

Social Network Based Substance Abuse Prevention via Network Modification (A Preliminary Study)

Aida Rahmattalabi

University of Southern California
rahmatta@usc.edu

Anamika Barman Adhikari

University of Denver
anamika.barmanadhikari@du.edu

Phebe Vayanos

University of Southern California
phebe.vayanos@usc.edu

Milind Tambe

University of Southern California
tambe@usc.edu

Eric Rice

University of Southern California
eric@usc.edu@usc.edu

Robin Baker

Urban Peak Organization
Robin.Baker@urbanpeak.org

ABSTRACT

Substance use and abuse is a significant public health problem in the United States. Group-based intervention programs offer a promising means of preventing and reducing substance abuse. While effective, unfortunately, inappropriate intervention groups can result in an increase in deviant behaviors among participants, a process known as *deviancy training*. This paper investigates the problem of optimizing the social influence related to the deviant behavior via careful construction of the intervention groups. We propose a Mixed Integer Optimization formulation that decides on the intervention groups to be formed, captures the impact of the intervention groups on the structure of the social network, and models the impact of these changes on behavior propagation. In addition, we propose a scalable hybrid meta-heuristic algorithm that combines Mixed Integer Programming and Large Neighborhood Search to find near-optimal network partitions. Our algorithm is packaged in the form of GUIDE, an AI-based decision aid that recommends intervention groups. Being the first quantitative decision aid of this kind, GUIDE is able to assist practitioners, in particular social workers, in three key areas: (a) GUIDE proposes near-optimal solutions that are shown, via extensive simulations, to significantly improve over the traditional qualitative practices for forming intervention groups; (b) GUIDE is able to identify circumstances when an intervention will lead to deviancy training, thus saving time, money, and effort; (c) GUIDE can evaluate current strategies of group formation and discard strategies that will lead to deviancy training. In developing GUIDE, we are primarily interested in substance use interventions among homeless youth as a high risk and vulnerable population. GUIDE is developed in collaboration with Urban Peak, a homeless-youth serving organization in Denver, CO, and is under preparation for deployment.

KEYWORDS

Multiagent Systems; Social Networks; Robust Optimization

1 INTRODUCTION

Peers have a direct influence in adolescents' risk behaviours. Substance use and abuse is a significant public health problem

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 2019, Montreal, Canada

© 2019 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.
<https://doi.org/doi>



Figure 1: Urban Peak site. We are collaborating with Urban Peak organization for conducting the group-based interventions

in the United States, particularly among youth. According to the Monitoring the Future study [12], around 54 percent of high school students have tried at least one illicit substance. Homeless youth, in particular, are shown to be disproportionately affected, with substantially higher levels of substance use compared to the housed youth [14].

Notably, interventions attempting to reduce substance use have successfully utilized social networks to disseminate and reinforce behavioral norms supportive of protective behaviors related to substance use (e.g., [19]). The way the social network is utilized is often through formation of subgroups where the individuals can talk, share experiences and engage in various constructive activities. Such social network-based approaches to substance abuse prevention are considered more promising because young people learn better from one another, find each other more credible and have a shared environment and culture [6].

Unfortunately, these social network-based efforts may inadvertently increase the chances of youth being exposed to negative social influences, as they do not explicitly structure the intervention groups. This can result in an effect known as deviancy training. Deviancy training occurs when high-risk youth are aggregated together and reinforce negative behaviors and attitudes.

Social network-based interventions have typically grouped participants into intervention groups in three ways [19]: (1) *random assignment*: participants are randomly assigned to the groups; (2) *network based assignment*: participants are assigned based on their own nominations; (3) *teacher nominated assignment*: groups are created based on teacher nominations, often corresponding to an assignment based on the participants' behaviors, i.e., even distribution of high-risk individuals. The random assignment and teacher nominated methods are less effective because they are not designed to leverage naturally occurring ties in participants' social networks. The network-based method on the other hand does leverage these

natural forms of influence. However, it is shown that it can have unintended effect of exacerbating the problem behavior [19].

Recently, successful applications of AI based decision aids such as [21] and [5] have encouraged efforts to address such complicated social problems using techniques in AI and optimization [15]. In the present work, we aim to tackle the problem of deviancy training in substance abuse intervention by structuring more effective groups, one that effectively partitions the participants' social network.

Building this decision aid, however, raises major challenges.

