Combinatorial Optimization Games Arise in Social Networks

Team Members:
Rui Zhang and Mustafa Sahin

May 8, 2014

Recently, the dynamic processes for the diffusion of influence has attracted significant interest from algorithmic researchers. Over the past ten years, the following viral marketing problem has attracted a significant amount of interest. Assume we want to promote a new product over a given social network and wish this product will be adopted by most people in this network. We can initialize the diffusion process by "targeting" some influential people. Then, a cascade will be caused by these initial adopters and other people start to adopt this product due to the influence they receive from earlier adopters. But how should we select these influential people who are targeted initially? Domingos and Richardson [2001] studied the problem in a probabilistic setting, and provided heuristic solutions. Kempe et al. [2003] were the first ones to model this problem as an optimization problem by using the threshold model proposed by Granovetter [1973], showed it is NP-hard to find the optimal initial set, and developed approximation algorithms for the problem. Subsequently, several different variants of this problem have been studied. Since then, several variants of this problem are studied. Among them, Chen [2009] presented the Target Set Selection (TSS) problem. Built on the TSS problem, Raghavan and Zhang [2013] proposed the Weighted Target Set Selection (WTSS) problem and Gunnec et al. [2013] studied the Least Cost Influence Problem (LCIP).

The WTSS problem is defined on a connected undirected graph $G = (V, E)$. The node set V can be interpreted as a set of people in a social network then the edge set E becomes the connections between people on the social network. For each $i \in V$, there is a threshold, denoted by g_i, which is between 1 and $\alpha(i)$, where $\alpha(i)$ denotes the degree of node i. Let g_i and b_i be the threshold value and the weight for node $i \in V$ respectively. The weight models the fact that different nodes require differing levels of effort to become initial adopters (in practice, it is reasonable to assume different persons would not always require the same amount of effort to be convinced). Initially we assume all nodes are inactive. We select a subset of nodes $A \subseteq V$, namely the target set, and for each node $i \in A$, the threshold value b_i is incurred so that i becomes active. After the target set becomes active, in each step, we update the state of the remaining nodes by the following rule: an inactive node i becomes active if at least g_i of its neighbors were active in the previous step. The goal is to find the target set A^* while ensuring that all nodes are active at the end that minimizes the total cost incurred, i.e. $\sum_{i \in A^*} b_i$. We present a Mixed Integer Programming (MIP) formulation for WTSS.

x_i is 1 if node i is selected in target set. 0 o.w. y_{ij} is 1 if node i becomes active before node j. 0 o.w.

WTSS(G): Minimize $\sum_{i \in V} b_i x_i$

$$\text{Subject To} \quad \sum_{j \in \alpha(i)} y_{ji} + g_i x_i \geq g_i \qquad \forall i \in V \tag{1}$$

$$\sum_{(i,j) \in C} y_{ij} \leq |C| - 1 \qquad \text{for each cycle } C \in G \tag{2}$$

$$x_i \in \{0, 1\} \qquad \forall i \in V \tag{3}$$

$$y_{ij}, y_{ji} \in \{0, 1\} \qquad \forall (i, j) \in E \tag{4}$$

The objective function calculates the total threshold value of the target set. By constraints (1), all nodes are eventually active. Constraints (3) ensure that for a given cycle C, one node in the cycle has to be active before the others, otherwise, it would be possible to activate nodes in a circular fashion without one node becoming active before the next node in the cycle. Constraints (3) and (4) are domain constraints.

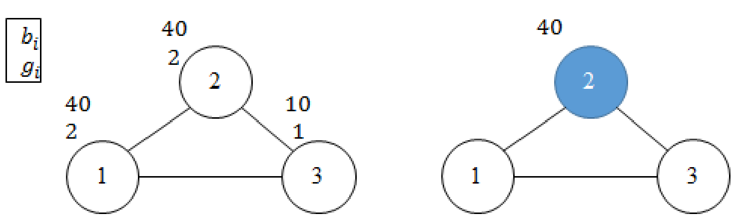

Figure 1 is an example of the WTSS problem. In the left, it is a WTSS instance. It has three persons in this network. The numbers beside a node are its weight and threshold. For instance, the weight of node 1 is 40 and its threshold is 2. In the right, it is the optimal solution for this instance. We pick node 2 and pay its weight, 40. Then, node 3 becomes adopted. After that, node 1 adopts as well. So, the whole network adopts.

