JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 40, 217-230 (2024)
DOI: 10.6688/JISE.202403_40(2).0001

Opinion Optimization for Two Different Social Objectives: Combinatorial Algorithms and Linear Program Rounding*

PO-AN CHEN+, YI-LE CHEN AND WEI LO
Institute of Information Management
National Yang Ming Chiao Tung University
Hsinchu City, 300 Taiwan
E-mail: poanchen@nycu.edu.tw; {jairachen78; lo.nuwa}@gmail.com

In this paper, we aim to optimize the two different social objectives of opinion optimization at equilibrium by controlling some individuals. This is usually called "Stackelberg games", in which a centralized authority is allowed to assign the strategies to a subset of individuals. The *Stackelberg strategies* of the centralized authority are the algorithms to select a subset of individuals and decide the actions for them in order to palliate the cost caused by the selfish behavior of the uncontrolled individuals. We give some combinatoral algorithms and linear program rounding algorithms as Stackelberg strategies for approximately optimizing the objective of utilitarian social cost (on special cases) and the objective of total expressed opinion (on general directed graphs), respectively.

Keywords: opinion optimization, Stackelberg strategies, combinatorial algorithms, linear program rounding, randomized algorithms

1. INTRODUCTION

Consider the process of how people make the decisions or form opinions in a social network composed by the individuals and their relationships in the society or community. It is common to see the opinions of the individuals, such as the viewpoints of public issues or the valuations of some products, are influenced by people surrounding them. The opinion forming process could be then naturally seen as some updating dynamics with respect to the interaction between one and his/her neighbors. DeGroot [1] gave a framework to model the opinion formation dynamics. In this model, each individual has a numeric-value opinion, which might be the satisfaction degree for some services or products, or the level of interest in some issues. This opinion will be updated to a weighted average of the opinions from its friends and its own, where the weights represent the degree of influence by the opinions from her friends and its own. The dynamics will converge to a fixed point that all individuals hold the same opinions, *i.e.*, a consensus.

However, we can easily observe that in the real world, the consensus is difficult to be reached. Friedkin and Johnsen [2] proposed a variant of the DeGroot model, in which each individual in the network updates its *expressed opinion* every time step while holding

Received July 19, 2022; revised September 13, 2022; accepted October 18, 2022.
Communicated by Yen-Hung Chen.
+Corresponding author.
*This work was supported in part by Ministry of Science and Technology under Grand No. MOST 111-2410-H-A49-022-MY2.

a persistent *internal opinion* remaining unchanged. In this model, the individual is always affected by her inherent belief, and the dynamic converges to an unique equilibrium point, which is not necessarily a consensus reached by the society. Some costs may occur when the disagreement exists, such as the collision or time wasting to meet the consensus. The work by Bindel *et al*. [3] thus proposed that the updating rule given by Friedkin and Johnsen can be equivalently viewed as each individual in a network updating its expressed opinion to minimize her cost, which consisted of the disagreement between its expressed opinion and those of its friends and the difference between its expressed opinion and internal opinion.

Formally, a social network is mapped to a weighted graph $G = (V, E)$, where V is the set of n nodes as selfish individuals, and E is the set of edges, each of which represents a relationship between a pair of nodes. The edge weight $w_{ij} \geq 0$ represents the extent of the influence from j to i. w_{ii} is the weight on an individual's internal opinion. Each node holds an internal opinion s_i, which is a constant unchanged during the updating dynamic. Each node has an expressed opinion z_i, which is updated every time step. Both of s_i and z_i are continuous. The cost function is defined as $C_i(\mathbf{z}) = w_{ii}(z_i - s_i)^2 + \sum_{j \in N(i)} w_{ij}(z_i - z_j)^2$, where $\mathbf{z}$ is the vector of expressed opinions of all nodes, and $N(i)$ represents the set of the neighbors of i. The goal of each node is to minimize its cost. The social cost $C(\mathbf{z})$ is defined as the sum of the costs $\sum_{i=1}^{n} C_i(\mathbf{z})$. Given an expressed opinion vector $\mathbf{z} = (z_i)_i$, and another social objective is the total opinion in the entire network $g(\mathbf{z})$ as $g(\mathbf{z}) = \sum_{i=1}^{n} z_i$.

Our Results. In some types of social networks, there exists a single opinion leader, like a celebrity, that draws the attention of all individuals. For example, an in-directed star network. When considering the social cost in this kind of network, the stubborn leader can affect everyone while its own opinion remaining unchanged. If the opinion of the leader is extremely different from the public, the price of anarchy for the utilitarian social cost might be unbounded. As for the other social objective, the goal of the so-called *opinion maximization* problem is to maximize the total expressed opinion, which has been shown to be NP-complete [4].

Thus, this work aims to optimize the two different social objectives at equilibrium through controlling some individuals. This is usually called *Stackelberg games*, in which a centralized authority or the *leader* is allowed to assign the strategies to a subset of individuals. The *Stackelberg strategies* of the centralized authority are the algorithms to select a subset of individuals (called the *controlling set*) and decide the actions for them to take in order to palliate the cost caused by the selfish behavior of the uncontrolled (free) individuals. We give some combinatoral algorithms and linear program rounding algorithms as Stackelberg strategies for approximately optimizing the objective of social cost (on special cases) and the objective of total expressed opinion (on general directed graphs), respectively.

