



## Finding a Weighted Positive Influence Dominating Set in E-learning Social Networks

Guangyuan Wang, Hua Wang, Xiaohui Tao, Ji Zhang and Guohun Zhu

Department of Maths and Computing

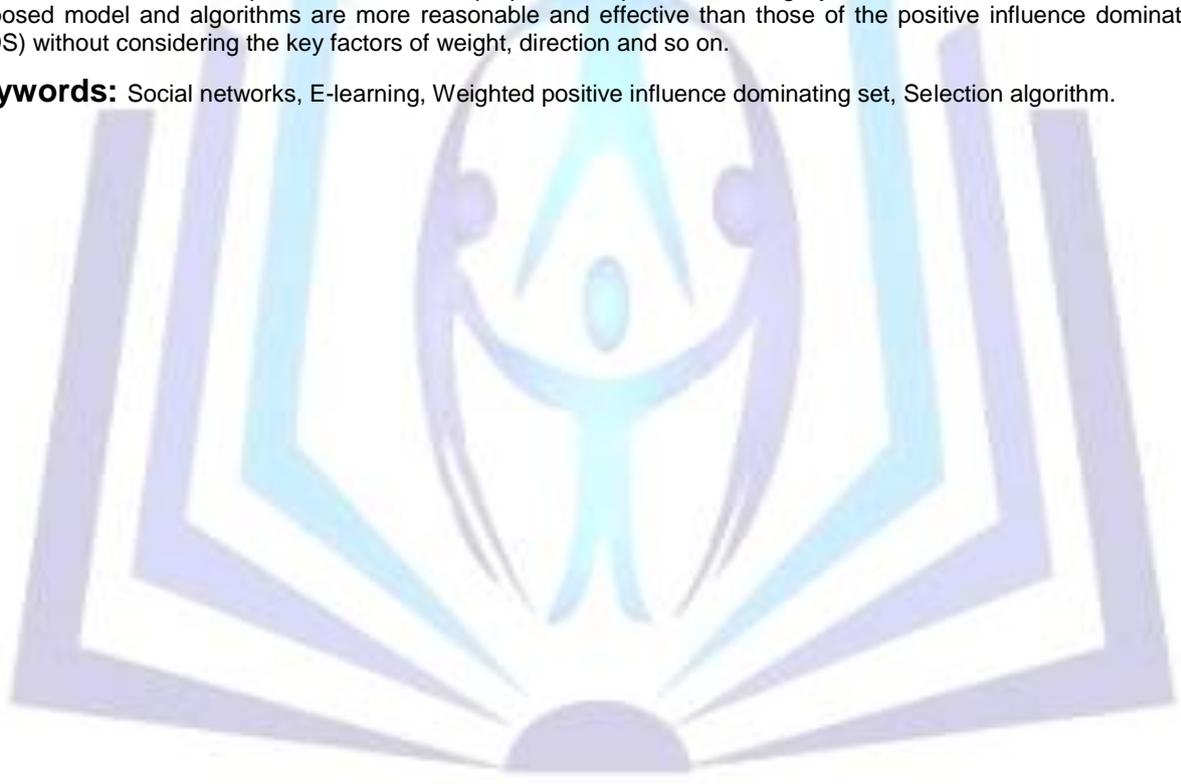
University of Southern Queensland, QLD, 4350, Australia

{guangyuan.wang,hua.wang,xiaohui.tao,ji.zhang,guohun.zhu}@usq.edu.au

### ABSTRACT

Online social network has developed significantly in recent years. Most of current research has utilized the property of online social network to spread information and ideas. Motivated by the applications of dominating set in social networks (such as e-learning), a variation of the dominating set called positive influence dominating set (PIDS) has been studied in the literature. The existing research for PIDS problem do not take into consideration the attributes, directions and degrees of personal influence. However, these factors are very important for selecting a better PIDS. For example, in a real-life e-learning community, the attributes and the degrees of their influence between a tutor and a student are different; the relationship between two e-learning users is asymmetrical. Hence, comprehensive, deep investigation of user's properties become an emerging and urgent issue. The focus of this study is on the degree and direction between e-learners' influence. A novel dominating set model called weighted positive influence dominating set (WPIDS), and two selection algorithms for the WPIDS problem have been proposed. Experiments using synthetic data sets demonstrate that the proposed model and algorithms are more reasonable and effective than those of the positive influence dominating set (PIDS) without considering the key factors of weight, direction and so on.

**Keywords:** Social networks, E-learning, Weighted positive influence dominating set, Selection algorithm.



## Council for Innovative Research

Peer Review Research Publishing System

**Journal:** INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 10, No 10

[editor@cirworld.com](mailto:editor@cirworld.com)

[www.cirworld.com](http://www.cirworld.com), [member.cirworld.com](http://member.cirworld.com)



## 1 Introduction

As the Internet becomes widespread, e-learning communities have become more and more popular [10, 11]. E-learning is an attractive and efficient way for modern education since e-learning environments are more convenient and source saving to build compared with the traditional learning environments. In such learning environments, almost all the resources are provided through the computers and networks and students can learn anytime and anywhere. The interaction and collaboration of tutors and students also play important roles in the e-learning program. Social interaction within an online framework can help students share experiences and collaborate on relevant topics.