First, it has been observed that in the process of being a part of the network-based intervention, the sub-networks (peer-groups) undergo a transformation, whereby network ties change in strength [17]. Such changes will consequently change the influence spread patterns and therefore constitute an important component of the model. Second challenge is understanding the influence process. In this regard, we propose an influence spread model which is based on the competitive Linear Threshold (LT) Model in [3] to explain both positive and negative social influences within interventions. The third major challenge is to strategically choose the right peer-groups from the larger social network to mitigate the challenge of deviancy training. In the following, we will refer to those engaged in substance use as "users", that we wish to have socially influenced by "non-users" peers in the smaller groups, so as to reduce/prevent such substance use and abuse.

We view this problem as an influence maximization problem, with the goal of maximizing/minimizing the total positive/negative influence by careful construction of the intervention groups, considering the consequent changes in the network topology. Influence maximization in networks has been widely studied in the past decade and several works have tried to address the problem of modifying the network topology in order to maximize positive influence spread or curtail undesirable behaviors. In particular, [9, 10] study the problem of edge deletion/addition under the LT model. Our problem however is different: we do not directly decide on edges to add/delete. Instead, we can only select how participants are partitioned into intervention groups. Depending on this choice, there is a disciplined process that controls how social network changes. More precisely, we will cluster the given social network into intervention groups, which causes further changes in the network topology as a byproduct, and subsequently controls how the positive and negative influences propagate.

Social influence optimization and network dynamics distinguish this work from the existing literature on network clustering. For example, [2] studies clustering graphs with both positive and negative edges, but the structure of the network is static. Further, unlike [7] that considers the network dynamic, we are interested in optimizing a stochastic influence function, plus, the evolution of the network topology is not just a function of time but a function of the decisions variables or clusters. To best of our knowledge no work has addressed such influence-based clustering of networks to this date.

To address these challenges, we propose an AI-based decision aid, called GUIDE (GroUp-based Intervention DEcision aid). GUIDE assists interventionists in substance abuse prevention by giving recommendations regarding the intervention groups. In what follows we detail our main contributions: (I) First, we introduce a novel problem to AI researchers; We propose a mathematical model for this type of intervention, which includes the key aspects of the

influence spread and network dynamics. The model enables us to predict, both the expected success of the intervention, measured as the expected number of "non-users" at the end of the intervention, and the possibility of harm, or deviancy training. (II) We show that the problem of finding the optimal partition to minimize the expected substance use is NP-hard. (III) Therefore, we propose both a Mixed Integer Linear Program (MILP) to solve this problem and a hybrid meta-heuristic algorithm that combines MIP and Large Neighborhood Search (LNS) to find near-optimal network partitions. (IV) We provide extensive analysis of the model, using both real-world social network data and synthetic graphs and we show that our proposed partitions can significantly outperform all common practices such as random assignment of individuals or participants choice.

2 PROBLEM STATEMENT

Our goal is to create an assistant that will aid an interventionist for substance abuse prevention. The intervention that we consider is a 6-week program, in which participants are placed into different groups. The size of each group is bounded between 4 and 7. An interventionist interacts with each group. Pre-intervention, the participants are asked to report both their hard-drug using behaviors, and their network relationships (i.e., who they know) and what the strength of that relationship is (whether a person is a strong or a weak tie). Over the course of the intervention and based on the groups assignments, these social ties will change: some may become weaker, while others may grow stronger. We aim to maximize the number of "non-users" one-month after the intervention is complete.

We view the problem of group configuration for substance abuse prevention as a graph partitioning problem. Given a social network, the goal is to find a partition such that the expected number of "non-users," at the end of the intervention, is maximized. More formally, let graph $G = \langle V, E \rangle$ be a directed graph representing the given social network, with V as the set of all nodes (individuals in the social network), and $E \subseteq V \times V$ as the set of all edges (social ties). Arc $i, j \in E$ indicates existence of an arc pointing from i to j , signifying that j has reported i as a friend pre-intervention. Associated with each node $i \in V$ is a *node behavior* indicating that node i is a substance "user" or a "non-user." The scalar $b_i \in \{0, 1\}$ encodes the node behavior with $b_i = 1$ if and only if node i is a "user." Additionally, associated with each edge $e = i, j \in E$ is the *edge strength* $s_e \in \{0, 1\}$, where $s_e = 1$ (resp. 0), indicates that j considers i as strongly (resp. weakly) influential for him.

