A Lagrangian Relaxation for the Generalized Target Set Selection Problem

Felipe de C. Pereira[1], **Pedro J. de Rezende**[2] , **Tallys Yunes**[3]

[1] Department of Computing, Federal University of Sergipe, São Cristóvão, SE, Brazil

[2]Institute of Computing, University of Campinas (UNICAMP), Campinas, SP, Brazil

[3]Miami Herbert Business School, University of Miami, Coral Gables, FL, USA

felipe@dcomp.ufs.br, pjr@unicamp.br, tallys@miami.edu

Abstract. *The Generalized Target Set Selection Problem (*`GTSSP`*) models the spread of information in social networks and is known to be* NP*-hard. We propose an integer programming formulation for the* `GTSSP`*, together with a promising Lagrangian relaxation approach. By relaxing a subset of constraints, we obtain a decomposable structure in which the relaxed problem can be split into a series of independent subproblems. Each subproblem corresponds to an instance of the Minimization Knapsack Problem, a weakly* NP*-hard problem solvable in pseudopolynomial time using a dynamic programming algorithm. We describe the decomposition scheme and discuss its computational complexity, highlighting its potential for use within an exact algorithm.*

1. Introduction

In the age of social media, word-of-mouth marketing plays a crucial role for promoting products through interpersonal influence. For example, when a gaming company releases a new title, it may strategically engage select individuals to initiate a cascading spread of information. This process, driven by posts and shares on social networks like Instagram and $\mathbb{X}$ (formerly Twitter), introduces important optimization challenges.

The primary objective is to minimize the cost of recruiting initial spreaders while ensuring that information reaches a broad audience. This scenario can be formulated as a combinatorial optimization problem on graphs, where vertices represent individuals, edges indicate influence relationships, and information propagates through the graph.

In this paper, we investigate the Generalized Target Set Selection Problem (GTSSP) [Ravelo and Meneses 2021], a generalization of the well-known Target Set Selection Problem (TSSP) [Chen 2009], an optimization problem that models the spread of influence in social networks. To the best of our knowledge, this version of the problem has not been studied before.

The TSSP is NP-hard and has inspired a collection of optimization problems that have been extensively studied [Ackerman et al. 2010, Shakarian et al. 2013, Spencer and Howarth 2013, Chen et al. 2013, Li et al. 2018, Raghavan and Zhang 2019, Pereira et al. 2021, Pereira and de Rezende 2023, Pereira et al. 2024]. Specifically, the GTSSP extends the TSSP by incorporating more realistic characteristics: non-reciprocal influence relationships; varying degrees of influence between individuals; different costs for recruiting seminal spreaders; and rewards for influencing individuals.

The `GTSSP` is defined as follows. Consider a portion of a social network represented by a directed graph $D = (V, E)$, where V and E are the sets of vertices and directed edges (arcs) in D. Each vertex represents an individual, and each arc $(u, v) \in E$ indicates that u can influence v during propagation. The *in-neighborhood* of each $v \in V$ is denoted by $N_{\text{in}}(v) = \{u \in V : (u, v) \in E\}$.

A set of vertices chosen to initiate the propagation is called a *target set*, containing the *targets*. As targets transmit information to their out-neighbors, some vertices may become influenced enough to forward it, triggering propagation. Each vertex assumes one of two states: *inactive*, if it has not been influenced yet, or *active*, if it is a target or has received enough influence from its in-neighbors to forward the information.

The *cost* of selecting a vertex v as a target is $c_v \in \mathbb{Z}^+$, and the cost of a target set $S \subseteq V$ is given by $c_S = \sum_{v \in S} c_v$. The *threshold* of v, denoted by $t_v \in \mathbb{Z}^+$, represents the amount of influence v must receive before it can spread the information. Also, the *reward* of v, denoted by $r_v \in \mathbb{Z}^+$, indicates the reward obtained when v is active. The *weight* of an edge $(u, v) \in E$, denoted by $w_{u,v} \in \mathbb{Z}^+$, quantifies the influence exerted by u on v.

Each time interval in a propagation is represented by a *round* $\tau \in \mathbb{N}$, and the subset of V containing the active vertices in round τ is denoted by S_τ. In round $\tau = 0$, all vertices are inactive except for the targets, *i.e.*, $S_0 = S$. For every $\tau \geq 1$, the set of active vertices updates as $S_\tau = S_{\tau-1} \cup \{v \in V \setminus S_{\tau-1} : \sum_{u \in N_{\text{in}}(v) \cap S_{\tau-1}} w_{u,v} \geq t_v\}$. It follows that $S_\tau \subseteq S_{\tau+1}$ for all $\tau \geq 0$.