Some research has been done to understand the properties of e-learning. Many educators and researchers have proposed their designs, described their implementation and shared their experiences from different points of view on e-learning environments [9,11,19,27]. In fact, the relationship between the e-learning users (e-learners) composes an online social network. In e-learning programs the tutors and students compose the set of users. There are some different studying groups according to their interests and purposes. The fact is that each user has different learning ability and it is very important to divide groups such that there are plenty of tutors or excellent students in each group to have positive influence to help others. For example, a user can be an authority such as a tutor who has heavy positive influence on others, an excellent student, an average student, or a poor student in terms of their academic records. An excellent student has positive influence on his direct friends (outgoing neighbors), but he might turn into a poor student and has negative influence on his outgoing neighbors if many of his friends (incoming neighbors) are poor students, and vice versa.

On the other hand, due to the financial limitations in budget, it is impossible to set lots of tutors in the study program. These issues are very intricate and complex problems that require a systemlevel approach where the dynamics of positive and negative influence resulting from individual-to-individual and from individual-to-group interactions as well as the evolving status of individuals can be fully captured. Therefore, it becomes an important research problem as to how to choose a subset of individuals to be part of the program so that the effect of the intervention program can spread

through the whole group under consideration. In an effort to address this issue, the specific problem we study in this paper is the following: given an online e-learning system and the set of users. We identify a subset of the individuals within the e-learning online social network to participate in an education/intervention program such that the education/intervention can result in a globally positive impact on the other users.

The rest of this paper is organized as follows. Section 2 describes some related work on theoretical and experimental analysis of e-learning and other related networks, followed by our motivations from the PIDS problem. In Section 3, the WPIDS problem arising from e-learning social networks is formalized. In section 4, one theorem for theoretical justification of our algorithms and two WPIDS selection algorithms are presented. In section 5, experiments are conducted on synthetic data sets to demonstrate that the proposed algorithms are efficient. Section 6 concludes this paper and discusses our future work.

## 2 Related Work

Much research has studied e-learning issues [7, 8, 23]. Garruzzo [7, 8] described their multi-agent elearning platform. The agents in tutor system can provide adaptive service by exploiting the device agents associated with the e-learning web site and the teacher agents. van Raaij and Schepers [23] built a conceptual model to explain the differences between individual students in the level of acceptance and use of a virtual learning environment in China. It indicates that perceived usefulness has a direct effect on virtual learning environment use and both personal innovativeness and computer anxiety have direct effects on perceived ease of use only. Wei et al. disclosed [28] that it will greatly improve e-learning efficiency if credible study materials can be accurately identified in the e-learning community. Hsiao et al. [11] designed a model of comparative social visualization for e-learning, which encourages information discovery and social comparisons.

On the other hand, some researches have been done to understand the e-learning online social network properties [1, 17, 18] and how to effectively utilize social networks to spread ideas and information within a group [3]. Among many exploiting researches, the relationships and influences among individuals in social networks might offer considerable benefit to both the economy and society. Domingos and Richardson [4] were the first ones to study the propagation of influence and the problem of identification of the most influential users in networks. They proposed the data mining

to viral marketing and first considered the customer's values which it may influence other customers to buy. Kempe et al. [13, 14] investigated the problem of maximizing the expected spread of an innovation or behavior within a social network based on the observation that individuals' decisions to purchase a product or adopt an innovation are strongly influenced by recommendations from their friends and acquaintances. Leskovec et al. [16] studied the influence propagation in a different perspective in which they aimed to find a set of nodes in networks to detect the spread of virus as soon as possible.

Among these research, finding a proper subset of most influential individual is formulated into a domination problem. For example, Eubank et al. [5] proposed a greedy approximation algorithm and proved that the algorithm gives a  $1 + O(1)$  approximation with a small constant in  $O(1)$  to the dominating set problem in a power-law graph. Zhu et al. [29] studied a new type of dominating set which satisfies the property that for every node not in the domination set has at least half of its neighbors which are in the dominating set. They presented results regarding the complexity and approximation in general graphs. Wang et al. [25] introduced a variation of dominating set, called positive influence dominating set (PIDS), originated from the context of influence propagation in social networks. Wang et al. [26] also proved that finding a PIDS of

minimum size is APX-hard and proposed a greedy algorithm with an approximation ratio of  $H(\Delta)$  where  $H$  is the harmonic function and  $\Delta$  is the maximum vertex degree of the graph representing a social network. Dinh et al. [21] provided tight hardness results and approximation algorithms for many existing domination problems, especially the PIDS problem and its variations.

Domination problems are all NP-Hard in general graphs [6]. More and more researchers move their attention to compute approximation solutions to domination problems [5, 22, 26]. Since finding a positive influence dominating set (PIDS) of minimum size is APX-hard [26] (APX-hardness of PIDS means that if  $NP \neq P$ , then PIDS has no PTAS (polynomial-time approximation scheme)), some greedy approximation algorithms have been proposed [25, 26], which are all limited to find approximate solutions to PIDS in large social networks. However, among these researches, they do not consider the asymmetry of their relationship, and fail to address the direction and degree of influence between their relationship [25, 26]. Another drawback in [25, 26] is they overlook the key persons' heavy influence during the procedure. In this paper, we study a typical e-learning social network and explore how to utilize e-learning networks topology properties to help e-learners to improve their achievements. Our research focuses on these reasonable factors between their relationships and aims to find a novel dominating set to positively dominate other users. Our simulation experiments and analysis indicate the effectiveness of our method.