We model the intervention as a partition of V into S subsets $P_s, s = 1; \dots; S$, such that $\bigcup_{s=1}^S P_s = V, \bigcap_{s=1}^S P_s = \emptyset$, where each subset must consist of at least \underline{C} and at most \bar{C} nodes. Note that the number of subsets S is a decision variable, but \underline{C} and \bar{C} are pre-specified. During the intervention and based on the group configurations, the ties undergo changes. Some ties are cut or weakened and some will be reinforced. We assume that, given a choice of partition $P := \{P_s\}_{s=1}^S$, the post-intervention structure of the network is known deterministically. We let $G^{+1}P = \langle V, E^{+1}P \rangle$ denote the post-intervention graph when groups are formed according to partition P , with $E^{+1}P$ corresponding to the new edge set. Thus, $i, j \in E^{+1}P$ if and only if j considers i as a friend post-intervention. Accordingly, for

Same Group	no-tie	weak	strong
(user, user)	strong	strong	strong
(non-user, non-user)	strong	strong	strong
(non-user, user)	weak	weak	strong
(user, non-user)	weak	weak	strong
Separate Groups	no-tie	weak	strong
(user, user)	none	none	strong
(non-user, non-user)	none	weak	strong
(non-user, user)	none	none	weak
(user, non-user)	none	none	weak

Table 1: Changes in tie strength post-intervention. The existing relationships, and the behavior of the individuals as well as their assignment to groups impacts the changes.

$e \in E^{+1P^0}$, we denote by $s_e^{+1P^0} \in \{0, 1\}$ the strength of edge e post-intervention.

The behavior of node $i \in V$ post-intervention is random. We let $B_i^{1P^0}$ denote the random variable that represents the behavior of node i post-intervention, i.e., $B_i^{1P^0} = 1$ if and only if node i is a “user.” This is a complicated stochastic function of the partition that depends on both the link formation and breakage rules and the influence model assumed. Figure 2 shows an example network pre- and post-intervention.

Mathematical Formulation. Mathematically, the problem of selecting the optimal partition P that maximizes the expected number of “non-users” in the network post-intervention can be formulated as:

$$\begin{aligned}
 & \text{maximize} && E \sum_{i \in V} (1 - B_i^{1P^0}) \\
 & \text{subject to} && S \subseteq \mathcal{N}_+ \\
 & && P_s \cup \bigcup_{s=1}^S P_s = V \\
 & && |P_s| \leq C \quad \forall s \in \{1, \dots, S\}
 \end{aligned} \tag{1}$$

Where $E \{ \cdot \}$ denotes the expectation operator with respect to the distribution of $\{B_i^{1P^0}\}_{i \in V}$. The objective counts the expected number of “non-users” in G^{+1P^0} . Note that we are considering the intervention period as one time step and therefore, we are assuming a single-stage influence process. The first three constraints ensure that P forms a partition of the node set. The last constraint enforces capacity constraints on each group.

Problem (1) is a combinatorial optimization problem which is hard to solve as formalized by Theorem 2.2. In order to solve this problem, both the random behavior B , as well as the network structural changes must be modeled. In the following two sections, we propose models for network and influence dynamics that are supported by the social work literature.

Tie Formation and Breakage. While these youth have pre-

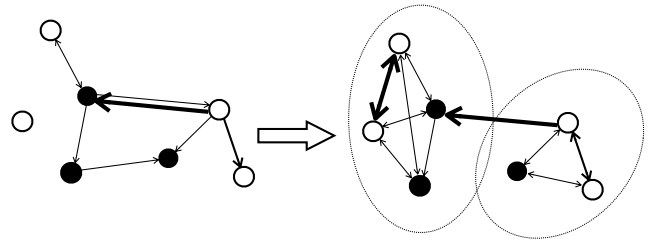


Figure 2: An example network pre- (left) and post- (right) intervention. The black (resp. white) circles indicate “user” (resp. “non-user”) nodes. Weak (resp. strong) links are denoted by thin (resp. thick) arrows. The ellipsoids represent the two groups that are formed for the intervention. As seen, new edges are created within the groups, while some edges are cut across the groups. partitions

existing strong or weak ties, especially centered around their substance use, the strength of these relationships would weaken or strengthen during the intervention [4]. There is empirical evidence to suggest that the more similar two individuals are, the stronger their ties are [1]. This is explained by their behavioral homophily. Homophily refers to the tendency of people to associate with people who are like themselves and these ties also often facilitate more communication and influence [13]. Therefore, if two individuals are both “users” or both “non-users,” we assume that they would develop a strong tie over time. Based on this rationale, we also model the interventionist who will exert a positive influence as an additional “non-user” who develops a weak tie to the “users” and a strong tie to the “non-users,” during the intervention period. An important component of the proposed intervention is to develop skills that would facilitate bonding with pro-social peers and discourage interactions with youth who have high-risk behaviors [18]. Therefore, if two individuals are separated and at least one of them has “user” behavior, their tie weakens or is cut. Table 1 summarizes how we expect strength of ties to change in response to graph partitioning. In this table, the rows indicate the behavior of the pair of nodes under consideration, and the columns show their pre-intervention tie. The changes are defined as a function of the ties pre-intervention (E), the behavior of the end nodes (B_i), and whether they are assigned to the same group or not.