1.1 Related Work

The study of *opinion formation games* was pioneered by Bindel *et al*. [3], which described the model by Friedkin and Johnsen [2] as games and defined a quadratic individual cost function to quantify the cost of the lack of agreement. They analyzed the quality of an equilibrium in terms of the *price of anarchy* [5], which means the ratio of social cost under the worst equilibrium to the optimal solution. Without any Stackelberg strategies, the upper bound for the ratio is $9/8$ in undirected graphs yet can be unbounded in directed graphs. Nonetheless, some good bounds can be obtained in Eulerian graphs, such as 2 in directed cycles and $d+1$ in d-regular Eulerian graphs; in [6], bounds on the price of anarchy were established beyond directed Eulerian graphs.

The term opinion maximization was coined by Gionis *et al*. [4], indicating to maximize the total expressed opinion (*i.e*., the overall favorable opinion) by assigning the maximal opinion for some individuals to express in a social network, for example, to employ some famous bloggers to recommend the product on their blogs. They introduced an alternative way to compute the equilibrium profile by connecting the problem to absorbing random walks. Moreover, they used greedy algorithms for the submodular objective of total expressed opinion and suggested some heuristics to improve the running times for large data.

Also to maximize the sum of individuals' opinions, Ahmadinejad *et al*. [7] had other goals. Apart from the Gionis' work that controls the expressed opinions of the individuals, this work aimed to reach some objectives by changing the initial opinions of some opinion leaders. There, they dealt with a set of problems that have different objectives other than maximizing the total expressed opinion, such as maximizing the number of individuals whose opinions are bigger than a threshold and maximizing the overall opinions of a target group of individuals while minimizing the size of the controlling set.

Although both [4, 7] identified the best subset of participants to control just as we aim to do, one of our social objectives (the utilitarian social cost) is different from theirs. The utilitarian social cost is more difficult to deal with than the linear social cost of total expressed opinion. One of the combinatorial algorithms that we consider is inspired by the work of Roughgarden [8] and the work of Fotakis [9]. The *Largest-Latency-First* strategy selects the players with the largest costs to control. In these works, the strategies are used for scheduling games and congestion games, and the bounds on the approximation ratio or equivalently, the price of anarchy, are also given in [10].

2. PRELIMINARIES

We measure an equilibrium that is induced with certain Stackelberg strategy by the social objectives. We provide the definitions of equilibrium, followed by the definition of the combinatorial optimization problem of opinion optimization, which we study in this paper in terms of two different social objectives.

Nash Equilibrium

In our work, the expressed opinion decided by each uncontrolled node depends on the individual cost. A Nash equilibrium is reached if each node expresses an opinion z_i that satisfies $C_i(z_i, \mathbf{z}_{-i}) \leq C_i(z'_i, \mathbf{z}_{-i})$, where $\mathbf{z}_{-i}$ represents all node opinions but i, and z'_i is any other opinion different from z_i. Thus, given the initial opinion s_i and the weight w_{ij} for the volume of influence from i to j, as the individual cost defined by $w_{ii}(z_i - s_i)^2 + \sum_{j \in N(i)} w_{ij}(z_i - z_j)^2$, each node updates its expressed opinion by the rule,

$z_i = \frac{w_{ii}s_i + \sum_{j \in N(i)} w_{ij} z_j}{w_{ii} + \sum_{j \in N(i)} w_{ij}}$.

2.1 Opinion Optimization

We study the following problem: given a graph $G = (V, E)$, select a set and assign the expressed opinions to them such that the different objectives could be optimized. Note that the uncontrolled nodes still update their opinions selfishly. Two social objectives are used in this paper.

2.1.1 Utilitarian social cost

Formally, the social cost with respect to a given set K is $\overline{C}(K) = \min_{\overline{\mathbf{z}}_K} C(\overline{\mathbf{z}}_K, \overline{\mathbf{z}}_{-K})$, where $\overline{\mathbf{z}}_K$ is a $|K|$-sized vector defined as the opinions expressed by the nodes in set K, which is assigned by some Stackelberg strategy, and the $|V \setminus K|$-sized vector $\overline{\mathbf{z}}_{-K}$ represents the opinions expressed by the free nodes. We aim to construct the set K^* such that $\overline{C}(K^*) = \min_{K \subseteq V} \overline{C}(K)$.

We propose combinatorial algorithms as Stackelberg strategies to approximately minimize the social cost. In Section 3, we introduce three algorithms and discuss approximation ratios for special cases due to the difficulty of showing submodularity for general directed graphs.

2.1.2 Total expressed opinion: Sum of expressed opinions

Referring to the work of Ginois *et al*. [4], the goal of opinion maximization, which they called a "campaign problem", is to make the sum of opinion as large as possible under a constraint limiting the number of controlled nodes in a network. Given an expressed opinion vector $\mathbf{z} = (z_i)_i$, we can define the total expressed opinion in the entire network $g(\mathbf{z})$ as $g(\mathbf{z}) = \sum_{i=1}^{n} z_i$.

The goal of the opinion maximization problem is to maximize g. We assume that influence spread counts on selecting a set of nodes T of at most k nodes. With selecting the set T to maximize g, we can let $g(\mathbf{z} \mid T)$ represent the sum of expressed opinions in a social network. The value of z_i(expressed opinion) for all nodes in T are fixed to 1 to obtain the Nash equilibrium vector z. We emphasize that our inputs are the weighted graph $G = (V, E)$, the internal opinions s_i of all nodes i, and the maximum number k of nodes that we can choose to set their external opinions to 1.