## 2.1 Motivation and Contribution

In [25], the authors introduced the notion of positive influence dominating set (PIDS) and proposed a greedy approximation PIDS selection algorithm in 2009. Recall that  $D \subseteq V$  is a positive influence dominating set (PIDS) [25, 26] if any node  $i$  in  $V$  is dominated by at least  $\lceil d(i)/2 \rceil$  nodes (that is,  $i$  has at least  $\lceil d(i)/2 \rceil$  neighbors) in  $D$  where  $d(i)$  is the degree of node  $i$ . Note that there are two requirements for PIDS: (1) every node not in  $D$  has at least half of its neighbors in  $D$ , (2) every node in  $D$  also has at least half of its neighbors in  $D$ . Wang et al. [25] revealed that approximately 60% of the whole group under consideration needs to be chosen into the PIDS to achieve the goal that every individual in the community has more positive neighbors than negative neighbors. If we consider some key factors, such as direction and degree of each person's influence. The size of the solution for selecting a proper subset of the whole group might be smaller and the algorithm might be more effective and economical.

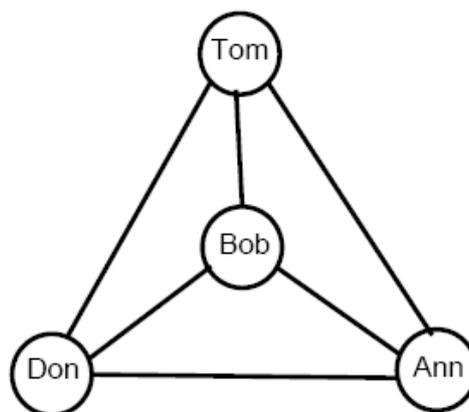


Fig. 1. An Example of PIDS Graph Model

Following the key factors as we analyzed above, Fig. 1 is a proper example which illustrates the scenario as discussed above. In Fig. 1, Bob, Tom, Don and Ann are four equal e-learners in a small leaning group. Any three of them form a PIDS satisfying its definition (conditions (1) and (2)). But if Bob is a tutor and has strong positive influence to others, only Bob can positively affect (dominate) others. In this paper we consider the degree and direction of each user's influence, propose a novel weighted positive influence dominating set (WPIDS) and develop two WPIDS selection algorithms.

The main idea of our research is as to how to effectively select positive e-learners to affect an individual in the network becomes "influence" if half of its neighbors are "positive" about adopting a product or behavior. We give both theoretical justification and empirical verification for the two proposed selection algorithms. Specifically, we prove that the feasibility of the two selection algorithms by a theorem. The contributions of this paper are as follows:

- A new weighted positive influence dominating set (WPIDS) model and two WPIDS selection algorithms have been presented. The model reasonably utilizes its online social network structure to help e-learning users to improve their achievements.
- The effectiveness of our two WPIDS selection algorithms has been evaluated by simulation experiments.
- The differences between WPIDS and PIDS models and the causes why our WPIDS model is better than PIDS model have been discussed.



### 3 Problem De\_nitions

In this section, we formulate the weighted positive influence dominating set (WPIDS) problem arising from the e-learning online social networks. We will use the following network model to illustrate the e-learning online social network in context of the improving achievement issue: A digraph  $G = (V; A; C; W)$  is used to represent the e-learning online social network.  $V$  is the set of nodes in which each node is a user in the e-learning systems.  $A$  is the set of arcs in which each direct arc represents the existence of a social connection/influence between the two endpoints.  $C$  is the compartment vector that saves the compartment of each node. The compartment of a node decides whether it has *positive* or *negative* influence on its outgoing neighbors. For example, for the improving e-learning users' achievements problem, the compartment of each node is one of the followings: authority (tutor), excellent student, average student, or poor student. A node in the authority or excellent student compartment has positive influence and all nodes in any of the other two compartments have negative influence.  $W$  is a set of weight values corresponding to arcs belong to  $A$ . Each arc's value is decided by the frequency of the two persons' interactions. In this paper, we assume that 1) if the total arcs weight of an individual's incoming neighbors has positive impact on him, then the probability that this individual positively impact others in the social network is high. 2) education/intervention programs can convert a negative influential individual to a positive influential person. 3) there are some authority users (tutors) with no incoming arcs which means that they are positive users without others' influence. Our first assumption comes from an extensive body of evidence suggesting that one of the most powerful predictors of habitual behavior

in individuals is whether an individual has friends who also engage in that behavior [2, 12, 20]. Due to outside competition in terms of personality traits attained from peer influence, the more neighbors/ friends exerting positive influence, an individual has, the more likely he is to impact others in a positive way. Our second assumption comes from the work in [2,15,24], where nearly every individual in the feedback intervention program showed an improving grade in studying. The third assumption comes from the fact that the tutors are authorities in the study program who can not be affected by other students' negative influence. With the above three assumptions, the problem is equivalent to selecting a subset of the whole e-learners in the e-learning program such that other e-learners in the system has more positive influence than negative influence. The formal definitions of the weighted positive influence dominating set (WPIDS) problem are as follows.