A non-target v is *activated* in round $\tau \geq 1$ if $v \in S_\tau \setminus S_{\tau-1}$. The propagation terminates in the earliest round ρ in which no vertex is activated. If $\sum_{v \in S_\rho} r_v \geq q$, where $q \in \mathbb{Z}^+$ is a given constant representing the required reward, then S is a *feasible* target set; otherwise, it is *infeasible*.

Problem 1 (`GTSSP`). *Given an instance (D, c, t, w, q), where $D = (V, E)$ is a directed graph, and $c : V \to \mathbb{Z}^+$, $t : V \to \mathbb{Z}^+$, and $w : E \to \mathbb{Z}^+$ are cost, threshold, and weight functions, respectively, and q is the required reward, the objective is to find a feasible target set of minimum cost.*

Since the `TSSP` is a special case of the `GTSSP`, where $(u, v) \in E$ iff $(v, u) \in E$, $w_{u,v} = 1$ for all $(u, v) \in E$, $c_v = r_v = 1$ for all $v \in V$, and $q = |V|$, it follows that `GTSSP` is NP-hard as well. The `GTSSP` was originally introduced in a study by Ravelo and Meneses (2021) [Ravelo and Meneses 2021], where they proposed an integer programming (`IP`) formulation, along with a Lagrangian relaxation approach and a subgradient-based heuristic.

Our contributions. In this work, we propose:

- a new `IP` model for the `GTSSP`;
- a promising Lagrangian relaxation method for this new formulation;
- a conjecture on the strength of the proposed relaxation versus the linear relaxation.

2. A New Lagrangian Relaxation for the `GTSSP`

First, we introduce a novel `IP` formulation for the `GTSSP`, denoted by `IP-GTSSP`. Let $\{x_v : v \in V\}$, $\{y_{u,v} : (u, v) \in E\}$ and $\{a_v : v \in V\}$ be sets of binary variables, where $x_v = 1$ iff v is a target, $y_{u,v} = 1$ iff u transmits influence to v during propagation, and

$a_v = 1$ iff v is active at the end of the propagation. Let $\mathcal{C}(D)$ be the set of all cycles in D, where each $C \in \mathcal{C}(D)$ is a set containing directed edges that form a cycle in D.

The objective function (1) minimizes the cost of the target set. Constraint (2) guarantees that the required reward is attained. Constraints (3) ensure that if a vertex v is a target, then v is active at the end of the propagation. Constraints (4) prevent the targets from being influenced by any of their in-neighbors. Constraints (5) force every vertex v that is activated during the propagation to receive an amount of influence of at least t_v. Lastly, constraints (6) forbid the occurrence of a circular sequence of influences during the propagation by avoiding directed cycles in the directed subgraph of D induced by each $(u, v) \in E$ with $y_{u,v} = 1$.

$$\texttt{IP-GTSSP}\begin{cases} \min \sum_{v\in V} c_v x_v & & (1)\\ \text{s.t.} \quad \sum_{v\in V} r_v a_v \geq q & & (2)\\ x_v \leq a_v & \forall v \in V & (3)\\ x_v + y_{u,v} \leq 1 & \forall (u,v) \in E & (4)\\ \sum_{u\in N_{\text{in}}(v)} w_{u,v} y_{u,v} \geq t_v(a_v - x_v) & \forall v \in V & (5)\\ \sum_{(u,v)\in C} y_{u,v} \leq |C| - 1 & \forall C \in \mathcal{C}(D) & (6) \end{cases}$$

We consider constraints (6) to be the ones that make it difficult to find an optimal solution in `IP-GTSSP`. Let $\{\lambda_C : C \in \mathcal{C}(D)\}$ be a set of non-negative Lagrange multipliers. By dualizing (6), we obtain the Lagrangian relaxation denoted by `LR-GTSSP`.

$$\texttt{LR-GTSSP}\begin{cases} \min \sum_{v\in V} c_v x_v + \sum_{C\in\mathcal{C}(D)} \lambda_C \left(\sum_{(u,v)\in C} y_{u,v} - (|C| - 1)\right) & (7)\\ \text{s.t. constraints (2), (3), (4) and (5).} \end{cases}$$