Example 2.1. Consider the example network in Figure 2. In this network, black circles indicate “user” individuals and white circles represent “non-user” people. Initially, individuals are connected with either weak or strong ties, where weak ties are shown via light arrows, and strong ties are depicted via bold arrows.

Given this social network as input, we need to group the individuals into two intervention groups. Based on this decision, new connections will form. Consider the grouping as depicted in Figure 2. Post-intervention, the ties undergo changes according to table 1. Therefore, new ties will be formed between individuals that did not have any ties before intervention and are placed in the same group. Also, some of the weak existing ties become stronger before the intervention. Also, some of the weak ties will become stronger

as a result of the grouping. Finally, the existing ties that are across groups will be broken or weakened.

Substance Abuse Prevention Influence Spread Model. Following the intervention, and depending on how the network evolves, we evaluate the influence to predict the changes in the nodes' behaviors. We use a variant of the popular Linear Threshold model [8] proposed in [3], in which they study the competitive influence processes in social networks. Our problem is similar to [3] as there are two concurrent influences, a positive influence which originates from the "non-users" and a negative influence from "users." However, it differs from their model in that the nodes in our network have already adopted one of these two behaviors.

Consider the post-intervention graph as $G^+ = (V, E^+)$. For notational convenience, we henceforth eliminate the dependence of G^+ and E^+ on P . For each $i, j \in V$, a weight value $w_{ij} \in \mathbb{R}$ is assigned, which characterizes the amount of influence from node i to node j . These weights are such that the sum of the weights of the incoming edges to a certain node is bounded. Specifically,

$$\sum_{i \in V} w_{ij} \leq 1; \quad \forall j \in V;$$

Initially, each node $i \in V$ is assigned a threshold value, denoted by τ^i , corresponding to the influence threshold for switching behavior (from "user" to "non-user" and vice-versa). We assume these thresholds are uniformly distributed in the range $[0, 1]$. A necessary condition for a node to change his behavior is for the sum of his neighbors' weights with opposite behavior to exceed the node's threshold. Mathematically, if for example, $b_j = 1$ and $\sum_{i: b_i=0} w_{i,j} > \tau^j$, node j

has the necessary condition to become a "non-user." Following the model from [3], if the threshold of a "user" (resp. "non-user") is exceeded, the node behavior will switch with probability u_j (resp. n_j). Thus, for any node i , we can express their post-intervention behavior distribution as:

$$P^i(b_j = 1) = u_j \sum_{i: b_i=0} w_{ij} > \tau^j, \text{ if } b_j = 0$$

$$P^i(b_j = 0) = n_j \sum_{i: b_i=1} w_{ij} > \tau^j, \text{ if } b_j = 1$$

THEOREM 2.2. *Problem (1) is NP-Hard, even under the presented influence model.*

PROOF. We construct a reduction from Set Cover problem with N items to be covered by M subsets. The reduction can be explained as follows:

For any item in the set cover problem, a "non-user" node, and for any set, a "user" node will be created. There is a tie from a "user" to a "non-user", if and only if the item is included in a set. In the drug prevention problem, we aim to partition the network into S sub-graphs. However, several assumptions are made. First, all ties across partitions are cut. Every pair of "user" in the same group form a tie. Every member of the network is connected to a "non-user" (the interventionist). All the ties have the same weight. The threshold

values for "non-users" are set to $\frac{\text{Number of "user" connections}}{\text{Number of "user" connections} + 1}$.