**De\_nition 1 (E-learner Social Network).** An E-learner Social Network is a weighted digraph  $G = (V; A; C; W)$ .  $V$  is the set of nodes in which each node is a user in the system.  $A$  is the set of arcs between the vertices:  $A = \{(u; v) | u; v \in V \text{ and the user } u \text{ influences the user } v\}$ .  $C$  is the compartment vector that saves the compartment of each node. The compartment of a node decides whether it has positive or negative influence on its outgoing neighbors.  $W$  is a set of weight values corresponding to arcs belong to  $A$ . The weight value  $W$  of an arc  $(u; v)$  is defined as:  $-w(u; v) \in [-1; 0)$ , if the user  $u$  is a negative user;  $-w(u; v) \in [0; 1]$ , if the user  $u$  is a positive user.

**De\_nition 2 (WPIDS).** With the above e-learner social network model, the weighted positive influence dominating set (WPIDS) of an e-learner online social network  $G$  is defined as finding a WPIDS  $P$  of  $V$  such that any node  $u \in V - P$  is positively dominated by  $P$ . That is,  $\forall u \in V - P; \sum_{v \in N^-(u)} w(u; v) \geq 0$ ; where  $N^-(u) = \{v | (v; u) \in A\}$  is the incoming neighbor nodes of node  $u$ .

The weighted positive influence dominating set (WPIDS) problem is to find a so-called minimum weighted positive influence dominating set (WPIDS) of  $G$ , which minimizes the total number of its vertices. Since the WPIDS problem is NP-hard, in this paper, we propose two selection algorithms for the WPIDS problem and find approximate solutions to WPIDS in large online social networks.

**Example 2.1** An example of WPIDS is shown in Fig. 2. Let node  $v_2$  represents an authority (tutor) and nodes  $v_1; v_3; v_4$  represent non-positive students, and let  $w(v_2; v_1) = w(v_2; v_3) = w(v_2; v_4) = 0.7$ , other arcs weight values are  $-0.3$ . According to the definition of WPIDS, the total incoming arcs weight values of nodes  $v_1; v_3; v_4$  are  $0.1$ . So the subset  $P = \{v_2\}$  is a WPIDS which shows the key person's influence.

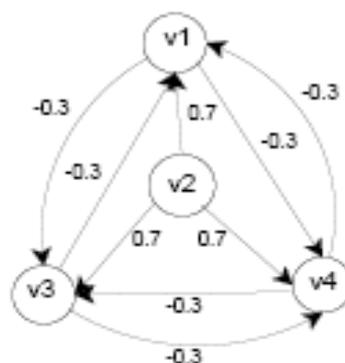


Fig. 2. An Example of WPIDS Graph Model with an Authority



## 4 WPIDS Selection Algorithm

In this section, we present one theorem and two WPIDS selection algorithms for the weighted positive dominating set problem formalized in the above section. To do so, we first define a function  $f$  as follows.

**Definition 3** Consider a weighted digraph  $G = (V; A; C; W)$  as an instance of WPI DS. For any vertex subset  $P \subseteq V$ , we define  $w(n-P(v)) = \sum_{u \in n-P(v)} w(u; v)$ ;  $f(P) = \sum_{v \in V-P} \min\{0; w(n-P(v))\}$ ; where  $n-P(v) = \{u | (u; v) \in P\}$  denotes the incoming neighbors nodes of  $v$  in  $P$ . The following theorem states important properties of the function  $f$ .

**Theorem 1.** (1)  $f(\emptyset) = 0$ .

(2)  $\forall v \in V - P; f(P) = 0$ ; if and only if  $P$  is a WPIDS of  $G$ .

(3) If  $f(P) < 0$ , then there exists a vertex  $u \in V - P$  such that  $f(P \cup \{u\}) > f(P)$ .

**Proof.** Note that  $n-\emptyset(v) = \emptyset$  for all  $v \in V$ . Therefore (1) holds.

For (2), we note that  $f(P) = 0$  if and only if  $0 \leq w(n-P(v))$  for every  $v \in V - P$  if only if  $P$  is a WPIDS.

To see (3), note that  $f(P) < 0$  implies the existence of  $v \in V - P$  such that  $0 > w(n-P(v))$ . Let  $u$  be an incoming neighbor of  $v$  which is not in  $P$  and select  $u$  into  $P$ , then  $f(P \cup \{u\}) = \sum_{v \in V} \min\{0; w(n-P(v)) + w(u; v)\} > f(P)$ .

Theorem 1 is the theoretical analysis to conduct two greedy algorithms for the WPIDS problem formalized in the earlier section. This is very important for running the algorithms in practice.

We define and explain a few terms and definitions used in the description of our algorithms. Let a weighted digraph  $G = (V; A; C; W)$  be an instance of WPIDS. Each node of  $V$  can have either positive or negative impact on its neighbor nodes. The positive degree of a node  $v$  affects an outgoing neighbor node  $u$  is the positive weight value of  $v$ 's outgoing arc weight value  $w(u; v)$ . The same holds for negative degree. The compartment  $C$  of a node decides whether the node is a positive or a negative node. For example, in the e-learning communist, a node in the tutor compartment is a positive node and a node in any other compartment is a negative node. Nodes that are chosen into the WPIDS are marked as positive nodes. Thus a e-learning user  $u$  is a *positive* user if  $u$  is initially a positive node or  $u$  is selected into the WPIDS. A PIDS  $P$  of a graph  $G$  is a subset of nodes in  $G$  that any node  $u$  in  $G$  is dominated by at least  $\lceil \frac{d(u)}{2} \rceil$  positive nodes (that is,  $u$  has at least  $\lceil \frac{d(u)}{2} \rceil$

positive neighbors) in  $P$  where  $d(u)$  is the degree of node  $u$ . A WPIDS  $P$  of a graph  $G$  is a subset of nodes in  $G$  such that any node  $u$  in  $V - P$  is positively dominated by the positive nodes in  $P$  (that is, the total influence value of  $u$ 's incoming neighbors is no less than zero).