Note that Ginois *et al*. [4] have proven that $g(\mathbf{z} \mid T)$ is monotone and submodular, so a greedy algorithm has been proposed with an approximation guarantee of $(1 - 1/e)$. Nevertheless, we have studied opinion maximization in directed acyclic graphs through a totally different approach, which is mixed integer programming (not necessarily efficiently) for exact solutions and linear program randomized rounding (efficiently) for approximations in [11]. We will generalize this approach for any directed graphs in Section 4. The notations used in this paper are summarized in Table 1.

Table 1. Summary of the notations.

Notations	Definitions
$G = (V, E)$	weighted (directed) graph of size n
w_{ij}	influence weight from j to i
$\mathbf{s} = (s_i)_i$	vector of internal opinions individually indexed by i
$\mathbf{z} = (z_i)_i$	vector of expressed opinions individually indexed by i
$C_i(\mathbf{z})$	individual cost of i given the expressed opinions
$C(\mathbf{z})$	utilitarian social cost given the expressed opinions
$\overline{C}(K)$	utilitarian social cost by the optimal expressed opinions $\overline{\mathbf{z}}_K$ given the controlled subset
$g(\mathbf{z})$	K value of total expressed opinion
$g(\mathbf{z} \mid T)$	value of the total expressed opinion given a controlled subset T

3. COMBINATORIAL ALGORITHMS (APPROXIMATELY) MINIMIZING THE UTILITARIAN SOCIAL COST

3.1 Stackelberg Strategies

3.1.1 Controlling the nodes violating the conditions

The first strategy to decide the controlling set is inspired by the result in [6]. Since the upper bounds on price of anarchy might be unbounded if there exist some nodes who violates the conditions, we have motivation to select all the nodes whose ratio $\frac{w_{ii}+\sum_{j\neq i} w_{ij}}{\sum_{j\neq i} w_{ji}}$ is less than $1+1/\varepsilon$ to control, where $\varepsilon > 0$ is the largest possible number such that the conditions hold for all uncontrolled nodes. Yet, the size of the controlling set can be unbounded this way. The expressed opinions to assign to the controlling set is shown in Section 3.2.1.

3.1.2 Greedy algorithm

Next, we use a greedy algorithm to decide the set to control and compute the optimal values of opinions to assign. The algorithm constructs the controlling set K of k nodes by selecting the node that can improve the result the most at each step iteratively. That is, define K^t as the set constructed by the greedy algorithm at time t and $K^0 \to \emptyset$. For $1 \le t \le k-1$, $K^t \cup \{x\} \to K^{t+1}$ in which $\{x\} = \arg\max_{\{y\}\in V\setminus K^t} [\overline{C}(K^t) - \overline{C}(K^t \cup \{y\})]$.

After the set is constructed, the optimal opinions for nodes in K to express could be shown as a closed form. Actually, the form is for any given controlling set. We first re-sort all nodes according to whether they are selected or not, and define two n-sized vectors, $\mathbf{z}_{-K}$ and $\mathbf{z}_K$. The former is composed of the expressed opinions of the free nodes $\overline{\mathbf{z}}_{-K}$ and a $|K|$-sized zero vector, *i.e.*, $\mathbf{z}_{-K} = (z_1, z_2, ..., z_{|V\setminus K|}, 0, ..., 0)$. In contrast, the latter is composed by a zero vector of length $|V \setminus K|$ and $\overline{\mathbf{z}}_K$, which means $\mathbf{z}_K = (0, ..., 0, z_{|V\setminus K|+1}, z_{|V\setminus K|+2}, ..., z_n)$. Observe that $\mathbf{z} = \mathbf{z}_{-K} + \mathbf{z}_K$. Thus, we need to solve $\mathbf{z}_K^* = \arg\min_{\mathbf{z}_K} C(\mathbf{z}_{-K} + \mathbf{z}_K)$. Referring to the work of Bindel *et al*. [3], let $w_{ii} = 1$ for all i.

Theorem 1 *We first define an $n \times n$ matrix L by setting $L_{i,i} = \sum_{j\in N(i)} w_{i,j}$ and $L_{i,j} = -w_{i,j}$, then define a matrix A of the same size, in which $A_{i,i} = \sum_{j\in N(i)} w_{i,j} + w_{j,i}$ and*

$A_{i,j} = -w_{i,j} - w_{j,i}$ for $i \neq j$, and also use an $n \times |K|$ matrix M defined as $M_{n-|K|+i,i} = 1$ and 0 for remaining entries. The values of opinions $\overline{\mathbf{z}}_K$ assigned to the controlling set $K \subseteq V$ could be written as a closed form. That is, for a vector $\mathbf{s}$ of internal opinions

$$\overline{\mathbf{z}}_K^* = [M^\top (L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}M]^{-1}M^\top (L_{-K}+I)^{-\top}[\mathbf{s}-(A+I)(L_{-K}+I)^{-1}\mathbf{s}].$$

***Proof*:** We first rewrite the social cost to a function of $\mathbf{z}_K$, the vector of assigned opinions. Then, by taking the derivatives, the optimal $\mathbf{z}_K$ could be obtained. The social cost could be thus rewrite as $C(\mathbf{z}) = \sum_i[\sum_{j\neq i} w_{ij}(z_i - z_j)^2 + (z_i - s_i)^2] = \mathbf{z}^\top A\mathbf{z} + \|\mathbf{z}-\mathbf{s}\|^2$.

