TU WIEN Informatics

Exact methods for the Selective Assessment Routing Problem

Joan Salvà Soler, 12223411

September 11, 2023

Supervision:
Vera C. Hemmelmayr, PhD, Wirtschaftsuniversität Wien, main supervisor
Günther Raidl, Ao.Univ.Prof, TU Wien, co-supervisor

Project Title	Exact Methods for the Selective Assessment Routing Problem
Area	Operations Research
Student	Joan Salvà Soler, 12223411
Semester	WS 2023/2024
Domain lecture	Humanitarian Logistics
Research group	Institute for Transport and Logistics - WU Wien
Lecture done or in parallel	In parallel

***This is a pre-print. The project is expected to be finished by December 2023.**

Abstract

In the immediate aftermath of a disaster, relief agencies perform rapid needs assessments to investigate the effects of the disaster on the affected communities. Since assessments must be performed quickly, visiting all of the sites in the affected region may not be possible. The Selective Assessment Routing Problem (SARP) was first defined by [1] and addresses the site selection and routing decisions of the rapid needs assessment teams which aim to evaluate the post-disaster conditions

of different community groups, each carrying a distinct characteristic. SARP constructs an assessment plan that maximizes the covering of different characteristics in a balanced way. While some approximate solutions have been already proposed ([1], [2]), this project explores exact approaches that maximally exploit or extend the capabilities of numerical solvers. Different mathematical formulations will be discussed, and different exact optimization techniques will be considered to improve the performance of numerical solvers.

1 Motivation

Once a disaster occurs, humanitarian agencies first carry out rapid assessments of the affected region. The objective of the rapid assessment stage is to perform a broad evaluation of the disaster impact and population needs [3]. The rapid assessment process may start as early as a few hours after the disaster occurrence and is typically completed within three days [3].

Since the rapid needs assessment stage must be completed quickly, assessment teams do not make observations at all of the affected sites; therefore, sites to be visited are sampled. The general aim of sampling in the rapid needs assessment stage is to choose a limited number of sites that will allow the assessment teams to observe and compare the post-disaster conditions of different community groups. Random sampling is always an option, but purposive sampling constructs an assessment plan as follows [4]: (1) specify the critical characteristics that define the target community groups, (2) select sites to be visited according to sample richness, and (3) plan the logistics (assign teams and routes) according to routing and resources constraints. However, the before-mentioned process might become a trial and error procedure that oscillates between (2) and (3). The SARP addresses this problem by solving an optimization problem, the solution of which is the assessment plan.

To the best of our knowledge, exact methods for SARP remain largely unexplored in the existing literature. Though researchers in [1] propose a MILP formulation and solve some instances with an optimization solver, we do not have any record of another attempt to solve the SARP exactly.

Execution time is a key element to take into consideration in the context of rapid needs assessment after a humanitarian disaster, and it could be contended that it should be prioritized before the quality of the solution. However, we argue that it is still advantageous to pursue optimality despite the additional time investment. This approach yields lower bounds that can serve as benchmarks for subsequent research, enabling the evaluation of solutions provided by approximate methods. Moreover, working on exact methods for the SARP also offers potential inspiration for the development of exact methods in related problems like the Team Orienteering Problem (TOP) [5] or the so-called *routing problems with profits* [6].

2 Methodology

2.1 Mathematical formulations and algorithm design

Alternative MILP formulations to the one in [1] will be investigated. Due to the closeness of the problems, we will explore how the well-studied formulations for the VRP translate into different formulations for the SARP: vehicle flow formulations (MTZ constraints), commodity flow formulations (SCF, MCF), or exponentially sized formulations that use, for instance, Subtour Elimination Constraints [7].

For exponential formulations, we will study different classes of strengthening inequalities and separation procedures. Different relaxations for the problem will be studied. In particular, we will pay attention to combinatorial and Lagrangian relaxations. Additionally, we will discuss possible primal heuristics to construct feasible solutions during the optimization process and for starting. Finally, if considered adequate and promising, we explore advanced techniques like branching schemes and search strategies that exploit the properties of SARP.

2.2 Implementation

The computational study will be done using Gurobi with its Python API (using a student license). Depending on the feasibility and how promising the different intended techniques are, further options are to use Gurobi or SCIP in conjunction with C++.