We simply explain the following heuristic methods to obtain the two WPIDS selection algorithm. First we only need to consider the users (nodes) who are not positively dominated. It is easy to imagine that we can get a more "greedy" algorithm if we choose users (nodes) with the biggest outgoing negative influence as dominators into the positive dominating set because they can have more positive influence to others. And repeat this procedure until all persons (nodes) not in the positive influence dominating set are positively dominated by their incoming neighbors. The details

of this algorithm are presented in Algorithm 1. Considering the fact that the negative users (nodes) who have the highest accumulative weights from other neighbors' influence are easy to be educated into positive users, we propose a new algorithm.

The main idea of WPIDS Selection Algorithm 2 is to choose the users (nodes) from the negative users group with the highest accumulative weights of incoming arcs into the positive influence dominating set according to the fact that these selected persons are easier to be educated into positive users. The only difference between these two selection algorithms is that Algorithm 1 is to choose persons with the biggest outgoing negative influence as dominators and Algorithm 2 is to choose persons who are easy to change positive users as dominators. The pseudo codes of these

two WPIDS selection algorithms are described in Algorithms 1 and 2 respectively. The computing complexity of our approximation algorithms are  $O(n^2)$ . We list the complete pseudo codes of Algorithms 1 and 2 as follows.

**Example 4.1** Fig. 3 shows how to operate our two WPIDS selection algorithms. Fig. 3 is almost the same as Fig. 2 except one more negative node  $v_5$  and let  $w(v_4; v_5) = w(v_5; v_4) = -0.2$ . According to Algorithm 1, the nodes  $v_1$  and  $v_3$  have already been positively dominated by the node  $v_2$ . So we just consider the nodes  $v_4$  and  $v_5$ . The node  $v_4$  has the smallest total outgoing arcs weight value (-0.8) and the node  $v_5$  has total outgoing arcs weight values is -0.2, so the node  $v_4$  is selected as a positive node. The arc weight  $w(v_4; v_5)$  becomes 0.2 which can positively dominate the node  $v_5$ . Consequently,

the set  $\{v_2; v_4\}$  is a WPIDS which can positively dominate the whole nodes. According to Algorithm 2, The node  $v_4$  has the biggest total ingoing arcs weight value (-0.1) and the node  $v_5$  has total ingoing.

**Algorithm 1:** Weighted Positive Influence Dominating Set Selection Algorithm

**Input:** A digraph  $G = (V; A; C; W)$  where  $V$  is the set of nodes,  $A$  is the set of arcs that capture the social interaction of the nodes,  $C$  is the set of nodes that are initially in positive compartment.



**Output:** A subset  $P \cup C$  of  $V$  such that any node  $u$  in  $V - P \cup C$  has been positively dominated by  $P \cup C$ , i.e., each node's total incoming arcs weight value is non-negative.

- 1 let  $V' = V - C$  and set the status of nodes in  $C$  to POSITIVE;
- 2 initialize the status of all nodes in  $V'$  to NEGATIVE and  $P$  to empty;
- 3 calculate the sum of total incoming arcs weight of each node in  $V'$ ;
- 4  $T$  is the set of nodes in  $V'$  that have been positively dominated;
- 5 let  $V' = V' - T$ ;
- 6 while not every node in  $V'$  has been positively dominated
- 7 choose the node  $u$  with the minimum negative weight value of its total outgoing arcs in  $V'$  into the  $P$  set, set the status of nodes in  $P$  to positive; the node  $u$  has absolute weight value of outgoing arc weights;
- 8 update  $P = u \cup P$ ,  $V' = V' - T$ ;
- 9 update the positive values of each node in  $V'$ ;
- 10 end of while
- 11 the nodes in set  $V - P \cup C$  are positively dominated in  $G$ , and the positive influence of each node in  $V - P \cup C$  is also calculated.
- 12 OUTPUT: The nodes in set  $P \cup C$  positively dominate set of  $V - P \cup C$ , and the positive value of each node in  $V - P \cup C$  is also calculated.

**Algorithm 2:** Weighted Positive Influence Dominating Set Selection Algorithm

**Input:** A digraph  $G = (V; A; C; W)$  where  $V$  is the set of nodes,  $A$  is the set of arcs that capture the social interaction of the nodes,  $C$  is the set of nodes that are initially in positive compartment.

**Output:** A subset  $P \cup C$  of  $V$  such that any node  $u$  in  $V - P \cup C$  has been positively dominated by  $P \cup C$ , i.e., each node's total incoming arcs weight value is non-negative.