The Least Cost Influence Problem (LCIP) is defined on an undirected graph $G = (V, E)$. Each node in the network makes a decision about buying a product. An individual buys the product if the utility from the product exceeds a certain threshold for that individual. Such threshold is named as *hurdle* in marketing literature. Utility from the product is calculated as the sum of the utilities from different attributes of the product and represented by b_i for node $i \in V$. We introduce two more attribute-like measures to be added to the total utility from the product. The first is the *inducement* received from the seller. Inducements are considered as endorsements for the individual to buy the product, and counted in terms of additional positive units of utility. They correspond to promotions received from sellers as free samples, discount coupons, etc. We denote the inducement received by p_i, $i \in V$. The second is the additional utility from using the same product as one's neighbors. The additional utility is dependent on the person influencing (owning the product) and the person being influenced (making a decision about whether to buy the product). We denote this *influence factor* by d_i which captures how much a neighbor of i influences node i if the neighbor has already adopted the product. d_i is defined over all nodes $i \in V$.

The objective of LCIP is to minimize the total amount of incentives given while ensuring that all of the market will adopt the product at the end. We present an MIP formulation for LCIP.

In order to take into account social network effects, and to capture the order of buying among customers, we introduce time periods, $t = 1, 2, \ldots, T$ on when the product is bought by an individual. We do not allow for churns and once a product is adopted, it is kept the entire time. So the critical part (in terms of affecting neighbor's decisions) is when it is the first time to buy the product which we denote by v_{it} and equals 1 if node i adopts at time period t, and 0 otherwise. We also introduce an auxiliary binary decision variable y_{it} taking the value 1 if node i has adopted the product by time period t, 0 otherwise.

$$\textbf{LCIP}(G)\textbf{:}\ \text{Minimize} \sum_{i \in V} p_i$$

$$\text{Subject to } p_i + \sum_{j \in \alpha(i)} d_i y_{j,(t-1)} \geq b_i y_{it} \qquad i \in V, t = 2, ..., T \qquad (5)$$

$$y_{it} - y_{i,(t-1)} = v_{it} \qquad i \in V, t = 2, ..., T \qquad (6)$$

$$\sum_{t=1}^{T} v_{it} = x_1 \qquad i \in V \qquad (7)$$

$$p_i \geq 0 \qquad i \in V \qquad (8)$$

$$y_{it} \in \{0, 1\}, \quad v_{it} \geq 0 \qquad i \in V, t = 1, 2, \ldots, T \qquad (9)$$

The objective function calculates the sum of the incentives given over the network. Constraints (5) provide the buying condition for each person. The hurdle, b_i, for person i is compared with the sum of the total influence over all neighbors, $\sum_{j \in \alpha(i)} d_i y_{j,(t-1)}$, and the amount of incentives received, p_i. Here, we note that the utility a person gets from using the product (without any influence from neighbors or incentives from the seller) has already been deducted and omitted from the model. Ignoring the utility solely from using the product in this model does not change the solution of which nodes to select to give incentives to or the amount of incentives to give. Using network effects and inducements, our focus is on finding the least expensive way of closing the gap between a person's hurdle and the utility from the product. Therefore, the hurdles in this model are current (or updated) hurdles (which are therefore less than the general definition of hurdle in the marketing literature). With constraints (6), we set the value of v_{it} w.r.t. to the first time node i adopts the product. Adopters do not switch to not adopting at later points of time. By summing over all periods in constraints (7), we make sure that all nodes eventually adopt the product. Note that since there are $|V|$ nodes, the number of time indices required can be limited to $T \leq |V|$.