"User" thresholds are set equal to $\frac{1}{2} + \epsilon$ ($\epsilon > 0$, small positive number.) The capacity of each group is $N + M$. If the optimal solution is at least $\frac{1}{2}N + S + 1$, there is at most one group with more than two "users". This is because, if there are at least $N + S + 1$ "non-user" at the end of the intervention, at least S "users" must have flipped to "non-user", and the way the thresholds are set requires that the "users" are not put together in a group, in more than one group. Also, it is best to separate the "users" from "non-user". Therefore, the optimal S "user" to isolate will be equivalent to the solution of the Set Cover. Also, a solution to the Set Cover results in at least $N + S + 1$ "non-user". \square

3 MIXED-INTEGER PROGRAMMING FORMULATION

We present a MILP formulation of Problem (1). First, we define $s_{ij}^s, s_{ij}^w, s_{ij}^o$ as indicators for whether the pre-intervention tie from i to j is strong, weak, or there is no tie, respectively. The values for these variables are known from the initial data on the network structure and are determined as $s_{ij}^s := 1 \mid i, j \in E, s_{ij}^w := 1 \mid i, j \in E$ and $s_{i,j}^o = 1$, and $s_{ij}^w := 1 \mid i, j \in E$ and $s_{i,j}^o = 0$, for each $i, j \in V$, where 1 is the indicator function.

Let us first introduce the decision variables. We define Z_{ij} as a binary variable which is equal to 1 if nodes i and j are in the same group. Also, let $x_{ij}^{s+}, x_{ij}^{w+}, x_{ij}^{o+}$ be binary variables to encode the post-intervention tie strength. Finally, w_{ij} is the normalized post-intervention weight of the tie between i and j . Next, we introduce the constraints in this problem.

The groups must satisfy the capacity constraints, that is the sum of all the nodes that are in the same group must meet the following constraint, i.e.,

$$\sum_{i \in V} Z_{ij} \leq C; \quad \forall j \in V;$$

we require that if i and j are in the same group, and so are j and k , i and k must be in the same group as well. This can be expressed by saying that:

$$Z_{ij} + Z_{jk} - 1 \leq Z_{ki}; \quad \forall i, j, k \in V;$$

The post-intervention link strengths can be defined dependent on the pre-intervention edge strengths, the group assignments, and the behavior of the nodes. Therefore, according to Table 1, the link from i to j will become strong if for example, nodes i and j belong to the same group and either have similar behaviors, or they already have a strong tie, which is encoded as part of the following equation

$$x_{ij}^{s+} = Z_{ij} s_{ij}^s + j b_i + b_j - 1 + s_{ij}^o + s_{ij}^w$$

$$+ 1 - Z_{ij} j b_i + b_j - 1 + s_{ij}^s;$$

where the term $j b_i + b_j - 1$ is 1 iff i and j have the same behavior. Similarly,

$$x_{ij}^{w+} = Z_{ij} j b_i - b_j + s_{ij}^w + s_{ij}^o + 1 - Z_{ij} s_{ij}^s$$

$$+ s_{ij}^w - 1 + \max\{b_i, b_j\} + s_{ij}^o j b_i - b_j;$$

determines whether the new tie from i to j is weak. Once the post-intervention network is determined, the normalized weights are

calculated according to equation:

$$w_{ij} = \frac{W_s x_{ij}^{st} + W_w x_{ij}^{w+}}{i \sum_{j \in N_i} W_s x_{ij}^{st} + W_w x_{ij}^{w+}}$$

where W_s and W_w ; $W_w < W_s$ are numerical values for strong and weak ties, respectively. In this equation, the numerator is the strength of the tie from i to j , before normalization which is equal to W_s if it is a strong tie, or W_w if it is a weak tie. The denominator is the sum of all the incoming weights. While the above constraint consists of nonlinear terms, they can be linearized using standard techniques which require introducing a new decision variable for each product term.

Finally, the objective can be defined as follows:

$$\sum_{j: b_j=0} u_{jn} + \sum_{i: b_i=1} w_{ij} + \sum_{j: b_j=1} n_{ju} + \sum_{i: b_i=0} w_{ij}$$

This equation gives the closed form objective to our problem, which is the expected number of non-users post-intervention. The first term in the above equation is the expected number of non-users that remain non-users. The second term is the expected number of users that change. This equation is equivalent to the objective in Problem(1). Note that since in the influence model, the thresholds are drawn from a uniform distribution, the probability that a node changes its behavior equals the normalized incoming edge weights from neighbors with that behavior.

4 LARGE NEIGHBORHOOD SEARCH (LNS)

Local search algorithms target a set of feasible candidate solutions (neighborhood) and iteratively improve a solution by searching within that limited space. Small neighborhoods are faster to explore, but can make escaping a local optimum much harder. Large neighborhood search which was first proposed by [14], can alleviate this problem by exploring a carefully chosen large neighborhood. LNS iteratively improves a solution by alternatively destroying and repairing a solution.