To write the social cost of $\mathbf{z}_K$, recall the opinion formation dynamics. For each free node, the opinion is affected by the initial opinion and the opinions of the neighbors, including the free nodes and controlling nodes. Therefore we could rewrite the opinion profile of free nodes $\mathbf{z}_{-K}$ with the assigned opinions $\mathbf{z}_K$ as $\mathbf{z}_{-K} = (L_{-K}+I)^{-1}(\mathbf{s}-L_{-K}\mathbf{z}_K)$, where L_{-K} is defined as an $n\times n$ matrix that the upper $|V\setminus K|$ rows are the same as the upper $|V\setminus K|$ rows in L and the remaining entries are all 0. We now have a new expression of the opinion profile: $\mathbf{z} = \mathbf{z}_{-K}+\mathbf{z}_K = (L_{-K}+I)^{-1}(\mathbf{s}-L_{-K}\mathbf{z}_K)+\mathbf{z}_K = (L_{-K}+I)^{-1}(\mathbf{s}+\mathbf{z}_K)$.

Next, we rewrite the social cost.

$$\begin{aligned}
C(\mathbf{z}) &= \mathbf{z}^\top A\mathbf{z} + \|\mathbf{z}-\mathbf{s}\|^2 \\
&= [(L_{-K}+I)^{-1}(\mathbf{s}+\mathbf{z}_K)]^\top A[(L_{-K}+I)^{-1}(\mathbf{s}+\mathbf{z}_K)] + \|(L_{-K}+I)^{-1}(\mathbf{s}+\mathbf{z}_K)-\mathbf{s}\|^2 \\
&= [(L_{-K}+I)^{-1}(\mathbf{s}+\mathbf{z}_K)]^\top (A+I)[(L_{-K}+I)^{-1}(\mathbf{s}+\mathbf{z}_K)] \\
&\quad - 2[(L_{-K}+I)^{-1}(\mathbf{s}+\mathbf{z}_K)]^\top \mathbf{s} + \mathbf{s}^\top\mathbf{s} \\
&= \mathbf{s}^\top[(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}]\mathbf{s} + \mathbf{s}^\top[(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}]\mathbf{z}_K \\
&\quad + \mathbf{z}_K^\top[(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}]\mathbf{s} + \mathbf{z}_K^\top[(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}]\mathbf{z}_K \\
&\quad - 2\mathbf{s}^\top(L_{-K}+I)^{-\top}\mathbf{s} - 2\mathbf{z}_K^\top(L_{-K}+I)^{-\top}\mathbf{s} + \mathbf{s}^\top\mathbf{s}.
\end{aligned}$$

To take derivatives with the dimensions that we control, we use an $n\times|K|$ matrix M defined as $M_{n-|K|+i,i} = 1$ and 0 for remaining entries. Using M, we can define a $|K|$-dimension vector $\overline{\mathbf{z}}_K$ such that $\mathbf{z}_K = M\overline{\mathbf{z}}_K$. Therefore, the social cost becomes

$$\begin{aligned}
C(\mathbf{z}) &= \mathbf{s}^\top[(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}]\mathbf{s} + \mathbf{s}^\top[(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}]M\overline{\mathbf{z}}_K \\
&\quad + \overline{\mathbf{z}}_K^\top M^\top[(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}]\mathbf{s} + \overline{\mathbf{z}}_K^\top M^\top[(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}]M\overline{\mathbf{z}}_K \\
&\quad - 2\mathbf{s}^\top(L_{-K}+I)^{-\top}s - 2\overline{\mathbf{z}}_K^\top M^\top(L_{-K}+I)^{-\top}\mathbf{s} + \mathbf{s}^\top\mathbf{s}.
\end{aligned}$$

Taking the first derivatives with respect to $\mathbf{z}_K$,

$$\begin{aligned}
\frac{\partial C(\mathbf{z})}{\partial\overline{\mathbf{z}}_K} &= [\mathbf{s}^\top(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}M]^\top + [M^\top(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}\mathbf{s}] \\
&\quad + [M^\top(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}M]^\top\overline{\mathbf{z}}_K + [M^\top(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}M]\overline{\mathbf{z}}_K \\
&\quad - 2M^\top(L_{-K}+I)^{-\top}\mathbf{s} \\
&= 2M^\top(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}\mathbf{s} + 2M^\top(L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}M\overline{\mathbf{z}}_K \\
&\quad - 2M^\top(L_{-K}+I)^{-\top}\mathbf{s}.
\end{aligned}$$

Let $\frac{\partial c(\mathbf{z})}{\partial \overline{\mathbf{z}}_K} = 0$, we could finally obtain the closed form of optimal opinions to assign to the controlling set K,

$$\overline{\mathbf{z}}_K^* = [M^\top (L_{-K}+I)^{-\top}(A+I)(L_{-K}+I)^{-1}M]^{-1}M^\top (L_{-K}+I)^{-\top}[\mathbf{s}-(A+I)(L_{-K}+I)^{-1}\mathbf{s}].$$

3.1.3 Largest-cost-first algorithm

In the largest-cost-first algorithms, similar to the greedy algorithms, we construct the controlling set K of k nodes. The difference is that the selected set is equal to the set of the k individuals with the highest cost in the optimal solution of expressed opinions $\mathbf{o}$, and force them to express the values in $\mathbf{o}$.