- 1 let  $V' = V - C$  and set the status of nodes in  $C$  to POSITIVE;
- 2 initialize the status of all nodes in  $V'$  to NEGATIVE and  $P$  to empty;
- 3 calculate the sum of total incoming arcs weight of each node in  $V'$ ;
- 4  $T$  is the set of nodes in  $V'$  that have been positively dominated;
- 5 let  $V' = V' - T$ ;
- 6 while not every node in  $V'$  has been positively dominated
- 7 choose the node  $u$  with the maximum negative weight value of its total incoming arcs in  $V'$  into the  $P$  set, set the status of nodes in  $P$  to positive; the node  $u$  has absolute weight value of outgoing arc weights;
- 8 update  $P = u \cup P$ ,  $V' = V' - T$ ;
- 9 update the positive values of each node in  $V'$ ;
- 10 end of while
- 11 the nodes in set  $V - P \cup C$  are positively dominated in  $G$ , and the positive influence of each node in  $V - P \cup C$  is also calculated.
- 12 OUTPUT: The nodes in set  $P \cup C$  positively dominate set of  $V - P \cup C$ , and the positive value of each node in  $V - P \cup C$  is also calculated.

arcs weight values is -0.2, so the node  $v_4$  is selected as a positive node. The arc weight  $w(v_4; v_5)$

becomes 0.2 which can positively dominate the node  $v_5$ . The set  $\{v_2; v_4\}$  is a WPIDS.

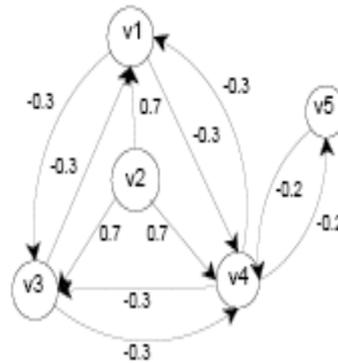


Fig. 3. An Example of WPIDS Selection Algorithm Graph Model

## 5 Evaluations

In order to clearly reveal the effectiveness of our proposed methods, we designed a simulation program to simulate e-learning environment. In this way, we can easily and clearly predefine the ground truth of each e-learner’s attribute to test the efficiency of our greedy WPIDS selection algorithms. In this section, we will discuss the experiments of WPIDS selection algorithms on data generated from a simulation program.

To evaluate the effect of WPIDS selection algorithms, we need to answer the following questions:

1) What is the difference between the size of WPIDS and PIDS and what is the difference between the influences of these two sets?

To answer the first questions, we compare the size of dominating sets [5] between PIDS and WPIDS. Generally speaking, the smaller of the size of the dominating set, the more effective and economical of the algorithm.

2) How many nodes need to be selected into the WPIDS and how influential these nodes can be?

To answer the second question, we calculate the average total incoming arcs weight of each node (called positive influence value) [28] to measure influence. The higher the average positive influence value is, the more influential of the WPIDS can be; and the possibility of the whole community turning into a positive community is higher.

3) what is the difference of the performance of our two greedy WPIDS selection algorithms in the WPIDS problem?

In order to test the third question, we compare the evaluation of these two WPIDS selection algorithms by parameter selection respectively.

### 5.1 Simulation Program Design

We conducted experimental evaluations of the proposed method on the data from simulations. Two types of users are defined in the simulation including: P type (positive users or tutors), N type (neutral users). P type of users have three characteristics as follows.

1) All P type users only have outgoing arcs which indicates they are positive users or tutors without others’ influence during the process.

2) All P type users have more outgoing arcs than N types users which indicates they are the very active users or instructor presenting distance learning courses.

3) A P type user has more heavy total influence degree than a N type user which means a tutor is the key person during the process. N type users are neutral users or students whose neighbors and influence degree are smaller than P type users and are assigned randomly.

In the designed simulation program, we simulated 300 e-learning users for 100 cycles. Among these 300 users, we used a parameter  $\alpha$  which means the percentage of all nodes initialized as positive. For example, if  $\alpha = 5\%$ ,  $10\%$ , or  $15\%$ , represent 15 P type users, 30 P type users or 45 P type users are initial definition respectively.

### 5.2 Evaluation Metrics

We used the size of dominating sets [5] as one of our evaluation parameters to evaluate the effect of WPIDS selection algorithms. The size of WPIDS is the proportion of selected e-learners in the whole e-learners group, i.e.,

$$Size = \frac{|WPIDS|}{|V|},$$



The smaller of the size of dominating sets, the more effective and economical of the algorithm. We used the average positive influence value of e-learners (APIV) [28] as the other evaluation parameter, which implies the possibility of the whole community turning into a positive community. The average positive influence value of e-learners is calculated as:

$$APIV = \frac{\sum_{v \in V-P} w(n^-(v))}{|V-P|},$$

where  $n^-(v) = \{u | (u, v) \in V\}$  denotes the incoming neighbors nodes of  $v$  in  $V$ . The high average positive influence value of e-learners can greatly increase the possibility of the whole community turning into a positive community.