Our solution approach is based on a mixture of LNS and MILP optimization which has been proven successful in [1]. Starting from an initial feasible solution, each loop starts by destroying the current solution by selecting two groups at random. It then repairs the solution by optimally deciding how to partition people between these two groups. Since this MILP subproblem is much smaller than the original MILP, it yields a significant improvement in the run time. It is important to note that in the destroy step, determining the amount of destruction is crucial. If too little is destroyed, the effect of large neighborhood is lost, whereas large destructions will result in a repeated re-optimization. We experimented with different neighborhoods, for example destroying the group assignment of random nodes or three groups at the same time (instead of two). In these cases, the results were not performing as well as the two-group neighborhood.

5 EXPERIMENTAL SETUP

Network Data. Social Network data primarily comes from the Youthnet study in which homeless youth were recruited from a drop-in center in Los Angeles County, CA. Clearly, our approach is independent of the input network data. Therefore, we use this

Figure 3: Comparison between small and large neighborhood search

dataset for our preliminary study. We also plan to deploy our approach in Denver and in collaboration with Urban Peak organization.

In order to explore different network structures, we use randomly generated Watts Strogatz (WS) graphs [25] with parameters $p = 0.25, k = 4$. WS graph models have properties such as short average path lengths which makes them close to real-world networks. Baselines. For evaluation, we compare MILP and LNS against three different baselines. These baselines are based on the three approaches that were detailed in the Introduction section (random assignment, network-based or by participants choice, and teacher nominated assignments).

Model Parameters. In the following experiments, and based on input from domain experts we choose the model parameters as:

$u_{jn} = 1:0, n_{ju} = 0:8, \underline{C} = 3, \bar{C} = 8, W_s = 3, \text{ and } W_w = 1$. We also explore the impact of variations of some of these parameters. For the LNS, 50 trials are performed, each starting from a different random initial solution and we report the solution with highest expected success rate. We also report averages over 25 different graph samples, both for the real and WS graphs.

6 RESULTS AND DISCUSSION

Solution Quality Metrics. Different solution strategies are compared based on success metric, which we define as:

$$\text{success} = \frac{E \sum_{i \in 2V} 1 - B_i P_i^{000}}{\max E \sum_{i \in 2V} 1 - B_i P_i^{000} \sum_{i \in 2V} 1 - b_i^0}$$

The numerator is the expected number of youth that have become non-users as the result of the intervention. The term $\sum_{i \in 2V} 1 - B_i P_i^{000}$

corresponds to the objective function of Problem(1). The denominator is its maximum possible value which is bounded above by $\sum_{i \in 2V} 1 - b_i^0$ it corresponds to the case where all users thresholds

are exceeded, and thus the maximum expected number of users that will change.

Neighborhood Search: Small vs. Large. We first compare the success rate of the two local search based algorithms. In each step of the small neighborhood search, a randomly chosen pair of nodes switch groups. If the resulting solution has a higher success rate, the solution is accepted and this process continues until a locally optimal solution is found. As explained previously, the LNS im-

Figure 4: Solution quality of MIP and LNS in GUIDE, against three baselines commonly employed by practitioners (Synthetic Graphs)

Figure 5: Solution quality of MIP and LNS in GUIDE, against three baselines commonly employed by practitioners (Real Graphs)

plementation is based on re-optimization of the randomly chosen groups. For both algorithms, each run is repeated for different initial random solutions until a time limit is reached. In this experiment, we use 10 different randomly generated WS graphs of size 30. The time limit is set to 1 hour. Figure 3 shows that the small neighborhood search can perform poorly relative to LNS under a variety of instances. This results indicates that between MILP, which is not easily scalable to larger networks, and pure small neighborhood local search which is favorable due its lower computational demand, LNS can achieve a balance between quality and speed and it is able to outperform local search, given the same time budget.

Solution Quality. Figures 4 and 5 compare the success rate of the optimization techniques, MIP and LNS against our baselines across 4 different network sizes, for both real network samples and WS samples. The MIP solver is given a cutoff time equal to the solution time of LNS (summed over the 50 trials). These results indicate that the solutions of both MIP and LNS are significantly better than any of the traditional methods for forming these groups, both statistically ($p < 0.01$) and practically. From the practical point of view, prevention science cares about effect size of interventions in addition to the statistical significance, and we observed in the results that we were able to provide solutions that outperform the traditional methods up to (20 - 41)%. Figure 5 also shows that in the real network samples (sizes 20 and 30), the baselines have a negative success rate. Negative success rate corresponds to the deviancy training effect, which means the expected number of non-users