2.3 Solution to problem instances

We will work with the exact same instances considered in [1]. We hope to obtain them directly from the author of the paper. We will test our developed approaches on all instances agreeing on a time limit, will validate the solutions, and will analyze and visualize the results (solving times, optimality gap, nodes explored...).

We will finally apply our best developed algorithm to a case study that was proposed by [1]. It consists of assessing the rapid needs assessment to affected towns after the earthquake in Van, which is a city in eastern Turkey. The earthquake occurred on 23 October 2011 and had a magnitude of 7.2; it killed more than 600 people and left more than 200,000 people homeless and in need. We will discuss how the developed solution aligns with the goals of Humanitarian Logistics, and how the different key concepts of this discipline are utilized in our work.

3 Mathematical Formulations

The following notation is common for all formulations. N is the set of affected sites. K is the set of vehicles (or teams), and each team $k \in K$ departs from the origin node $\{0\}$ and returns back to the origin after completing the route. the travel time between nodes is represented by t_{ij}, $\forall i, j \in N_0$. Each vehicle is allowed to travel at most $T_{\max}$ time units. Let C denote the set of critical characteristics of interest. A coverage parameter α_{ik} is defined, which takes value 1 if node $i \in N$ carries characteristic $c \in C$, and 0 otherwise. The total number of sites that carry characteristic $c \in C$ is represented by τ_c.

3.1 MTZ

This is the original formulation and uses Miller Tucker Zemlin constraints for cycle elimination. We will also refer to it as MTZ-3index or MTZ-3. The formulation uses the following variables:

- x_{ijk} takes value 1 if vehicle $k \in K$ visits site $j \in N_0$ right after $i \in N_0$, and 0 otherwise.
- y_{ik} takes value 1 if vehicle $k \in K$ visits site $i \in N_0$, and 0 otherwise.
- z is the continuous variable a maximize that models the coverage rate. It is bounded by 0 and 1.
- u_i is a continuous variable that defines a sequence on the visited nodes $i \in N$ that is valid for all routes.

$$\max z \tag{1}$$

$$\text{s.t. } z \cdot \tau_c \leq \sum_{i \in N} \sum_{k \in K} \alpha_{ic} \cdot y_{ik} \qquad \forall c \in C \tag{2}$$

$$\sum_{j \in N_0} x_{ijk} = y_{ik} \qquad \forall i \in N_0, k \in K \tag{3}$$

$$\sum_{j \in N_0} x_{jik} = y_{ik} \qquad \forall i \in N_0, k \in K \tag{4}$$

$$\sum_{k \in K} y_{ik} \leq 1 \qquad \forall i \in N \tag{5}$$

$$\sum_{i \in N_0} \sum_{j \in N_0} t_{ij} x_{ijk} \leq T_{\max} \qquad \forall k \in K \tag{6}$$

$$u_i - u_j + |N| \cdot x_{ijk} \leq |N| - 1, \qquad \forall k \in K, i \neq j \in N \tag{7}$$

$$x_{iik} = 0 \qquad \forall i \in N, k \in K \tag{8}$$

$$x_{ijk} \in \{0, 1\} \qquad \forall i, j \in N_0, k \in K \tag{9}$$

$$y_{ik} \in \{0, 1\} \qquad \forall i \in N, k \in K \tag{10}$$

$$0 \leq z \leq 1 \tag{11}$$

$$0 \leq u_i \leq |N| \qquad \forall i \in N \tag{12}$$

3.2 SCF

This is a Single Commodity Flow formulation. It uses the following variables:

- f_{ij} is the flow circulating from $i \in N_0$ to $j \in N_0$.
- x_{ij} takes value 1 if some vehicle visits site $j \in N_0$ right after $i \in N_0$, and 0 otherwise.
- y_i takes value 1 if some vehicle visits site $i \in N$, and 0 otherwise.
- z is the continuous variable a maximize that models the coverage rate. It is bounded by 0 and 1.

