Maximum influence in signed social networks

Ikram Bouras[2], Rosa Figueiredo[1], Cristina Requejo[3], Michael Poss[2]

[1] University of Avignon, LIA, 84911 Avignon, France {firstname.lastname}@univ-avignon.fr
[2] University of Montpellier, LIRMM, 34090 Montpellier, France {firstname.lastname}@lirmm.fr
[3] Universidade de Aveiro, 3810-193 Aveiro, Portugal {crequejo@ua.pt}

Keywords: *Signed social networks, Influence, Bi-level programming, Integer programming*

1 Introduction

In the last years, the number of users of social networks such as Facebook, YouTube, and Twitter rapidly increased (about 2.4 Billion users for Facebook, 1.9 billion for YouTube, and 330 million users of Twitter). In these social networks, people share and receive information, advertisements, and ideas from friends or subscribers in “word-of-mouth" form of communication. From the users perspective, this provides to the users new and comfortable channels for exchanging information and expressing views and opinions [2]. From the marketing perspective, this allows the social media to penetrate all aspects of everyday lives. In this context, the study of influence propagation through a social network gained importance when deciding whether or not to adopt an innovation (such as a political idea, a new product, or medical and technological innovations).

In the literature, works studying the influence and effects of “word-of-mouth" in the promotion of new products in social networks are motivated by applications like the spreading of ideas or innovations in a network and viral marketing of products. Definitely, influence maximization has become a relevant problem on social networks.

The Maximum influence problem in social networks consists of selecting a subset of seeds (users of the network) to spreed information or ideas through given diffusion models, in order to maximize the spread of influence in the network.

2 Problem description

We consider a signed social network, which can be modeled as an undirected signed graph $G = (V, E, s)$. The set of nodes V represents the individuals of the considered network. The set of edges E represents polarized relationships among individuals such as trust and distrust relationship, each edge $\{i, j\} \in E$ in the graph is associated with a positive or a negative signs according to the type of relationship between individuals i and j.

We suppose the existence of an information to be spread in the network which can take two opposite states in $\{I_0, I_1\}$. Ideally, the owner of the information wants it to be spread at state I_0 to each node (individual) in the graph (social network).

The information can be sent at state I_0, from the owner, to a node $i \in V$ at a sending cost f_i; in this case i is called a *seed* (i.e. an individual in charge of diffusing the initial information from the owner). A penalization C_i is defined for each node i receiving the information at state I_1, i.e. the reverse of the one initially sent by the owner.

A value d_{ij} is associated with each edge $\{i, j\} \in E$. Assuming i (respectively, j) receives the information at a time t_0, then j (respectively, i) receives the information at time $t_0 + d_{ij}$ through edge $\{i, j\}$. A negative edge $\{i, j\} \in E^-$ means the information inverts as it flows on the edge, while it keeps its state whenever $\{i, j\} \in E^+$.

Maximum influence problems amount to select a subset of nodes $S \subseteq V$ in the network (seeds) to diffuse a given information to each node $i \in V \setminus S$ that minimizes the associated sending and penalization costs for all nodes in V. Once the set of seeds S is defined, the information retained by each non-seed node $i \in V \setminus S$ is the first information arriving to this node, i.e., the information retained arrives through a shortest path linking a seed $j \in S$ to i. This assumption is motivated by many psychology and marketing researches which prove the effect of the first impression on decision making, this called the “Halo effect" for positive impressions and “Horn effect" for negative ones.

The bi-level model, presented in following, selects at the first level the set of seed nodes which will receive the information directly from the owner. The second level computes the shortest paths connecting each node to the selected seed.

$$\min \quad \sum_{k \in V} f_k z_k^k + \sum_{k \in V} C_k \pi_k \tag{1a}$$

$$\text{s.t.} \quad z_i^k \in \{0,1\}, \; z_i^k \leq z_k^k, \qquad \forall i, k \in V, \tag{1b}$$

$$\left\lceil \frac{\sum_{(i,j) \in A^-} y_{ij}^k}{2} \right\rceil - \frac{\sum_{(i,j) \in A^-} y_{ij}^k}{2} \leq \pi_k \leq \left\lceil \frac{\sum_{(i,j) \in A^-} y_{ij}^k}{2} \right\rceil - \frac{\sum_{(i,j) \in A^-} y_{ij}^k}{2} + \frac{1}{2}, \qquad \forall k \in V, \tag{1c}$$

$$\min \quad \sum_{k \in K} \sum_{(i,j) \in A} d_{ij} y_{ij}^k \tag{1d}$$

$$\text{s.t.} \quad \sum_{j:(k,j) \in A} y_{kj}^k - \sum_{j:(j,k) \in A} y_{jk}^k = z_k^k - 1, \qquad \forall k \in V, \tag{1e}$$

$$\sum_{j:(i,j) \in A} y_{ij}^k - \sum_{j:(j,i) \in A} y_{ji}^k = z_i^k, \qquad \forall i \neq k \in V, \tag{1f}$$

$$y_{ij}^k \leq 1 - z_k^k, \qquad \forall (i,j) \in A, \forall k \in V. \tag{1g}$$

The objective function of the leader (1a) minimizes the sum of sending costs associated with the selected seeds and penalty costs associated with non-seed nodes. Constraints (1b) allows node i to receive information through a path beginning at node k only if k is selected as a seed node. The variable π_k defined by constraint (1c) is a binary variable which represents the state of the information arriving to the node k according to the number of negative arcs in the path of k (1 for shifted information and 0 for the original one).

Constraints (1e) and (1f) force each non-seed node to receive an information from exactly one seed. Constraints (1g) force variables y^k to be equal to 0 whenever node k is a seed.

The second level problem is a shortest path problem between all pairs of nodes. We solve the above problem exactly by reformulating it as a one-level MILP using three different types of optimality conditions [1]: KKT optimality conditions, Bellman's optimality conditions, and inequalities eliminating unfeasible paths.

These formulations have been strengthened by adding a set of valid inequalities. In addition, preprocessing and polynomial time solution were proposed for particular cases of networks. Computational experiments are performed using random instances to compare the different proposed formulations. The obtained results showed the efficiency of the formulation based on Bellman's optimality condition over the other formulations for the majority of the instances.

References

[1] Ikram Bouras, Rosa Figueiredo, Michael Poss, and Fen Zhou. On two new formulations for the fixed charge network design problem with shortest path constraints. *Computers & Operations Research*, 108:226–237, 2019.

[2] David Kempe, Jon Kleinberg, and Éva 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.