Positive Influence Dominating Set in Online Social Networks

Feng Wang, Erika Camacho, and Kuai Xu

Arizona State University
Email: {fwang25, erika.camacho, kxu01}@asu.edu

Abstract. Online social network has developed significantly in recent years as a medium of communicating, sharing and disseminating information and spreading influence. Most of current research has been on understanding the property of online social network and utilizing it to spread information and ideas. In this paper, we explored the problem of how to utilize online social networks to help alleviate social problems in the physical world, for example, the drinking, smoking, and drug related problems. We proposed a *Positive Influence Dominating Set (PIDS)* selection algorithm and analyzed its effect on a real online social network data set through simulations. By comparing the size and the average positive degree of PIDS with those of a 1-dominating set, we found that by strategically choosing 26% more people into the PIDS to participate in the intervention program, the average positive degree increases by approximately 3.3 times. In terms of the application, this result implies that by moderately increasing the participation related cost, the probability of positive influencing the whole community through the intervention program is significantly higher. We also discovered that a power law graph has empirically larger dominating sets (both the PIDS and 1-dominating set) than a random graph does.

1 Introduction

Online social network is a network composed of individuals who share the same interest and purpose which provides a powerful medium of communicating, sharing and disseminating information, and spreading influence beyond the traditional social interactions within a traditional social network setting. Online social network has developed significantly in recent years. For example, online social network sites like Facebook, MySpace are among the most popular sites on the Internet; online social networks have also raise special interest among commercial businesses, medical and pharmaceutical companies as a channel to influence the opinion of their customers; even police has utilize the information in online social network sites to track down crimes.

Some research has been done to understand the properties of online social networks [1][2][3] and how to effectively utilize social networks to spread ideas and information within a group [4]. In this paper, we explore the problem of *how to utilize online social networks to help alleviate social problems in the*

physical world. Some examples of these type of problems include the drinking, smoking, and drug related problems. These social issues are very intricate and complex problems that require a system-level 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. In a social setting, people can have both positive and negative impact on each other and a person can take and move among different roles since they are affected by their peers. For example, within the context of drinking problem, a person can be an abstainer, or a binge drinker. An abstainer has positive impact on his direct friends (called neighbors) but he might turn into a binge drinker and have negative impact on his neighbors if many of his friends are binge drinkers.

Alcohol intervention strategies and programs that consist of disseminated education and therapy via mail, Internet, or face-to-face interviews such as *motivational feedback*[5] are important tools to help combat some of the social problems within today's society. In an ideal world to truly alleviate the main source of the drinking problem, one must educate as many binge drinkers as possible. This will prevent an abstainer who might adopt the negative influence of his close binge drinker friends from eventually turning into a binge drinker. If too few people are selected to participate in an intervention program, there is a high likelihood that the positive effect of the intervention will be overrun by the negative effect exerted by the binge drinkers, not included in the intervention, on those who are vulnerable. On the other hand, due to the financial limitations in budget, it is impossible to include all the binge drinkers in the intervention program. Therefore, how to choose a subset of individuals to be part of the intervention program (or to educate) so that the effect of the intervention program can spread through the whole group under consideration becomes the key item of inquiry. In an effort to address this question, the specific problem we study in this paper is the following: given an online social network of a community and a specific social problem, how to identify a subset of the individuals within the online social network to participate in an intervention program such that the intervention/education can result in a globally positive impact on the entire social network. We assume that 1) if more than half neighbors of an individual have positive impact on him, then the probability that this individual positively impact others in the network is high. 2) intervention program can convert a negative influential individual to a positive influential person. Our first assumption comes from an extensive body of evidence suggesting that one of the most powerful predictors of negative/risky behavior in individuals is whether an individual has friends who also engage in that behavior. In fact research [6] has shown that about 50% of the variance in adolescent personality primarily reflects the influence of peers. 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. The overall average effect of his negative neighbors will be overpowered by the contributions of his positive neighbors. Our second assumption comes from

the work in [7] [8], where nearly every individual in the feedback intervention program showed a reduction in drinking. With the above two assumptions, the problem is equivalent to selecting a subset of the individuals to participate in the intervention program such that each individual in the social network has more positive neighbors than negative ones.