### 5.3 Experimental Results

Table 1 illustrates the result of the simulation experiments for WPIDS Selection Algorithm 1. The size of the e-learning users is 300 nodes, we performed the simulation experiments with three different settings of the parameter  $\alpha = 5\%$  (15 P type users),  $10\%$  (30 P type users), and  $15\%$  (45 P type users). The result in the Table 1 is the average over 100 runs. As we can see, applying WPIDS Selection Algorithm 1, the experiment results shows that  $\alpha = 5\%$  conducts the best performance: 17.3% (52 P type users, the smallest) of the nodes are chosen into the dominating set and the resulting Average Positive Influence Value (APIV) is 0.218 (the biggest), which means more effective to positively control the community. For PIDS Selection Algorithm [25], 58% of the nodes are chosen into the dominating set, which is more than three times as big as our WPIDS Selection Algorithm 1. Another observation is that the WPIDS Selection Algorithm 1 displays a tendency that the smaller of the parameter  $\alpha$ , the better dominating set can be obtained (for both the two evaluation parameters).

Table 2 illustrates the result of the simulation program under similar settings (the same node size (300 nodes) and the similar parameter  $\alpha$  (5%, 10%, 15%) by WPIDS Selection Algorithm 2. On the contrary, it is clear shown in Table 2 that  $\alpha = 15\%$  gives the best performance: 22.7% (68 P type users, the smallest) of the nodes are chosen into the dominating set which is also much smaller than 58% (PIDS selection algorithm [25]), and the resulting Average Positive Influence Value (APIV) is 0.324 (the biggest). From the Table 2 we can also see a tendency that the bigger of the parameter  $\alpha$ , the better result can be achieved (for both the two evaluation parameters).

In summary, from Tables 1 and 2, we can see that the size of WPIDS Selection Algorithm 1 is smaller than that of WPIDS Algorithm 2 when the parameter  $\alpha$  is small ( $\alpha = 5\%$  and  $\alpha = 10\%$ ). This result meet our conjecture. Because greedy WPIDS Selection Algorithm 1 is more "greedy" than greedy WPIDS Selection Algorithm 2, and it may find a WPIDS more smaller and faster than Algorithm 2. On the other hand, the Average Positive Influence Values of WPIDS Selection Algorithm 1 are all smaller than those of WPIDS Selection Algorithm 2 respectively. This phenomenon contradicts our conjecture. Heuristically, to select persons with biggest negative influence into positive ones in Algorithm 1 can provide a bigger positive influence than to select persons with smallest negative influence in Algorithm 2. An explanation can be that it is more effective to change persons who are easily changed than to change persons who are hard to change in an education/intervention program. This interesting result needs further investigation.

**Table 1.** Result of the simulation program by WPIDS Selection Algorithm 1

P Type (percentage)	Size of WPIDS (percentage)	Average Positive Influence Value
15(5%)	52(17.3%)	0.218
30(10%)	67(22.3%)	0.212
45(15%)	79(26.3%)	0.211

**Table 2.** Result of the simulation program by WPIDS Selection Algorithm 2

P Type (percentage)	Size of WPIDS (percentage)	Average Positive Influence Value
15(5%)	76(25.3%)	0.306
30(10%)	74(24.7%)	0.324
45(15%)	68(22.7%)	0.324

### 5.4 Discussion

From the theoretical analysis in Section 4 we know the feasibility of the two proposed algorithms. Furthermore, from the experimental verification in Section 5 we can draw a conclusion that our WPIDS model is more reasonable and effective than PIDS model in papers [25, 26]. The reason is that the WPIDS in our model takes into consideration the fact that the key persons have important roles during the learning process in which the PIDS model in papers [25, 26] do not consider. In our model we consider the reasonable partition of persons who attended the program. Either in e-learning program or drinking (smoking, drug) intervention strategies and programs we should consider the authorities' effect such as tutors or



correctional officers who have heavy positive influence on other persons and they should be without interference from others as the Example 2.1 describes. Moreover, another fact is that the importance of directions. For example, the influence of their relationship between tutors and students can be considered one-way. The fact is that the degree of their influence between two persons is different and should be considered. Besides, our definition of the dominating problem is more reasonable than that of PIDS since one person's neighbors' level of influence is the decisive factor instead of the number of one's positive neighbors [25, 26]. Therefore, according to our research, in order to positively dominate the whole group we can strengthen positive influence on the key persons instead of increasing the number of positive persons [25, 26].

The algorithms we compared in this subsection are summarized in Table 3. As we can see, the PIDS selection algorithms [25, 26] can only work on undirected graph. The time complexity of Wang [25] is  $O(n^2)$ , and [26] gives a PIDS selection algorithm with an approximation ratio (AR) of  $H(\Delta)$  where  $H$  is the harmonic function and  $\Delta$  is the maximum vertex degree of the graph. Our WPIDS selection algorithms are the first work discussing PIDS on weighted directed graph. Furthermore, we give all e-learning users a reasonable partition in terms of the real-life besides the directions and degrees of their influence.

**Table 3.** Algorithms for dominating set

Reference	Output	Required topology	Greedy	Complexity
Wang et al. [25]	PIDS	Undirected graph	Yes	$O(n^2)$
Wang et al. [26]	PIDS	Undirected graph	Yes	$H(\Delta)$ AR
This paper	WPIDS	Weighted directed graph	Yes	$O(n^2)$

## 6 Conclusions and Future work

In this paper, we have introduced and studied the problem of how to utilize online social network as a medium to improve users' achievements in e-learning system. We have proposed a new dominating set model and two WPIDS selection algorithms to evaluate the effect of educating a subset of the entire target group susceptible to a social problem. The simulation experimental results have revealed that the WPIDS model and selection algorithm are more effective than those of PIDS [25, 26]. The main reason is that we consider the tutors who play important roles in the e-learning community. So the size of WPIDS is smaller than that of PIDS in an online social networks and our model is more reasonable and effective than the PIDS model [25, 26].