3.2 Upper Bounds

3.2.1 Analysis for controlling the nodes violating the conditions

Recall one of the results in [6]. Given a graph $G = (V,E)$ in which all $i \in V$ satisfy the conditions $(\sum_{j\neq i} w_{ij} + w_{ii})/\sum_{j\neq i} w_{ji} \geq 1+\varepsilon$, the upper bound on the price of anarchy is $1+1/\varepsilon$, where $\varepsilon > 0$ is the largest possible number such that the conditions hold for all uncontrolled nodes. For the arbitrary graphs, by the algorithms, nodes who violates the conditions are controlled. In other words, all of the free (uncontrolled) nodes satisfy the condition. We develop the bounds using this strategy.

Theorem 2 *For any graph G, the approximation ratio is upper bounded by* $1+1/\varepsilon$ *when the nodes whose*

$$\frac{\sum_{j\neq i} w_{ij} + w_{ii}}{\sum_{j\neq i} w_{ji}} < 1+\varepsilon$$

are all under controlled, where $\varepsilon > 0$ *is the largest possible number such that the conditions hold for all uncontrolled nodes.*

***Proof*:** We have the inequality that $1+1/\varepsilon \geq \sum_{i\in V\setminus K} C_i(\overline{\mathbf{z}}_K, \overline{\mathbf{z}}_{-K})/\sum_{i\in V\setminus K} C_i(\mathbf{o})$.

Then, consider the situation when the individuals in controlling set K could receive some payment as compensation for expressing their assigned opinions. The payment could be seen as the cost for the centralized mechanism. Here, we pay the amount equal to the cost that occurs in optimal solution $\mathbf{o}$ to those in set K, and we have the inequality

$$\begin{aligned} 1+1/\varepsilon &\geq \frac{\sum_{i\in V\setminus K} C_i(\overline{\mathbf{z}}_K, \overline{\mathbf{z}}_{-K})}{\sum_{i\in V\setminus K} C_i(\mathbf{o})} \\ &\geq \frac{\sum_{i\in V\setminus K} C_i(\overline{\mathbf{z}}_K, \overline{\mathbf{z}}_{-K}) + \sum_{j\in K} C_i(\mathbf{o})}{\sum_{i\in V\setminus K} C_i(\mathbf{o}) + \sum_{j\in K} C_i(\mathbf{o})} = \frac{\sum_{i\in V\setminus K} C_i(\overline{\mathbf{z}}_K, \overline{\mathbf{z}}_{-K}) + \sum_{j\in K} C_i(\mathbf{o})}{\sum_{i\in V} C_i(\mathbf{o})}. \end{aligned}$$

The upper bounds on the approximation ratio with this method is obtained.

3.2.2 Analysis for the greedy algorithm on special cases

According to [12], for minimizing a supermodular objectives, that is, the objective function satisfies that, for any set $K \subseteq M \subseteq V$ and $y \notin M$, $\overline{C}(K) - \overline{C}(K\cup\{y\}) \geq \overline{C}(M) - \overline{C}(M\cup\{y\})$, the upper bound on the approximation ratio of the greedy algorithms is $(e^t - 1)/t$, where the parameter t represents the steepness of the function. More precisely, $t = \max_{d_x(\{x\})} \frac{d_x(\{x\})-d_x(V)}{d_x(\{x\})}$, where $d_x(X) = \overline{C}(X\setminus\{x\}) - \overline{C}(X)$.

Lemma 3 ([12]) *For the cost-minimization problems, if the objective is a supermodular set function, an approximation ratio of the greedy algorithms is* $\overline{C}(K) \leq \frac{e^t-1}{t}\overline{C}(K_{OPT})$*, where* K *is the set constructed by the greedy algorithms and* K_{OPT} *represents the best* k*-sized set that can minimize the social cost.*

In a 2^k-ary structure of depth $\log_{2^k} n$ in which all edges are directed toward the root, the root has internal opinion 1 and the other nodes' are 0. For any nodes b at layer p in the graphs, there exists a sub-tree of which root is b. We define a as the parent of node b. The term z_b denotes the expressed opinion of node b no matter whether it's free or under controlled. At Nash equilibrium, the nodes in this sub-tree at layer $p+i$ has opinions $z_b \times 2^{-i}$, and the cost of each node except for the root is $(z_b \times 2^{-i} - 0)^2 + (z_b \times 2^{-i} - z_b \times 2^{-(i-1)})^2 = z_b^2 \times 2^{1-2i}$. The number of nodes in the sub-tree at layer $p+i$ is 2^{ik}. Thus, the social cost at Nash equilibrium is $\sum_i^{\log_{2^k} n-p} z_b^2 \times 2^{1-2i}2^{ik} = 2z_b^2 \sum_i^{\log_{2^k} n-p} 2^{(k-2)i}$. When node b is controlled, the expressed opinions should minimize

$$(z_b - 0)^2 + (z_b - z_a)^2 + 2z_b^2 \sum_i^{\log_{2^k} n-p} 2^{(k-2)i}.$$

By taking the derivatives, the opinion assigned is $z_b = z_a / 4\sum_i^{\log_{2^k} n-p} 2^{(k-2)i}$. As k growing, such as $k = \log\log n$, z_b approximates to 0, leading all nodes in this sub-tree to change their value to 0, and the cost in the sub-tree decreases.

To prove the supermodularity holds in this case, we start from the left hand side of the inequality. Given a set K, there are some sub-trees "occupied" by the nodes in K. If the node y is in any occupied sub-tree, controlling one more node y makes no difference from the original set. If y is not in any occupied sub-tree, it could lead nodes influenced by its opinion and decrease the social cost. On the other hand, if the node in $M \setminus K$ lays in the sub-tree of y, the difference between controlling M and $M \cup \{y\}$ would be smaller because of the reduction of free nodes in the sub-tree of y. Therefore in these cases, the objective function is supermodular.