Online social network can be represented as a graph of relationships and social interactions (edges) between individuals (nodes). We use the following network model to illustrate the online social network in context of the social problem: A undirected graph $G = (V, E, C)$ is used to represent the online social network. We use an undirected graph because friendship in an online social network are mostly bi-directional. V is the set of nodes in which each node is an individual in the online social network. E is the set of edges in which each edge represents the existence of a social connection 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 impact on its neighbors. For example, for the drinking problem [9], the compartment of each node is one of the followings: abstainer, problem drinker, social drinker or binge drinker. A node in the abstainer compartment has positive impact and all nodes in any of the other three compartments have negative impact. With the above network model, we define the problem of selecting a *positive influence dominating set (PIDS)* in the online social network G as finding a subset P of V such that any node u in V are dominated by at least $\lceil \frac{d(u)}{2} \rceil$ nodes in P where $d(u)$ is the degree of node u. We propose a PIDS selection algorithm and evaluate its effect on a real online social network data set through simulations by comparing the size and the average positive degree of PIDS with that of the traditional 1-dominating set. The results illustrate that by strategically choosing 26% more people in PIDS than in the 1-dominating set to participate in the intervention program, the average positive degree increases approximately by 3.3 times. This is a considerable increase in overall average positive influence emitted in an individual in comparison to the nominal increase in number of participants. Thus by moderately increasing the participation related cost, the probability of positive influencing the whole community through the intervention program is significantly higher. Our simulation results also reveal that a power law graph has empirically larger dominating set size than a random graph even though the dominating set problem is a theoretically easier problem in a power law graph than in a random graph.

Our contributions in this paper include: 1) we introduce the PIDS problem which formalizes the problem of utilizing online social network to help solving social problems; 2) we propose and evaluate the PIDS selection algorithm with a real online social network data set and compare the size of PIDS and traditional 1-dominating set; 3) we study the size of PIDS and 1-dominating set in both a power law graph and a random graph.

The rest of this paper is organized as follows. Section 2 describes the related work. In section 3, we present the PIDS selection algorithm. Section 4 shows the

simulation results of PIDS on a real online social network topology. Section 5 concludes this paper and discusses our future plan.

2   Related Work

Most of the current research in online social network fall in two categories: one is to understand the properties and characteristics of online social networks, such as the work in [1][2][3]. The other is to study how to utilize social network to spread ideas and information as presented in [4][10]. Our work focuses on exploring how to utilize online social networks to help alleviate social problems in the physical world. The social problem is different from spreading ideas and information. The spread of ideas and information is one direction in that once a person in influence to adopt an idea or learn some information, he cannot revert to his original state by means of future influence. The positive or negative influence in social problems can flow in two directions, that is a positive individual can convert to a negative individual then can move back and forth between these two states multiple times. Another different between our work and that of [4][10] is that we find a set of individuals that guarantees the positive effect of education will be injected into the entire group. They focus on finding a subset of a pre-established fix size that maximize the spread of information, but not the subset regardless of size that infuses information to the whole group.

3   Positive Influence Dominating Set Selection Algorithm

In this section, we present a PIDS selection algorithm for the positive dominating set problem formalized in the earlier section. First we define and explain a few terms and definitions used in the description of our algorithm. Each node can have either *positive* or *negative* impact on its neighbor nodes. We call a node with positive impact a *positive node* and a node with negative impact a negative node. The *positive degree* of a node is the number of its positive neighbors. The same holds for negative degree. The *compartment* of a node decides whether the node is a positive or a negative node. For example, in the context of college drinking [9], a node in the abstainer compartment is a positive node and a node in any other compartment is a negative node. Nodes that are chosen into the PIDS are marked as positive nodes. Thus a neighbor u of v is a *positive neighbor* if u is initially a positive node or u is selected into the PIDS. A *1-dominating set* is S of a graph G is a subset of nodes in G such that every node not in S has at least one neighbor in S. A *positive influence dominating set* P of a graph G is a subset of nodes in G that any node u in G are dominated by at least $\lceil \frac{d(u)}{2} \rceil$ nodes in P where $d(u)$ is the degree of node u.