$$\max\ z \tag{13}$$

$$\text{s.t.}\quad z \cdot \tau_c \leq \sum_{i \in N} \alpha_{ic} \cdot y_i \qquad \forall c \in C \tag{14}$$

$$f_{ij} \leq T_{\max} \cdot x_{ij} \qquad \forall i \in N_0, j \in N_0 \tag{15}$$

$$\sum_{i \in N_0} x_{0i} = |K| \tag{16}$$

$$\sum_{i \in N_0} x_{i0} = |K| \tag{17}$$

$$f_{i0} = x_{i0} \cdot t_{i0} \qquad \forall i \in N \tag{18}$$

$$\sum_{j \in N_0} (f_{ji} - f_{ij}) = \sum_{j \in N_0} x_{ji} \cdot t_{ji} \qquad \forall i \in N \tag{19}$$

$$\sum_{j \in N_0} (x_{ji} + x_{ij}) = 2y_i \qquad \forall i \in N \tag{20}$$

$$\sum_{j \in N_0} x_{ji} = y_i \qquad \forall i \in N \tag{21}$$

$$\sum_{j \in N_0} x_{ij} = y_i \qquad \forall i \in N \tag{22}$$

$$x_{ii} = 0 \qquad \forall i \in N_0 \tag{23}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i, j \in N_0 \tag{24}$$

$$y_i \in \{0, 1\} \qquad \forall i \in N \tag{25}$$

$$0 \leq z \leq 1 \tag{26}$$

$$0 \leq f_{ij} \leq |T_{\max}| \qquad \forall i, j \in N \tag{27}$$

3.3 Cutset

This formulation is essentially the MTZ formulation with the following changes:

- u_i variables are not defined
- The MTZ set of constraints (7) is replaced by the following set of constraints:

$$\sum_{(i,j)\in\delta^+(S)} x_{ijk} \geq y_{hk}, \qquad \forall S \subset N, h \in S, k \in K,$$

  where $\delta^+(S) = \{(i,j) : i \in S, j \notin S\}$.

The above-defined set of constraints is naturally exponential in size. In order to progressively add constraints to the initial model, we propose the following separation procedure to be applied at each node of the Branch & Bound tree that finds an integer solution.

Separation procedure. For each vehicle $k \in K$, we consider a copy of the original network that has edges ij with capacity x_{ijk}. For each node $i \in N$ in this graph, we calculate the Max Flow f between the depot 0 and i. If $f < y_{ik}$, we find a Min-Cut partition of the graph, we take S as the side that contains i, and we finally add the CutSet constraint for S.

3.4 MTZ-2index

This formulation simplifies the MTZ-3 formulation by making it vehicle-agnostic. We also refer to it as MTZ-2 or MTZOpt. In essence, we drop the vehicle subindex of the MTZ-3 formulation. This is the original formulation and uses MTZ constraints for cycle elimination, even though the variables u_i now model the time that some vehicle visits site $i \in N$. The formulation uses the following variables:

- x_{ij} takes value 1 if some vehicle visits site $j \in N_0$ right after $i \in N_0$, and 0 otherwise.
- y_i takes value 1 if some vehicle visits site $i \in N$, and 0 otherwise.
- z is the continuous variable a maximize that models the coverage rate. It is bounded by 0 and 1.
- u_i is a continuous variable that models the (last) time at which a vehicle visits site $i \in N_0$. On the origin, u_0 models the last time that a vehicle returns to the depot.

$$
\begin{aligned}
&\max\ z && && (28)\\
&\text{s.t. } z \cdot \tau_c \leq \sum_{i \in N} \alpha_{ic} \cdot y_i && \forall c \in C && (29)\\
&\sum_{j \in N_0} x_{ij} = y_i && \forall i \in N && (30)\\
&\sum_{j \in N_0} x_{ji} = y_i && \forall i \in N && (31)\\
&\sum_{i \in N} x_{0i} = |K| && && (32)\\
&\sum_{i \in N} x_{i0} = |K| && && (33)\\
&u_j - u_i \geq x_{ij}(T_{\max} + t_{ij}) - T_{\max}, && \forall i \in N, j \in N_0, i \neq j && (34)\\
&x_{ii} = 0 && \forall i \in N && (35)\\
&x_{ij} \in \{0,1\} && \forall i, j \in N_0 && (36)\\
&y_i \in \{0,1\} && \forall i \in N && (37)\\
&0 \leq z \leq 1 && && (38)\\
&x_{0i} \cdot t_{0i} \leq u_i \leq T_{\max} && \forall i \in N_0 && (39)
\end{aligned}
$$

