using algorithm  $X$ . The Fig. 8 depicts the results for the heuristics and SPIM. It may be observed from Fig. 7 that when extended benefit is considered, the average market share gain of  $B$  with each additional initial adopter diminishes even more drastically with increase of the number of initial adopters of  $A$ , when it uses the SPIM algorithm. Moreover, the gain of each additional initial adopter is smaller than 1 and implies that the additional adopter is not worth its cost.

In the second set of experiments we investigate different strategies for selection of initial adopters of  $A$  when  $P_2$  uses  $GWMI$ . The strategies that we consider for selection of initial adopters of  $A$  includes the greedy algorithm proposed in [11] and heuristics based on node degree and closeness centrality. In these experiments WPM is used as diffusion model and  $D = 10$ .

Fig. 9 depicts the results of these experiments. We observe that the closeness-centrality based heuristic performs poorly in comparison to other two algorithms. This is true because the number of initial adopters of  $B$  that it needs to defeat  $A$ 's overall influence (coverage) is much smaller than the size of initial adopters of  $A$ . More specifically, for closeness-centrality based heuristic, for  $|I_A|$  values greater than 60, the number of initial adopters of  $B$  is less than 50% of  $|I_A|$ . This set of results show that if the influence maximization algorithm (IM) proposed in [11] is used for the selection of  $I_A$ , it forces  $P_2$  to select a large set for  $I_B$  in order to be able to defeat  $P_1$  within  $D$  time steps.

## 8. CONCLUSION

In this paper we have introduced a new influence propagation problem in an adversarial setting where the goal of the second player is to defeat the first within  $D$  time steps and least cost, measured in terms of the number of seed nodes. Considering two different influence propagation models, we provided the NP-Hardness proof for the problem and an approximation algorithm with a tight performance bound. In addition, we evaluated the performance of the approximation algorithm with collaboration network data.



**Figure 9: Size of initial adopters of  $B$  for different values of  $|I_A|$**

We can envisage at least two new directions of research with this problem. In the first direction,  $P_2$  is not aware of  $P_1$ 's choice. In the second direction, back and forth transition of the nodes between two competing products is allowed.

## 9. ACKNOWLEDGMENTS

The research was supported in part by a grant to the Center for the Study of Religion and Conflict at Arizona State University (N00014-09-1-0815). The award was funded through the Office of the Secretary of Defense Minerva program, and managed out of the Office of Naval Research. The content is solely the responsibility of the authors and does not necessarily represent the views of the Office of Naval Research. In addition, it was also supported in part by the DTRA grant HDTRA1-09-1-0032 and the AFOSR grant FA9550-09-1-0120.

## 10. REFERENCES

- [1] S. Bhagat, A. Goyal, and L. V. Lakshmanan. Maximizing product adoption in social networks. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 603–612, 2012.
- [2] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. In *Proceedings of the 3rd international conference on Internet and network economics*, WINE'07, pages 306–311, 2007.
- [3] T. Carnes, C. Nagarajan, S. M. Wild, and A. van Zuylen. Maximizing influence in a competitive social network: a follower's perspective. In *Proceedings of the ninth international conference on Electronic commerce*, ICEC '07, pages 351–360, 2007.
- [4] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 199–208, 2009.
- [5] K. Dave, R. Bhatt, and V. Varma. Modelling action cascades in social networks. 2011.
- [6] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '01, pages 57–66, 2001.
- [7] A. Goyal, F. Bonchi, L. V. S. Lakshmanan, and S. Venkatasubramanian. Approximation analysis of influence spread in social networks. *arXiv:1008.2005v4*, 2011.
- [8] H. Habiba, Y. Yu, T. Y. Berger-Wolf, and J. Saia. Finding spread blockers in dynamic networks. In *Proceedings of the Second international conference on Advances in social network mining and analysis*, SNAKDD'08, pages 55–76, 2010.
- [9] G. Istrate, M. V. Marathe, and S. S. Ravi. Adversarial models in evolutionary game dynamics. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, SODA '01, pages 719–720, 2001.
- [10] Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, and K. Xie. Simulated annealing based influence maximization in social networks. 2011.
- [11] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '03, pages 137–146, 2003.
- [12] H. Li, S. S. Bhowmick, and A. Sun. Casino: towards conformity-aware social influence analysis in online social networks. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, pages 1007–1012, 2011.
- [13] M. Mathioudakis, F. Bonchi, C. Castillo, A. Gionis, and A. Ukkonen. Sparsification of influence networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 529–537, 2011.
- [14] M. Newman. <http://networkdata.ics.uci.edu/data/hep-th/>.
- [15] M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):404–409, 2001.
- [16] M. A. Nowak, C. E. Tarnita, and T. Antal. Evolutionary dynamics in structured populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537):19–30, 2010.
- [17] 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 international conference on Knowledge discovery and data mining*, KDD '10, pages 1039–1048, 2010.
- [18] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Number 8 in Structural analysis in the social sciences. Cambridge University Press, 1 edition, 1994.