Figure 6: Success rate with respect to different $\frac{u_{jn}}{n_{ju}}$ ratios

Figure 7: LNS solution time analysis. This plot indicates how the solution evolves with respect to the number of runs.

post-intervention is fewer than pre-intervention. This observation can be explained by the structure of the network as well as the behavior of the nodes. In fact, in these samples, a significant ratio of the individuals are users (68%), hence the negative influence dominates. Additionally, many of these users are strongly connected with their user peers, making it hard for them to be positively influenced. We should note that the optimized strategies are always strictly positive, though the success rate is limited due to the certain network structure. It can also be observed that the common intuition of evenly distributing the users across groups is in fact performing relatively well as compared to the other baselines. The importance of this result is twofold; First, it validates our model by reflecting the intuition of the social work partners that such a strategy is helpful. Secondly, we can clearly see that there is room for further improvements over these baselines.

There are computational lessons learned as well. For example, MIP is guaranteed to find the optimal solution by searching the entire solution space, but as shown here, it is not a practical solution due to time constraints. In fact, LNS outperforms MIP solution given the same time budget. Figure 6 compares the intervention success with respect to different $\frac{u_{jn}}{n_{ju}}$ ratios. This experiment is performed on synthetic graphs of size 40. It illustrates that the increase of this ratio will result in a decline in expected success of our intervention.

Scalability.

Figure 7 shows the LNS solution quality in a 40-node real-world graphs against the number of runs. Recall that LNS is run for 50

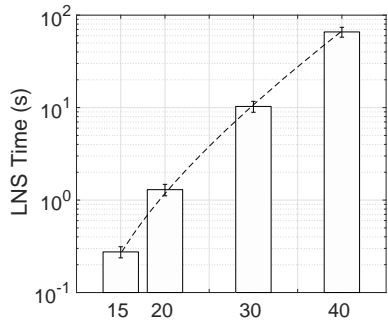


Figure 8: LNS solution time analysis. This plot shows how the time required to find a locally optimal solution changes with respect to the network size.

times, each starting from a different random initial solution. Figure 7 (a) shows that almost after 20 runs, the solution does not change significantly. Figure 8 shows the time needed for the LNS to find a locally optimal solution for different network sizes. This plot shows that as the size of the network increases, the LNS time increases polynomially. These results suggest that LNS is an efficient choice for our purpose, mainly because a decision aid, such as GUIDE, needs to be adjustable to unanticipated occurrences. For example in cases a person does not show up, or refuses to accept his assignment, GUIDE needs to recalculate a new solution relatively quickly based on the imposed constraints. Therefore we will utilize LNS as the core optimization algorithm in GUIDE.

7 DEPLOYMENT

GUIDE is developed in collaboration with Urban Peak, a homeless-youth serving organization in Denver, CO, and it is under preparation for deployment. We have developed an application that encapsulates the data collection phase as well as the implementation of the LNS, the optimization algorithm. This application has a user-friendly interface that facilitates the use of our algorithm for the practitioners.

In the first phase, detailed data about the social network and substance use behavior of each volunteer is collected. Then, a personalized ID is assigned to each individual and the data is stored on the local machine for future use. See Figure 9 (a). Next, the data is fed into the LNS algorithm to generate a suggestion for intervention groups. The practitioners can easily upload or choose an existing social network data file. After the optimization is performed, the results will appear on the screen, See Figure 9 (b). The results include a group assignment for each individual which is formatted as a table. Finally, the output can be printed and saved to a file for future references.

8 CONCLUSION

Substance abuse is a very significant public health and social problem in the United States. We have identified a niche in this wicked social problem where AI can make a contribution, illustrated the many modeling complexities, and provided algorithmic improvements over current practice. The result is GUIDE, an AI-based

(a) A snapshot of the data collection form (b) A snapshot of the GUI for the group recommendation

Figure 9: Two snapshots of the GUIDE application. (a) First, detailed information about the substance-use behavior of each individual is collected and saved in a data base. (b) Second, the practitioners will query the application for a group recommendation.

decision aid which leverages social network to structure intervention groups. We evaluated our approach against different traditional methods ranging from those based on intuitions such as distributing “users” to purely random strategies with no effort to understand the social circle of these youth. We showed that these traditional strategies significantly under-perform relative to our proposed method, emphasizing the importance of social influence in composition of these groups. In fact, here our focus is on the life choices of an extremely vulnerable section of the population. The result, as we discussed, is a model that is of tremendous value to prevention scientists, providing them with a quantitative means of understanding phenomena for which until now only qualitative observations were feasible. In particular, as these interventions are very costly and time consuming, the ability to forecast the likely impact of different strategies is very helpful.