4 Theoretical study

Proposition 4.1: SCF is stronger than MTZ-2, i.e., $P_{SCF} \subset P_{MTZ-2}$ and they are not equal.
Proof: We project the two formulations to the common space $(\dots y_i \dots, \dots x_{ij} \dots)$. We start with an SCF solution and define

$$u_0 = T_{\max} \tag{40}$$
$$u_i = T_{\max} + \sum_{k \in N_0} x_{ki} t_{ki} - \sum_{k \in N_0} f_{ki}, \qquad i \in N, \tag{41}$$

which gives a solution for MTZ-2. The goal is to see that it is feasible for MTZ-2, i.e., that it satisfies (34) and (39). The remaining MTZ-2 constraints are already satisfied because x_{ij}, y_i define a feasible VRP solution.

$$u_j - u_i \geq x_{ij} t_{ij} - T_{\max}(1 - x_{ij}), \qquad \forall i \in N, j \in N_0$$
$$x_{0i} t_{0i} \leq u_i \leq T_{\max}, \qquad \forall i \in N_0.$$

Constraint (39) is easy to check:

$$u_i = T_{\max} + \sum_{k \in N_0} x_{ki} t_{ki} - \sum_{k \in N_0} f_{ki} \underset{(15)}{\geq} T_{\max} + \underbrace{\sum_{k \in N_0} x_{ki} t_{ki}}_{\geq x_{0i} t_{0i}} - T_{\max} \geq x_{0i} t_{0i}$$

$$u_i = T_{\max} + \sum_{k \in N_0} x_{ki} t_{ki} - \sum_{k \in N_0} f_{ki} \underset{(19)}{=} T_{\max} - \sum_{k \in N_0} f_{ik} \leq T_{\max}$$

For proving (34), we start with the case $j = 0$.

$$u_0 - u_i = -\sum_{k \in N_0} x_{ki} t_{ki} + \sum_{k \in N_0} f_{ki} \underset{(19)}{=} \sum_{k \in N_0} f_{ik} = f_{i0} + \sum_{k \in N} f_{ik} \underset{(18)}{=} x_{i0} t_{i0} + \sum_{k \in N} f_{ik}$$

Since $\sum_{k \in N} f_{ik} \geq 0$ and $-T_{\max}(1 - x_{ij}) \leq 0$, (34) holds for $j = 0$.
We now assume $j \neq 0$.

$$u_j - u_i = \underbrace{\sum_{k \in N_0} x_{kj} t_{kj}}_{\geq x_{ij} t_{ij}} - \sum_{k \in N_0} x_{ki} t_{ki} + \sum_{k \in N_0} f_{ki} - \sum_{k \in N_0} f_{kj} \underset{(19)}{\geq}$$

$$\geq x_{ij}t_{ij} + \underbrace{\sum_{k\in N_0} f_{ik}}_{\geq f_{ij}} - \sum_{k\in N_0} f_{kj} \geq x_{ij}t_{ij} - \sum_{k\in N_0:k\neq i} f_{kj} \underset{(15)}{\geq}$$

$$\geq x_{ij}t_{ij} - T_{\max}\sum_{k\in N_0:k\neq i} x_{kj} \geq x_{ij}t_{ij} - T_{\max}(1-x_{ij}).$$

In the last inequality, we used (21):

$$\sum_{k\in N_0} x_{kj} = y_j \leq 1 \iff \sum_{k\in N_0:k\neq i} x_{kj} \leq 1 - x_{ij}.$$

This proves that $P_{SCF} \subset P_{MTZ-2}$. To see that they are not equal, consider the following example:

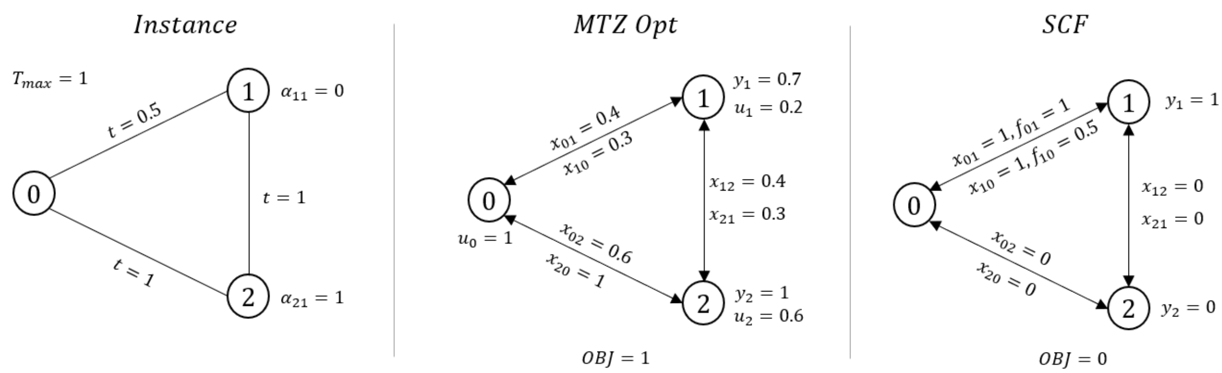

Figure 1: On the left, the instance presented, which only considers one characteristic: $C = \{0\}$. In the middle, the MTZ Opt and SCF solutions to the respective Linear Relaxations

.

The optimal LR of the MTZ-2 formulation gives a worse bound than the LR of the SCF formulation, which proves $P_{SCF} \neq P_{MTZ-2}$. $\square$

Proposition 4.2: SCF is stronger than MTZ-3, i.e., $P_{SCF} \subset P_{MTZ-3}$ and they are not equal.
Proof: We start with an SCF solution that we note as $(...x_{ij}..., ...y_i..., ...f_{ij}...)$, and we will build an MTZ-3 noted as $(...x_{ijk}..., ...y_{ik}..., ...u_i...)$. The x and y variables have different subindices in each formulation, so there should be no confusion regarding the formulation they belong to.

We define the MTZ-3 solution as follows:

$$x_{ijk} = \frac{x_{ij}}{|K|}, \qquad i, j \in N_0, k \in K \tag{42}$$

$$y_{ik} = \sum_{j \in N_0} x_{ijk}, \qquad i \in N_0, k \in K \tag{43}$$

$$\tilde{u}_i = T_{\max} + \sum_{k \in N_0} x_{ki} t_{ki} - \sum_{k \in N_0} f_{ki}, \qquad i \in N, \tag{44}$$

$$u_1, \dots u_n = F(\tilde{u_1}, \dots \tilde{u_n}) \tag{45}$$

Note that (43) is defined based on (42). Definition (45) uses F, a map that enumerates its arguments based on its relative order, starting from 1. For example, $F(0.3, 0.1, 0.6) = (2, 1, 3)$.

We need to prove that the definition above corresponds to an MTZ-3 solution, i.e, we have to check constraints (2)-(12)
Before we start, a useful result is the following:

$$\sum_{k \in K} y_{ik} = \sum_{k \in K} \sum_{j \in N_0} x_{ijk} = \sum_{k \in K} \sum_{j \in N_0} \frac{x_{ij}}{|K|} \underset{(22)}{=} y_i \tag{46}$$

We start with (2):

$$z \cdot \tau_c \leq \sum_{i \in N} \alpha_{ic} \cdot y_i \underset{(22)}{=} \sum_{i \in N} \alpha_{ic} \cdot y_i \underset{(46)}{=} \sum_{i \in N} \sum_{k \in K} \alpha_{ic} \cdot y_{ik}.$$

(3) is true by definition. We again use the proved relationship (46) to prove (4):

$$\sum_{k \in K} \sum_{j \in N_0} x_{jik} = \sum_{k \in K} \sum_{j \in N_0} \frac{x_{ji}}{|K|} \underset{(21)}{=} y_i = \sum_{k \in K} y_{ik}.$$

(5) follows naturally from (46):

$$\sum_{k \in K} y_{ik} = y_i \leq 1$$

(6) is less trivial and uses the fact that $\sum_{i \in N_0} (f_{ji} - f_{ij}) \leq 0$, which states that nodes consume flow. This can be easily derived from (19).

$$\sum_{i\in N_0}\sum_{j\in N_0} x_{ijk}t_{ij} = \frac{1}{|K|}\sum_{i\in N_0}\sum_{j\in N_0} x_{ij}t_{ij} = \frac{1}{|K|}\sum_{i\in N_0}\sum_{j\in N_0} x_{ji}t_{ji} \underset{(19)}{=} \frac{1}{|K|}\left(\sum_{i\in N_0} x_{0i}t_{0i} + \sum_{i\in N_0}\sum_{j\in N}(f_{ji}-f_{ij})\right) \leq$$

$$\underset{(18)}{\leq} \frac{1}{|K|}\left(T_{\max}\underbrace{\sum_{i\in N_0} x_{0i}}_{|K|} + \sum_{j\in N}\underbrace{\sum_{i\in N_0}(f_{ji}-f_{ij})}_{\leq 0}\right) \leq \frac{1}{|K|}\cdot T_{\max}|K| = T_{\max}$$