The main idea of PIDS algorithm is as follows: first prune the original graph by removing the initial positive nodes, then iteratively choose a 1-dominating set of the graph consisting of nodes with less than half neighbors as positive neighbors until all nodes in the original graph are either positive nodes or have

more positive neighbors than negative ones. To choose a 1-dominating set of a graph, we use a greedy algorithm similar to the one in [11]. This algorithm selects the node with the largest node degree into the dominating set. In our greedy algorithm, we choose the node that can dominate most negative nodes into the dominating set. [11] has proven that the simple greedy algorithm gives a $1 + o(1)$ approximation with a small constant in $o(1)$ to the 1-dominating set problem in a power law graph. Algorithm 1 gives the details of the PIDS algorithm. Compartment vector C contains the nodes that are initially in positive compartment. $C = \phi$ means every node is an negative node.

Algorithm 1 Positive Influence Dominating Set Selection Algorithm

---

1: INPUT: A graph $G = (V, E, C)$ where V is the set of nodes, E is the set of edges that capture the social interactions of the nodes, C is the set of nodes that are initially in positive compartment.
2: OUTPUT: A subset P of V such that any node u in V has at least $\lceil \frac{d(u)}{2} \rceil$ neighbors in $P \bigcup C$, where $d(u)$ is the degree of node u.
3: initialize the status of all nodes in V to NEGATIVE and P to empty
4: let $V' = V - C$ and set the status of nodes in C to POSITIVE
5: calculate the degree and the positive degree of each node in V
6: T is the set of nodes in V' that have more positive neighbors, set the status of nodes in T to POSITIVE
7: let $V' = V' - T$
8: while not every node in V' has more positive neighbors
9:     find a 1-dominating set S that dominates all nodes in V' using greedy algorithm as in [11]
10:     update $P = P \bigcup S$
11:     $V' = V' - S$
12:     update the positive degree of each node in V
13:     T is the set of nodes in V' that have more positive neighbors, set the status of nodes in T to POSITIVE
14:     $V' = V' - T$
15: end of while
16: let $U = C \bigcup P$
17: while not every $w \in U$ has more positive neighbors
18:     x is a neighbor of w that has maximum negative degree among all the neighbors of w
19:     let $P = x \bigcup P$
20:     update the positive degree of each node in V
21: end of while
22: The nodes in set P are positive influence dominating set of G, and the positive degree of each node is also calculated.

---

4 Performance Evaluations

To evaluate the effect of our PIDS algorithm, we first collect the data from one of the popular gaming applications, Fighter's Club (FC) [2] on Facebook social networking site. The FC game has attracted over 3.4 million Facebook users since it was initially launched on June 2007. The gaming application records the players as well as their IP addresses. See [2] for a detailed descriptions of the game. In our study, we choose a subset of 2,334 Facebook users that play the online games together, and their IP addresses belong to the same IP network. Figure 1 illustrates the distribution of the node degree in this online community. Similar to the observations in prior studies [12], we find that the node degree in this Facebook application community also follows a power-law distribution.
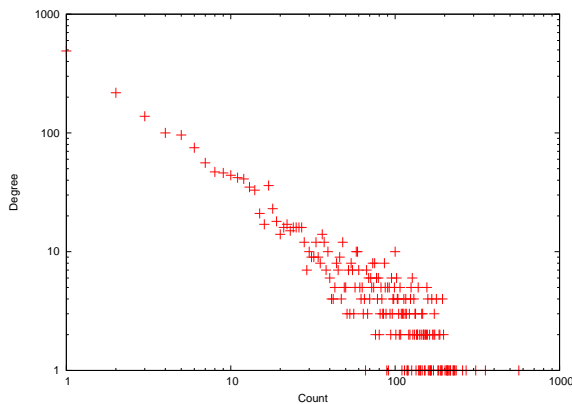

Fig. 1. Node degree distribution of an gaming application of 2334 users on facebook