Define $K = \sum_{C\in\mathcal{C}(D)} \lambda_C(|C| - 1)$ and, additionally, for each $(u, v) \in E$, define $\kappa_{u,v} = \sum_{C\in\mathcal{C}(D):(u,v)\in C} \lambda_C$. Note that these are non-negative constants. Then, the objective function (7) can be rewritten as $\min \sum_{v\in V} c_v x_v + \sum_{(u,v)\in E} \kappa_{u,v} y_{u,v} - K$. Moreover, since K is constant, it can be omitted from the objective function. We decompose `LR-GTSSP` into $|V|$ independent subproblems, one for each $v \in V$, denoted by `SUB-MODEL`. Next, we show how to solve `SUB-MODEL` for a fixed vertex v.

$$\texttt{SUB-MODEL}\begin{cases} \min c_v x_v + \sum_{u\in N_{\text{in}}(v)} \kappa_{u,v} y_{u,v} & & (8)\\ \text{s.t.} \quad x_v \leq a_v & & (9)\\ x_v + y_{u,v} \leq 1 & \forall u \in N_{\text{in}}(v) & (10)\\ \sum_{u\in N_{\text{in}}(v)} w_{u,v} y_{u,v} \geq t_v(a_v - x_v) & & (11) \end{cases}$$

If $a_v = 0$, then an optimal solution for `SUB-MODEL` satisfies $x_v = 0$ and $y_{u,v} = 0$ for each $(u, v) \in N_{\text{in}}(v)$. In this case, the value of the objective function (8) is zero. On the other hand, if $a_v = 1$, then we can have either $x_v = 1$ or $x_v = 0$. If $x_v = 1$, an optimal solution for `SUB-MODEL` satisfies $y_{u,v} = 0$ for each $(u, v) \in N_{\text{in}}(v)$. In this case, the value of the objective function (8) is c_v. It remains to analyze the case where $a_v = 1$ and $x_v = 0$. In this scenario, an optimal solution for `SUB-MODEL` satisfies $\sum_{u\in N_{\text{in}}(v)} w_{u,v} y_{u,v} \geq t_v$ while $\sum_{u\in N_{\text{in}}(v)} \kappa_{u,v} y_{u,v}$ is minimum.

To meet these conditions and determine optimal values for the y variables, we solve an instance of a variant of the Knapsack Problem, named the Minimization Knapsack Problem (MinKP) by Kellerer et al. [Kellerer et al. 2004]. In the MinKP, we are given a collection of items, each with a *profit* and a *burden*, and the goal is to select a subset of these items such that the total profit is at least a given required value while minimizing the total burden. Here, t_v represents the required profit, and each arc (u, v) with $u \in N_{\text{in}}(v)$ corresponds to an item, where $w_{u,v}$ is its profit and $\kappa_{u,v}$ is its burden.

The MinKP is weakly NP-hard, and an exact dynamic programming algorithm exists for this problem, running in $\mathcal{O}(\alpha \cdot \beta)$ time, where α is the number of items and β is the sum of the profits of the items minus the required profit [Kellerer et al. 2004]. This algorithm can be applied to solve SUB-MODEL in the case where $a_v = 1$ and $x_v = 0$ in $\mathcal{O}\left(|N_{\text{in}}(v)| \cdot \left(\sum_{u \in N_{\text{in}}(v)} w_{u,v} - t_v\right)\right)$ time.

Thus, when $a_v = 1$, there are two possible solutions: one for $x_v = 1$ and another for $x_v = 0$. We then evaluate these solutions based on their objective values according to objective function (8) and select the one that provides the optimal solution for SUB-MODEL in the case where $a_v = 1$.

Now, we return to the LR-GTSSP model, where we need to determine which vertices must be active at the end of the propagation. Thus, we must decide whether $a_v = 0$ or $a_v = 1$ for each vertex v while respecting constraint (2) and minimizing the objective function (7). To tackle this problem, we solve one final instance of the MinKP, where q represents the required profit, and each vertex $v \in V$ corresponds to an item, with r_v as its profit and the optimal value for SUB-MODEL (when $a_v = 1$) as its burden. This instance of the MinKP can be solved in $\mathcal{O}\left(|V| \cdot \left(\sum_{v \in V} r_v - q\right)\right)$ time.