We finally need to prove the MTZ constraint (7). The proof of proposition 4 shows that given an SCF solution, the definition of visiting times like (44) satisfies the MTZ-2 constraint. Therefore, our $\tilde{u}_i$ values satisfy:

$$\tilde{u}_i - \tilde{u}_j + x_{ij}t_{ij} \leq T_{\max}(1 - x_{ij}), \forall i, j \in N$$

The choice of the function F implies that the mapping of the times $\tilde{u}_i$ to integer ordering u_i satisfies

$$u_i - u_j + 1 \leq |N|(1 - x_{ijk}).$$

This proves that $P_{SCF} \subset P_{MTZ-3}$. To see that they are not equal, the same example as in 4 works:

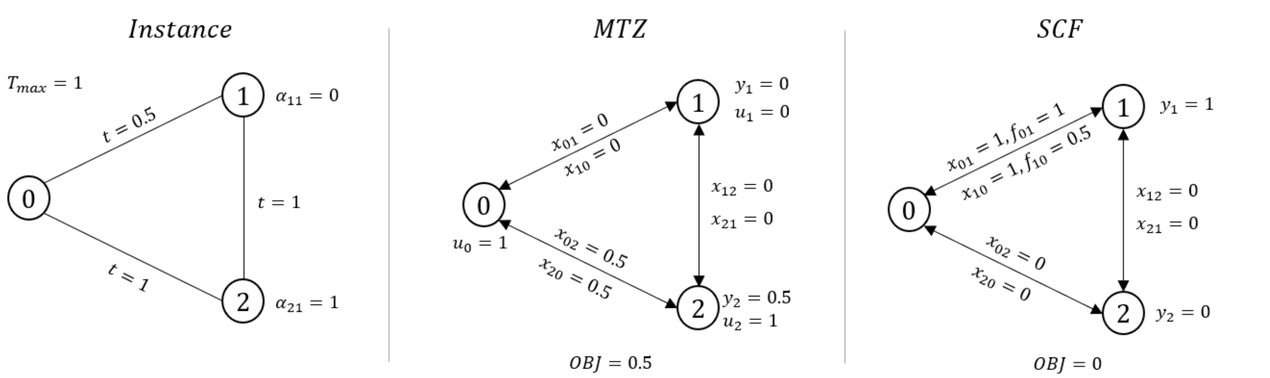

Figure 2: On the left, the instance presented. In the middle, the MTZ-3 and SCF solutions to the respective Linear Relaxations

.

The optimal LR of the MTZ-3 formulation gives a worse bound than the LR of the SCF formulation, which proves $P_{SCF} \neq P_{MTZ-3}$. $\square$

Proposition 4.3: SCF and CutSet are not comparable, i.e., $P_{SCF} \not\subset P_{CutSet}$ and $P_{Cutset} \not\subset P_{SCF}$.
Proof: Consider the following examples.