3.2.3 Analysis for the largest-cost-first algorithm on special cases

The same, consider the 2^k-ary tree structure. The algorithms simply control the root node and assign the opinion 0 when k is large enough. In this case, $\overline{C}(K_{LCF}) = \overline{C}(V) + \alpha$, where K_{LCF} denotes the set constructed by the largest-cost-first algorithms, and the term α represents the difference that occurs from the optimal solution of the root, which is close to 0 but not equal to 0. The approximation guarantee of the greedy algorithms is then

$$\overline{C}(K) \leq \frac{e^t-1}{t}\overline{C}(K_{OPT}) \leq \frac{e^t-1}{t}\overline{C}(K_{LFC}) \leq \frac{e^t-1}{t}(\overline{C}(V) + \alpha),$$

where the second inequality comes from the definition of K_{OPT}, and α would decrease and approach 0 as k increases.

Proposition 4 *For the specific* 2^k*-ary tree* G *in which all edges point toward the root. With the internal opinion is* 1 *for the root and* 0 *for the other nodes, the objective* $\overline{C}(K)$ *is a* *supermodular set function. As $k = o(\log n)$, for example, $k = \log\log n$, the approximation ratio is upper bounded by $(e^t - 1)/t$. Also, $k = \omega(1)$.*

3.3 Lower Bound on the Size of Controlling Set

An interesting question to ask is that, at least how many nodes to control to get a bounded price of anarchy in directed graphs. We could conclude that in an in-directed star or tree graphs, at least we have to control one individual. Therefore, $\sqrt{n}$ is be a lower bound for graphs consisting of $\sqrt{n}$ stars.

4. LINEAR PROGRAM ROUNDING ALGORITHMS (APPROXIMATELY) MAXIMIZING THE TOTAL EXPRESSED OPINION

On directed trees and directed acyclic graphs, the influence between nodes is somehow directional; in other words, the external opinion of each node will *not* be updated repeatedly and cyclically in the process of opinion formation. This is the key reason that we can model the opinion maximization problem as mixed integer program (MIP) and linear program (LP) relaxations in [11]. Surprisingly, we find that this formulation actually does *not* depend on being acyclic so we can generalize the approach to any directed graphs quite easily. The difference is that there may or may not be any nodes $b \in i$ that have no out-degree in directed (cyclic) graphs. Note that MIP is an exact algorithm (not necessarily efficient) and LP randomized rounding algorithms are efficient approximation ones. For LP randomized rounding Algorithm 2, we give an analysis to bound the approximation ratio with high probability. Note that in the following, our approach is shown for unweighted directed graphs for the simplicity of exposition, it can easily carry over for weighted directed graphs.

4.1 Mixed Integer Linear Program

Note that each node v_i with its internal opinion s_i has multiple parent nodes $v_{j_1}, ...,$ $v_{j_{|N(i)|}}$ and $w_{ij} = 1$ for all i, j. Nodes $v_0, ..., v_{b-1}$ are the nodes with no outgoing edges.

$$\max \sum_i z_i$$

$$\text{s.t.} \sum_i y_i \leq k, z_i = y_i + s_i - s_i y_i \forall i \in [b], z_i = y_i + \frac{s_i + \sum_{j \in N(i)} z_j}{|N(i)| + 1} - \frac{s_i y_i}{|N(i)| + 1} - \frac{\sum_{j \in N(i)} x_{i,j}}{|N(i)| + 1} \forall i > b$$

$$x_{i,j} \leq z_j \forall i, j, z_i \in [0,1] \forall i, y_i \in \{0,1\} \forall i, x_{i,j} \in [0,1] \forall i, j$$

Note that if $y_i = 1$ meaning that node v_i is selected, then the expressed opinion of node v_i is $z_i = 1 + \frac{\sum_{j \in N(i)} z_j - x_{i,j}}{|N(i)|+1} \geq 1$ where x_i is a nonnegative fractional auxiliary variable for setting z_i; since $z_i \in [0,1]$, $z_i = 1$. If $y_i = 0$ meaning that node v_i is not selected, $z_i = \frac{s_i + \sum_{j \in N(i)} z_j - x_{i,j}}{|N(i)|+1}$; since the objective is $\sum_i z_i$, $x_{i,j} = 0$ for all $j \in N(i)$ helps at maximum and thus $z_i = \frac{s_i + \sum_{j \in N(i)} z_j}{|N(i)|+1}$.

4.2 LP Randomized Rounding Algorithms

Randomized rounding is a method that relaxes an integer program (IP) and converts the fractional solution obtained from the relaxation of IP to an approximation solution. We gave mixed integer linear program (MIP) for problem of opinion maximization in the previous subsection. In this subsection, we inherit the mixed integer program that we gave. Let y_i be fractional $\in [0,1]$ for each node i to relax the mixed integer program to a linear program.