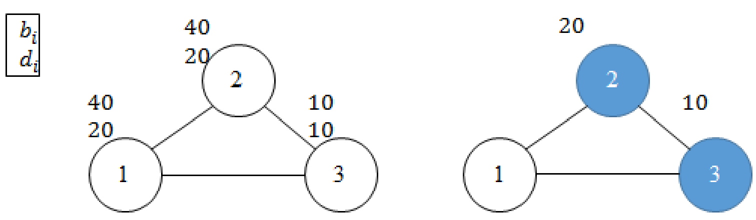

Figure 2 is an example of the LCIP. In the left, it is a LCIP instance. It has three persons in this network.

The numbers beside a node are its weight and influence factor. For instance, the weight of node 1 is 40 and its influence factor is 20. In the right, it is the optimal solution for this instance. We pay node 2 and node 3. After we pay node 3, it reduces node 1 and node 2's hurdle to 20. Then, we pay node 2. After it adopts, the node 1 adopts as well. The total cost is 30. The difference between the LCIP and the WTSS problem is that we are allow to hand out partial incentive in the LCIP. We can understand it as coupons or discounted price. But in the WTSS, you have to pay the full amount of weight to a node. In other words, the WTSS problem is the binary version and the LCIP is the fractional one.

Although these two models are considered in viral marketing setting, they can also be applied in in Epidemiological setting, see Gunnec et al. [2013]. We have a social network where each person has a risk to be affected by a kind of disease. For a person i, let r_i be the safe risk level for it. Suppose that e_{ji} denotes the risk factors of an untreated neighbor j on node i (e.g., if δ_{ji} denotes the probability of node i getting infected by an untreated neighbor j, then $e_{ji} = -\log(1 - \delta_{ji})$). Let f_i denote the reduction of the risk on node i if one of its neighbor j is treated so that its risk level is less than or equal to a threshold risk level r_j. We would like to ensure that the sum of all $e_{ji} - f_i$ for node i minus the intervention or treatment strategy z_i reduces the overall risk of node i below the threshold risk level r_i. This may be equivalently modeled by the LCIP in the marketing setting with $b_i = \sum_{j \in N(i)} e_{ji} - r_i$ and $d_i = f_i$, with a discrete set intervention or treatment strategy choices at each node (e.g. $z_i = p_i$).

There is a key difference between the marketing setting and the Epidemiological one. In the marketing setting, there is a centralizer (a company) who want to promote a new product. Hence, they are willing to pay the extra incentives to each potential customer. The cost can be considered as a marketing campaign cost. In the Epidemiological setting, each person wants to reduce his/her risk level. Then, the cost is afforded by the person who receives the treatment. However, there are many free riders. For instance, in Figure 1, person 1 and person 3 are free riders in the small example because they have not received any treatment but their risk level are reduced due the fact that only person 2 receives treatment. In the small example, the cost is 40. Hence, naturally, this leads to the question that how should we allocate the cost? Is it fair that only person 2 pays the cost?

To answer above questions, we combine cooperation game theory with these two combinatorial optimization problems. Cooperative game theory is concerned with situations in which at least two decision makers can increase their profits or decrease their cost by cooperation. To be concrete, we can think of a case where one person has the resources to make a certain product, another one has the know-how to make it, and yet a third one has the means to transport it to a market where it can be sold. Alone, none of them can generate a profit. By working together they can. A major part of the cooperative games theory is built about the question about how to allocate the total profit or costs among the group of decision makers who are willing to cooperate. Among others, combinatorial optimization games form an important subclass of cooperative games, see Nisan [2007]. We would like to ask these questions: How can we apply cooperation game theory to our model in the Epidemiological setting? Can it be tested in polynomial time whether a given instance of the game has a nonempty core? Is it possible to find an imputation in the core in polynomial time? Furthermore, we would like to explore the possibility that allocating the cost by other solution concepts such as Shapley value.

References

N. Chen. On the approximability of influence in social networks. *SIAM Journal on Discrete Mathematics*, 23(3):1400–1415, 2009.

P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66. ACM, 2001.

M. S. Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

D. Gunnec, S. Raghavan, and R. Zhang. The least cost influence problem. *Working paper*, 2013.

D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

N. Nisan. *Algorithmic game theory*. Cambridge University Press, 2007.

S. Raghavan and R. Zhang. Weighted target set selection. *Working paper*, 2013.