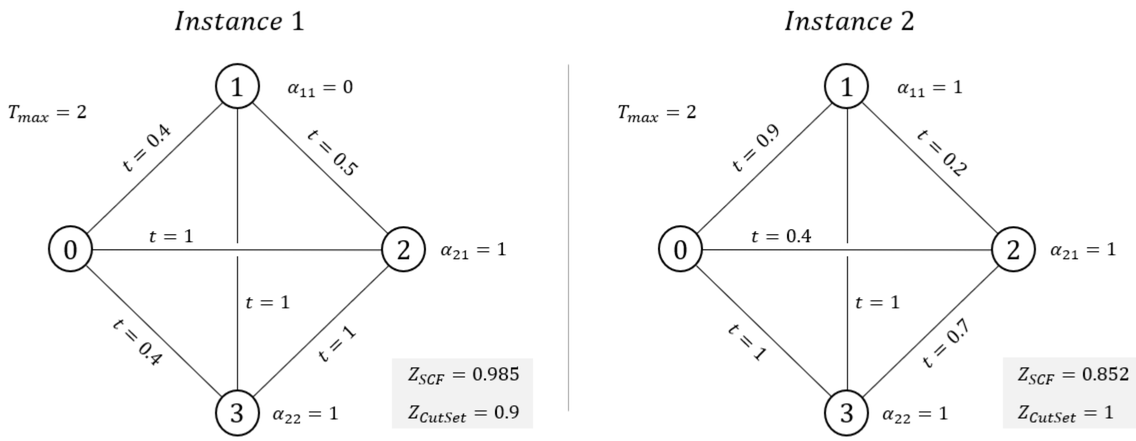

Figure 3: Instance 1 shows that $P_{SCF} \not\subset P_{CutSet}$. Instance 2 shows $P_{Cutset} \not\subset P_{SCF}$.

Proposition 4.4: MTZ-3 and CutSet are not comparable, i.e., $P_{MTZ-3} \not\subset P_{CutSet}$ and $P_{Cutset} \not\subset P_{MTZ-3}$.
Proof: Consider the following examples.

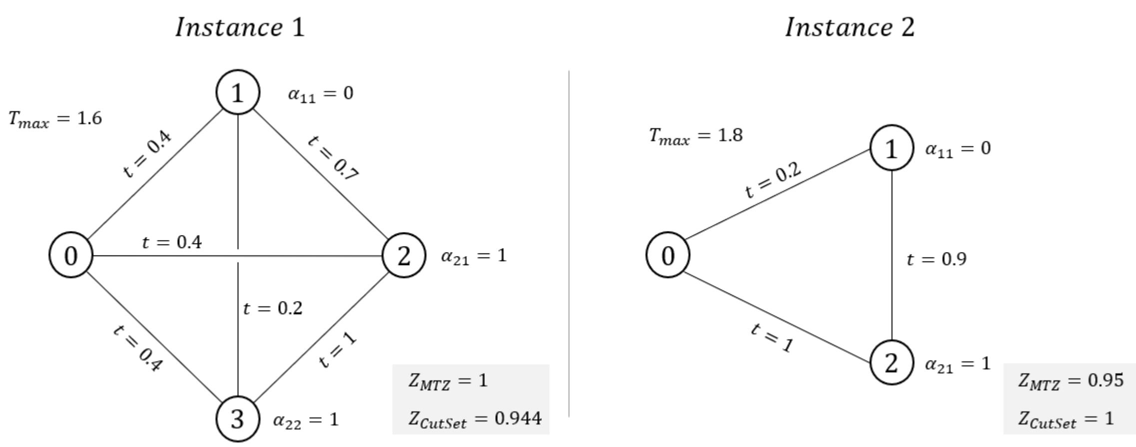

Figure 4: Instance 1 shows that $P_{MTZ-3} \not\subset P_{CutSet}$. Instance 2 shows $P_{Cutset} \not\subset P_{MTZ-3}$.

References

[1] B. Balcik, "Site selection and vehicle routing for post-disaster rapid needs assessment," *Transportation Research Part E: Logistics and Transportation Review*, vol. 101, pp. 30–58, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554516303106

[2] M. Hakimifar, V. C. Hemmelmayr, and F. Tricoire, "A lexicographic maximin approach to the selective assessment routing problem," *OR Spectr.*, vol. 45, no. 1, p. 205–249, jul 2022. [Online]. Available: https://doi.org/10.1007/s00291-022-00687-8

[3] A. Maya, "Rapid assessment in disasters," *JMAJ*, vol. 56, pp. 19–24, 2013.

[4] ACAPS, "Technical brief: Purposive sampling and site selection in phase 2," 2011. [Online]. Available: https://www.acaps.org/sites/acaps/files/resources/files/purposive_sampling_and_site_selection_in_phase_2.pdf

[5] I.-M. Chao, B. L. Golden, and E. A. Wasil, "The team orienteering problem," *European journal of operational research*, vol. 88, no. 3, pp. 464–474, 1996.

[6] D. Feillet, P. Dejax, and M. Gendreau, "Traveling salesman problems with profits," *Transportation science*, vol. 39, no. 2, pp. 188–205, 2005.

[7] P. T. D. Vigo, *The Vehicle Routing Problem*, ser. Monographs on Discrete Mathematics and Applications. SIAM, 2001.