Each node v_i with its internal opinion s_i has multiple parent nodes $v_{j_1},...,v_{j_{|N(i)|}}$ and $w_{ij} = 1$ for all i,j. Nodes $v_0,...,v_{b-1}$ are the nodes with no outgoing edges. The LP relaxation of the MIP for a directed graph is as follows.

$$\max \sum_i z_i$$

$$\text{s.t. } \sum_i y_i \leq k, z_i = y_i + s_i - s_i y_i \forall i \in [b], z_i = y_i + \frac{s_i + \sum_{j \in N(i)} z_j}{|N(i)|+1} - \frac{s_i y_i}{|N(i)|+1} - \frac{\sum_{j \in N(i)} x_{i,j}}{|N(i)|+1} \forall i > b$$

$$x_{i,j} \leq z_j \forall i,j, z_i \in [0,1] \forall i, y_i \in [0,1] \forall i, x_{i,j} \in [0,1] \forall i,j$$

With the variables and the solution that we obtain from linear programming, we use randomized rounding to convert some values of variables to integers in $\{0,1\}$ and obtain an approximation solution. We will give two LP randomized rounding algorithms in the following.

4.2.1 LP randomized rounding Algorithm 1

Algorithm 1 : LP Randomized Rounding Algorithm 1 (Repetition without Replacement)

Solve the LP above to get the optimal fractional solution $\{\bar{x}_i\}_i$, $\{\bar{y}_i\}_i$, $\{\bar{z}_i\}_i$ and the optimal objective value OPT_f.
Initially set the selected set $S := \phi$ and set $V' := V$.
while Repeat the following process k times **do**
  Select a node v_i to set $z_i = y_i = 1$ with distribution $\{\frac{\bar{y}_i}{\sum_i \bar{y}_i}\}_i$ of nodes in V'.
  Add the selected node to S.
  Remove the selected node from V'.
end while
Efficiently compute Nash equilibria according to S ([4]).

The concept of LP randomized rounding algorithm 1 is selecting the node that we want to add in set S with distribution $\{\frac{\bar{y}_i}{\sum_i \bar{y}_i}\}_i$ and setting their expressed opinion $z_i = 1$. We select one node in each round until k nodes are selected. Regarding the performance of LP randomized rounding algorithm 1, we have shown how close it is to the optimal solution through experiments in [11] for directed acylic graphs. The result would carry over to general directed graphs.

4.2.2 LP randomized rounding Algorithm 2

Algorithm 2 : LP Randomized Rounding Algorithm 2 (Independence)

Solve the LP above to get the optimal fractional solution $\{\bar{x}_i\}_i$, $\{\bar{y}_i\}_i$, $\{\bar{z}_i\}_i$ and the optimal objective value OPT_f.
while Repeat the following process $c \log n$ rounds **do**
  for $v_i \in V$ **do**
    Round z_i to 1 with probability of $\bar{y}_i$ and set it to $\frac{s_i + \sum_{j \in N(i)} \bar{z}_j}{|N(i)|+1}$ with probability of $1 - \bar{y}_i$.
    Round y_i to 1 if $z_i = 1$ and to 0 otherwise.
  end for
end while
for $z_i \in V$ **do**
  Set z_i to the value at the same arbitrary round.
end for

For the analysis, we upper bound the probability of violating the constraints so we can ensure that LP randomized rounding algorithm 2 has high probability of finding a feasible solution with a good approximation ratio. Note that the algorithm and analysis work for general directed graphs.

Theorem 5 *Algorithm 2 outputs a feasible solution and has an objective value at least* $(1-\delta)(OPT-\varepsilon)$ *with high probability for some constant* $0 < \delta \leq 1$ *and a constant* $\varepsilon \geq 0$ *defined in terms of the optimal fractional solution.*

***Proof*:** In one of the $c \cdot \log n$ rounds by Markov's inequality

$$\mathbf{Pr}[\sum_i y_i \geq k+1] \leq \frac{\mathbf{E}[\sum_i y_i]}{k+1} = \frac{\sum_i \mathbf{E}[y_i]}{k+1} = \frac{\sum_i \bar{y}_i}{k+1} \leq \frac{1}{1+1/k}. \tag{1}$$

For the $c \cdot \log n$ times of the process, the probability that $\sum_i y_i \geq k+1$ every round is at most for some proper constant $c' > 0$

$$(\frac{1}{1+1/k})^{c \cdot \log n} \leq \frac{1}{c'n}. \tag{2}$$

The expected objective is

$$\begin{aligned}
& \sum_i \mathbf{E}[z_i] \\
= \; & \sum_i \bar{y}_i \cdot 1 + (1-\bar{y}_i)\frac{s_i + \sum_{j \in N(i)} \bar{z}_j}{|N(i)|+1} \geq \sum_i \bar{y}_i + \frac{s_i + \sum_{j \in N(i)} \bar{z}_j}{|N(i)|+1} - \frac{s_i \bar{y}_i}{|N(i)|+1} - \frac{\sum_{j \in N(i)} \bar{z}_j}{|N(i)|+1} \\
= \; & \sum_i \bar{y}_i + \frac{s_i + \sum_{j \in N(i)} \bar{z}_j}{|N(i)|+1} - \frac{s_i \bar{y}_i}{|N(i)|+1} - \frac{\sum_{j \in N(i)} \bar{x}_{ij}}{|N(i)|+1} - \frac{\sum_{j \in N(i)} \bar{z}_j - \bar{x}_{ij}}{|N(i)|+1} \\
= \; & \sum_i \bar{z}_i - \varepsilon = OPT_f - \varepsilon \\
\geq \; & OPT - \varepsilon \geq k - \varepsilon
\end{aligned}$$

for $\varepsilon = \sum_i \frac{\sum_{j\in N(i)} \bar{z}_j - \bar{x}_{ij}}{|N(i)|+1}$. By the Chernoff bound, for some constant $0 < \delta \leq 1$

$$\mathbf{Pr}[\sum_i z_i \leq (1-\delta)\mathbf{E}[\sum_i z_i]] \leq e^{-\frac{\mathbf{E}[\sum_i z_i]\delta^2}{2}} \leq \frac{1}{e^{(k-\varepsilon)\delta^2/2}}.$$