REFERENCES

- [1] Sinan Aral and Dylan Walker. 2014. Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science* 60, 6 (2014), 1352–1370.
- [2] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. 2004. Correlation clustering. *Machine Learning* 56, 1-3 (2004), 89–113.
- [3] Allan Borodin, Yuval Filmus, and Joel Oren. 2010. Threshold models for competitive influence in social networks. In *Proceedings of the 6th international conference on Internet and network economics*. Springer-Verlag, 539–550.
- [4] Damon Centola and Michael Macy. 2007. Complex contagions and the weakness of long ties. *American journal of Sociology* 113, 3 (2007), 702–734.
- [5] John P Dickerson, Ariel D Procaccia, and Tuomas Sandholm. 2012. Optimizing kidney exchange with transplant chains: Theory and reality. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*. International Foundation for Autonomous Agents and Multiagent Systems, 711–718.
- [6] J Georgie, Harrison Sean, M Deborah, Hickman Matthew, Campbell Rona, et al. 2016. Peer-led interventions to prevent tobacco, alcohol and/or drug use among young people aged 11–21 years: a systematic review and meta-analysis. *Addiction* 111, 3 (2016), 391–407.
- [7] Robert Görke, Tanja Hartmann, and Dorothea Wagner. 2009. Dynamic Graph Clustering Using Minimum-Cut Trees. In *Proceedings of the 11th International Symposium on Algorithms and Data Structures*. Springer-Verlag, 339–350.
- [8] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.
- [9] Elias B. Khalil, Bistra Dilikina, and Le Song. 2014. Scalable Diffusion-Aware Optimization of Network Topology. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

- [10] Masahiro Kimura, Kazumi Saito, and Hiroshi Motoda. 2008. Solving the contamination minimization problem on networks for the linear threshold model. *PRICAI 2008: Trends in Artificial Intelligence* (2008), 977–984.
- [11] Boon Ping Lim, Menkes Van Den Briel, Sylvie Thiébaux, Russell Bent, and Scott Backhaus. 2015. Large neighborhood search for energy aware meeting scheduling in smart buildings. In *International Conference on AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*. Springer, 240–254.
- [12] Sean Esteban McCabe, Brady T West, Philip Veliz, Kenneth A Frank, and Carol J Boyd. 2014. Social contexts of substance use among US high school seniors: a multicohort national study. *Journal of Adolescent Health* 55, 6 (2014), 842–844.
- [13] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 27, 1 (2001), 415–444.
- [14] Adeline Nyamathi, Angela Hudson, Barbara Greengold, and Barbara Leake. 2012. Characteristics of homeless youth who use cocaine and methamphetamine. *The American journal on addictions* 21, 3 (2012), 243–249.
- [15] Aida Rahmattalabi, Phebe Vayanos, and Milind Tambe. 2018. A Robust Optimization Approach to Designing Near-Optimal Strategies for Constant-Sum Monitoring Games. In *Decision and Game Theory for Security*, Linda Bushnell, Radha Poovendran, and Tamer Başar (Eds.). Springer International Publishing, Cham, 603–622.
- [16] Paul Shaw. 1998. Using constraint programming and local search methods to solve vehicle routing problems. In *International Conference on Principles and Practice of Constraint Programming*. Springer, 417–431.
- [17] Jennifer R Spoor and Janice R Kelly. 2004. The evolutionary significance of affect in groups: Communication and group bonding. *Group processes & intergroup relations* 7, 4 (2004), 398–412.
- [18] Steve Sussman, Clyde W Dent, Alan W Stacy, and Sande Craig. 1998. One-year outcomes of project towards no drug abuse. *Preventive Medicine* 27, 4 (1998), 632–642.
- [19] Thomas W Valente, Beth R Hoffman, Annamara Ritt-Olson, Kara Lichtman, and C Anderson Johnson. 2003. Effects of a social-network method for group assignment strategies on peer-led tobacco prevention programs in schools. *American journal of public health* 93, 11 (2003), 1837–1843.
- [20] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of “small-world” networks. *nature* 393, 6684 (1998), 440.
- [21] Amulya Yadav, Hau Chan, Albert Xin Jiang, Haifeng Xu, Eric Rice, and Milind Tambe. 2016. Using social networks to aid homeless shelters: Dynamic influence maximization under uncertainty. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 740–748.