To deeply understand the effect of WPIDS, we will compare our WPIDS selection algorithm with the PIDS algorithm in some real-life e-learning communities in future work. Since it is very important to specify the reasonable arc weight values of the e-learning users' influence, the proper selection of parameter  $\Delta$ , and the cost of each tutor, we will also design the weighted positive influence dominating set (WPIDS) models under these factors. We can also discuss these problems with the total WPIDS case.

## References

1. A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *KDD*, pages 7–15, 2008.
2. D. J. Crandall, D. Cosley, D. P. Huttenlocher, J. M. Kleinberg, and S. Suri. Feedback effects between similarity and social influence in online communities. In *KDD*, pages 160–168, 2008.
3. D. David, J. M. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.
4. P. Domingos and M. Richardson. Mining the network value of customers. In *KDD*, pages 57–66, 2001.
5. S. Eubank, V. Kumar, M. V. Marathe, A. Srinivasan, and N. Wang. Structural and algorithmic aspects of massive social networks. In *SODA*, pages 718–727, 2004.
6. M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
7. S. Garruzzo, D. Rosaci, and G. M. L. Sarn`e. Isabel: A multi agent e-learning system that supports multiple devices. In *IAT*, pages 485–488, 2007.
8. S. Garruzzo, D. Rosaci, and G. M. L. Sarn`e. Masha-el: A multi-agent system for supporting adaptive e-learning. In *ICTAI (2)*, pages 103–110, 2007.
9. N. Ghasem-Aghaee, S. Fatahi, and T. I. " Oren. Agents with personality and emotional filters for an e-learning environment. In *SpringSim*, page 5, 2008.



10. I. Giannoukos, I. Lykourantzou, G. Mpardis, V. Nikolopoulos, V. Loumos, and E. Kayafas. Collaborative e-learning environments enhanced by wiki technologies. In *PETRA*, page 59, 2008.
11. I. H Hsiao, J. Guerra, D. Parra, F. Bakalov, B. König-Ries, and P. Brusilovsky. Comparative social visualization for personalized e-learning. In *AVI*, pages 303–307, 2012.
12. J. Jaccard, H. Blanton, and T. Dodge. Peer influences on risk behavior: analysis of the effects of a close friend. *Developmental Psychology*, 41(1):135–147, 2005.
13. D. Kempe, J. M. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD*, pages 137–146, 2003.
14. D. Kempe, J. M. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In *ICALP*, pages 1127–1138, 2005.
15. M. E. Larimer and J. M. Cronce. Identification, prevention, and treatment revisited: individual-focused college drinking prevention. *Addictive Behaviors*, 32:2439–2468, 2007.
16. J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. M. VanBriesen, and N. S. Glance. Cost-effective outbreak detection in networks. In *KDD*, pages 420–429, 2007.
17. A. Mislove, M. Marcon, P. K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Internet Measurement Conference*, pages 29–42, 2007.
18. A. Nazir, S. Raza, and C. N. Chuah. Unveiling facebook: a measurement study of social network based applications. In *Internet Measurement Conference*, pages 43–56, 2008.
19. H. C. Shi, S. Revithis, and S. S. Chen. An agent enabling personalized learning in e-learning environments. In *AAMAS*, pages 847–848, 2002.
20. J. B. Standridge, R. G. Zylstra, and S. M. Adams. Alcohol consumption: an overview of benefits and risks. *Southern Medical Journal*, 97(7):664–672, 2004.
21. M. T. Thai T. N. Dinh and D. T. Nguyen. *A Unified Approach for Domination Problems on Different Network Topologies*. Springer Publisher, (p. pardalos, d. du, and r. graham eds) edition.
22. J. A. Torkestani and M. R. Meybodi. Finding minimum weight connected dominating set in stochastic graph based on learning automata. *Information Sciences*, 200:57–77, 2012.
23. E. M. van Raaij and J. J. L. Schepers. The acceptance and use of a virtual learning environment in china. *Computers & Education*, 50(3):838–852, 2008.
24. S. T. Walters and M. E. Bennett. Addressing drinking among college students: a review of the empirical literature. *Alcoholism Treatment Quarterly*, 18(1):61–67, 2000.
25. F. Wang, E. Camacho, and K. Xu. Positive influence dominating set in online social networks. In *COCOA*, pages 313–321, 2009.
26. F. Wang, H. Du, E. Camacho, K. Xu, W. Lee, Y. Shi, and S. Shan. On positive influence dominating sets in social networks. *Theor. Comput. Sci.*, 412(3):265–269, 2011.
27. H. Wang, Y. C. Zhang, and J. L. Cao. Effective collaboration with information sharing in virtual universities. *IEEE Trans. Knowl. Data Eng.*, 21(6):840–853, 2009.
28. W. Wei, J. Ho-Man Lee, and I. King. Measuring credibility of users in an e-learning environment. In *WWW*, pages 1279–1280, 2007.
29. X. Zhu, J. Yu, W. J. Lee, D. Kim, S. Shan, and D. Z. Du. New dominating sets in social networks. *J. Global Optimization*, 48(4):633–642, 2010.