To understand the effect of PIDS, we need to answer the following questions: 1)how many nodes need to be selected into the PIDS and how influential these nodes can be. To measure influence, we calculate the average over the number of positive neighbors of each node (called positive degree). The higher the average positive degree is, the more influential the PIDS can be. 2) what is the difference between the size of PIDS and a 1-dominating set and what is the difference between the influence of these two sets. 3) what is the difference between the size of dominating sets (both the PIDS and the 1-dominating set) of a power-law graph and that of a random graph. There are two main 1-dominating set construction algorithms, greedy algorithm (which is used in the step of generating 1-dominating set in the PIDS algorithm) or tree based algorithm as the one proposed in [13]. Tree based algorithm first builds a depth first tree then chooses a node that has the lowest level and most negative neighbors among all nodes on the same level in the dominating set. We are also interested in the performance of these two algorithms in the PIDS problem. In the remaining of this paper, we call them greedy PIDS algorithm and tree-based PIDS algorithm respectively.

Table 1. Dominating set of a topology retrieved from the FC gaming application on Facebook social network

Algorithm	Size of dominating set (percentage)	Avg. positive degree (percentage)
greedy positive influence dominating set	1358(58.2%)	22.5(79.3%)
greedy 1-dominating set	744(31.9%)	6.87(24.1%)
tree-based positive influence dominating set	1478(63.3%)	23.1(81.4%)
tree-based 1-dominating set	851(36.5%)	6.85(24.1%)

Table 1 illustrates the results on the topology retrieved from the FC gaming application on the Facebook social network, whose node degree follows a power-law distribution. The size of the community is 2334, and average degree is 28.42, with 15% of all nodes initialized as positive. The result in the table is the average over 100 runs. Note that for each run, the size and average degree do not change. Since the abstainer nodes are randomly chosen, the network topology is different for each run. As we can see, the greedy algorithm has better performance over the tree-based algorithm in this power law graph. Applying greedy PIDS algorithm, 58% of the nodes are chosen into the dominating set and the resulting average positive degree is 22.5, which means that for a given node, on average, approximately 79% of its neighbors are positive nodes. This outcome of 79% is significant larger than our set goal, which is 50%. The high average positive degree can greatly increase the possibility of the whole community turning into a positive community. Furthermore, it indicates that we can prune the dominating set and reduce its size while keeping the average positive degree at around 50%. For greedy 1-dominating set algorithm, 31.9% nodes are chosen into the dominating set and the resulting average positive degree is only 6.85, which means that for a given node, approximately 76% of its neighbors are negative neighbors. Thus there exists a great possibility for the node of becoming a negative node and negatively impact others. The simulations reveal that by strategically choosing 26% more people into PIDS to participate in an intervention program, the average positive degree increases approximately by 3.3 times. In terms of the application, we can be confident that by moderately increasing the participation related cost, the probability of positive influencing the whole community through an intervention program where participants selection is determined by the greedy PIDS is significantly higher. Another observation is that the greedy algorithm gives smaller dominating sets than the tree-based algorithm does in an online social network, in which the node degree follows a power law distribution.

Table 2 illustrates the size of dominating sets in random graphs. We simulate a random graph by throwing a set of nodes to a square area and randomly generate the (x, y) coordinates of each node. Each node is associated with a variable range and if two nodes are within each other's range, there exists a link connecting them. We set the size of network to 2334 and initialize 15% nodes as positive, which is the same as the real online social network retrieved from the gaming application on the Facebook website, and adjust the parameters (area

Table 2. Dominating set of random networks

Algorithm	Size of dominating set (percentage)	Avg. positive degree (percentage)
greedy positive influence dominating set	1166(50%)	19.3(67%)
greedy 1-dominating set	233(10%)	6.47(22.5%)
tree-based positive influence dominating set	1196(51.3%)	18.8(65.2%)
tree-based 1-dominating set	253(10.9%)	6.32(21.9%)

edge length, minimum range and maximum range) to get an average node degree close to the online social network topology (Note it is hard to generate a random graph that has exactly the same average node degree as that in the online social network). The average of the average degree over 100 runs is 28.82, and Comparing Table I and II, we can see that under similar settings (same node size and similar density), a random graph has significantly smaller dominating set than a online social network which is a power law graph. This contradicts our conjecture. Heuristically, in a power law graph, a small set of nodes with high degree should dominate most of the nodes in the graph and thus resulting in a smaller dominating set. An explanation can be that even though the network is clustered around the most influential (or connected) nodes in the power law graph, there are more nodes that are sparse (has few neighbors) so more dominator nodes are chosen to dominate these sparse nodes. This interesting result needs further investigation.