Next, we provide insights into the strength of the proposed Lagrangian relaxation compared to the linear relaxation. If Conjecture 1 holds, then there exist λ multipliers such that the LR-GTSSP yields a stronger lower bound for the IP-GTSSP model than the linear relaxation [Fisher 1981], leading to Conjecture 2.

Conjecture 1. *No integer optimal solutions exist for LR-GTSSP when the integrality constraints are relaxed.*

Conjecture 2. *There exist non-negative multipliers λ_C for each $C \in \mathcal{C}(D)$ such that the optimal solution value for LR-GTSSP is strictly greater than the optimal solution value for the linear relaxation of IP-GTSSP.*

3. Conclusion and Future Work

In this work, we introduced an IP model for the GTSSP, along with a Lagrangian relaxation. We showed how to solve this relaxation exactly by decomposing the formulation into a series of sub-models, each corresponding to an instance of the Minimization Knapsack Problem. We conjecture that there exist Lagrange multipliers for which the relaxation provides stronger dual bounds for the IP formulation than its linear relaxation.

As the next steps, we aim to prove Conjectures 1 and 2. Additionally, we plan to develop an algorithm to compute Lagrange multipliers for the LR-GTSSP model that yield the tightest bounds. A promising approach is the subgradient method [Zhao et al. 1999], which iteratively updates the multipliers based on constraint violations. By fine-tuning the step size and exploring alternative strategies, we aim to accelerate convergence and improve bound quality.

Finally, we plan to evaluate our approach through computational experiments on a diverse set of instances. In particular, we aim to compare the dual bounds for the `GTSSP` with those reported by Ravelo and Meneses (2021) [Ravelo and Meneses 2021], obtained via their relaxation of a different `IP` model for the problem. The experimental results will offer insights into the practical effectiveness of our method and guide further refinements.

References

Ackerman, E., Ben-Zwi, O., and Wolfovitz, G. (2010). Combinatorial Model and Bounds for Target Set Selection. *Theoretical Computer Science*, 411(44):4017–4022.

Chen, N. (2009). On the Approximability of Influence in Social Networks. *SIAM Journal on Discrete Mathematics*, 23(3):1400–1415.

Chen, W., Lakshmanan, L., and Castillo, C. (2013). Information and Influence Propagation in Social Networks. *Synthesis Lectures on Data Management*, 5(4):1–177.

Fisher, M. L. (1981). The lagrangian relaxation method for solving integer programming problems. *Management Science*, 27(1):1–18.

Kellerer, H., Pferschy, U., and Pisinger, D. (2004). *Knapsack Problems*. Springer.

Li, Y., Fan, J., Wang, Y., and Tan, K. (2018). Influence Maximization on Social Graphs: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872.

Pereira, F. C. and de Rezende, P. J. (2023). The Least Cost Directed Perfect Awareness Problem: Complexity, Algorithms and Computations. *Online Social Networks and Media*, 37-38:100255.

Pereira, F. C., de Rezende, P. J., and de Souza, C. C. (2021). Effective Heuristics for the Perfect Awareness Problem. In *Proocedings of the XI Latin and American Algorithms, Graphs and Optimization Symposium*, pages 489–498.

Pereira, F. C., de Rezende, P. J., and Yunes, T. (2024). Minimizing the Cost of Leveraging Influencers in Social Networks: IP and CP Approaches. In *Proceedings of the 21st International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, pages 111–127.

Raghavan, S. and Zhang, R. (2019). A Branch-and-Cut Approach for the Weighted Target Set Selection Problem on Social Networks. *INFORMS Journal on Optimization*, 1(4):304–322.

Ravelo, S. V. and Meneses, C. N. (2021). Generalizations, Formulations and Subgradient Based Heuristic with Dynamic Programming Procedure for Target Set Selection Problems. *Computers & Operations Research*, 135:105441.

Shakarian, P., Eyre, S., and Paulo, D. (2013). A Scalable Heuristic for Viral Marketing Under the Tipping Model. *Social Network Analysis and Mining*, 3(4):1225–1248.

Spencer, G. and Howarth, R. (2013). Maximizing the spread of stable influence: Leveraging norm-driven moral-motivation for green behavior change in networks. *CoRR*, abs/1309.6455.

Zhao, X., Luh, P. B., and Wang, J. (1999). Surrogate gradient algorithm for lagrangian relaxation. *Journal of Optimization Theory and Applications*, 100(3):699–712.