Coming all these bad events such that $\sum_i y_i \geq k+1$ every time or $\sum_i z_i \leq (1-\delta)\mathbf{E}[\sum_i z_i]$, we have that the objective value by the algorithm's feasible solution is at least $(1-\delta)(OPT-\varepsilon)$ with probability of at least

$$1 - \frac{1}{e^{(k-\varepsilon)\delta^2/2}} - \frac{1}{c'n}.$$

5. CONCLUSIONS AND FUTURE WORK

We consider that, if we control a set of people and force them to express some specific opinions, how the social cost would change. We introduced three algorithms as Stackelberg strategies: controlling those who violates the conditions, the greedy algorithm, and the largest-cost-first algorithms. In the first two methods, a closed form of assigned opinion was given. In the last method, the opinion to assign is the values in the optimal solution. We propose the MIP and two different LP randomized rounding algorithms to solve the problem of opinion maximization proposed by Gionis *et al*. [4] for general directed graphs. With mixed integer programming, the optimal solution can be obtained. With two LP randomized rounding algorithms, we obtain approximations. We proved that we can obtain an approximate solution with high probability via the LP randomized independently rounding algorithm.

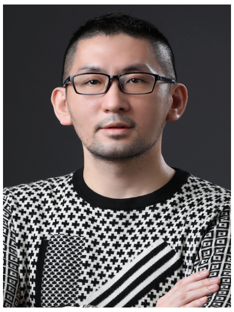
As for future work, the cost for bribing the individual to express the assigned opinion is not considered in this work. Taking the cost into consideration, the choice of the nodes to be controlled may be different because some individuals would take higher cost for changing their minds. Different ways to control the nodes is another possible direction. For example, we may force the nodes to break off the relations, *i.e.*, deleting some edges. For the case in which the nodes refuse to change, deleting the nodes might be an efficient method. The cost of deleting the edges or nodes could also be considered. In [4, 7, 11] and this paper, there is only one corporation in the society. The case that there are two or more competitors in the society might be interesting. For example, two competitive corporations have their own goals; one might hope to maximize the $\sum_i z_i$ while the other one might expect to minimize the objectives. The existence of the Nash equilibrium, the bounds on price of anarchy, and the strategies for the two corporations to play are some issues worth exploring.

ACKNOWLEDGMENT

We would like to thank Ling-Wei Wang for useful discussions.

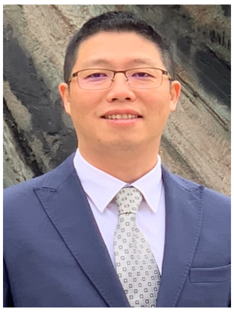
REFERENCES

1. M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, Vol. 69, 1974, pp. 118-121.
2. N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *Journal of Mathematical Sociology*, Vol. 15, 1990, pp. 193-206.
3. D. Bindel, J. Kleinberg, and S. Oren, "How bad is forming your own opinion?" *Games and Economic Behavior*, Vol. 92, 2015, pp. 248-265.
4. A. Gionis, E. Terzi, and P. Tsaparas, "Opinion maximization in social networks," in *Proceedings of SIAM International Conference on Data Mining*, 2013, pp. 387-395.
5. E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *Proceedings of the 16th International Symposium on Theoretical Aspects of Computer Science*, 1999, pp. 404-413.
6. P.-A. Chen, Y.-L. Chen, and C.-J. Lu, "Bounds on the price of anarchy for a more general class of directed graphs in opinion formation games," *Operations Research Letters*, Vol. 44, 2016, pp. 808-811.
7. A. Ahmadinejad and H. Mahini, "How effectively can we form opinions?" in *Proceedings of International World Wide Web Conference*, 2014, pp. 213-214.
8. T. Roughgarden, "Stackelberg scheduling strategies," *SIAM Journal of Computing*, Vol. 33, 2004, pp. 332-350.
9. D. Fotakis, "Stackelberg strategies for atomic congestion games," *Theoretical Computer Science*, Vol. 47, 2010, pp. 218-249.
10. V. Bilò and C. Vinci, "Stackelberg strategies for atomic congestion games," *Theory of Computing Systems*, Vol. 63, 2015, pp. 1228-1249.
11. P.-A. Chen, Y.-W. Cheng, and Y.-W. Tseng, "Mixed integer programming and LP rounding for opinion maximization on directed acyclic graphs," in *Proceedings of the 10th International Conference on Complex Networks and Their Applications*, 2021, pp. 867-878.
12. V. P. Ilév, "An approximation guarantee of the greedy descent algorithm for minimizing a supermodular set function," *Discrete Applied Mathematics*, Vol. 114, 2001, pp. 131-146.



Po-An Chen is currently an Associate Professor in the Institute of Information Management, National Yang Ming Chiao Tung University. He is generally interested in economics and computation, artificial intelligence, and operations research, specifically including algorithmic game theory, machine learning (particularly, online learning), social networks, and multiagent and distributed systems. His current focus include the performance analysis and computation of equilibria and learning algorithms in repeated games, markets or multiagent systems.

Yi-Le Chen graduated from the Institute of Information Management, National Chiao Tung University. She currently works at Taiwan Semiconductor Manufacturing Company (TSMC).



Wei Lo graduated from the Institute of Information Management, National Yang Ming Chiao Tung University.