In summary, in a typical online social network where average node degree is about 28.42 and 15% nodes as initially positive nodes, approximately 60% nodes are chosen into the PIDS, which results in an average percentage of positive degree over node degree as high as approximately 79%. Furthermore, there are 26% more people in the PIDS than those in 1-dominating set, while the average positive degree in the PIDS is approximately 3.3 times that in 1-dominating set. In the case of college drinking where participants of an intervention program are selected accord with the greedy PIDS, by moderately increasing the participation related cost, the probability of positive influencing the entire community is significantly higher. We also found that power law graphs have larger dominating set size than random graphs even though dominating set problem is theoretically an easier problem in a power law graph than in a random graph [14].

5 Conclusions

In this paper, we introduced and studied the problem of how to utilize online social network as a medium to help alleviate a certain social problem. We proposed the PIDS selection algorithm to evaluate the effect of educating a subset of the entire target group susceptible to a social problem. Our simulation results reveal 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. We also discovered that the dominating sets of a power-law graph is larger than that of a random graph.

Our future work includes applying a mathematical model to understand the influence of each individual and study the effect of PIDS over a period of time. Since the PIDS is fairly large in power law graph, we will investigate what is an empirically proper positive degree threshold that can spread the positive education influence throughout the entire community under consideration through additional modeling, experiment and data analysis.

References

1. Alan, M., Massimiliano, M., P., G.K., Peter, D., Bobby, B.: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on Internet Measurement Conference(IMC). (2007) 29–42
2. Nazir, A., Raza, S., Chuah, C.N.: Unveiling facebook: A measurement study of social network based applications. In: Proceedings of ACM SIGCOMM Internet Measurement Conference (IMC). (2008)
3. Aris, A., Ravi, K., Mohammad, M.: Influence and correlation in social networks. In: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD). (2008) 7–15
4. David, K., Jon, K., Éva, T.: 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). (2003) 137–146
5. Walters, S.T., Neighbors, C.: Feedback interventions for college alcohol misuse: What, why and for whom? Addictive Behaviors (2005) 1168–1182
6. Jaccard, J., Blanton, H., Dodge, T.: Peer influences on risk behavior: Analysis of the effects of a close friend. Developmental Psychology (2005) 135–147
7. E., L.M., M., C.J.: Identification, prevention and treatment: A review of individual-focused strategies to reduce problematic alcohol consumption by college students. (2002) 148–163
8. T., W.S., E., B.M.: Addressing drinking among college students: A review of the empirical literature (2000)
9. L.Almada, E.Camacho, R.Rodriguez, M.Thompson, L.Voss: Deterministic and small-world network models of college drinking patterns. In: AMSSI Technical Report at http://www.amssi.org/reports/alcohol2006.pdf. (2007)
10. Shishir Bharathi, David Kempe, M.S.: Competitive influence maximization in social networks. In: Proceedings of 3rd International Workshop On Internet And Network Economics(WINE). (2007)
11. Stephen, E., Anil, K.V.S., V., M.M., Aravind, S., Nan, W.: Structural and algorithmic aspects of massive social networks. In: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms (SODA). (2004) 718–727
12. Albert-Lszl, B., Bonabeau, E.: Scale-free networks. (2003) 60–69
13. Wan, P.J., Alzoubi, K.M., Frieder, O.: Distributed construction of connected dominating set in wireless ad hoc networks. In: Proceedings of 21th Annual Joint Conference of the IEEE Computer and Communication Societies(InfoCom). (2002)
14. Ferrante, A., Pandurangan, G., Park, K.: On the hardness of optimization in power-law graphs. Theoretical Computer Science (2008) 220–230