



University of Glasgow | School of
Computing Science

The Traveller's Problem

Iva Stefanova Babukova

School of Computing Science
Sir Alwyn Williams Building
University of Glasgow
G12 8QQ

Masters project proposal

18 December 2016

Contents

1	Introduction	3
2	Problem Formulation	3
2.1	Hard Constraints	4
2.2	Soft Constraints	4
3	Worked Examples	5
4	List of Problems	7
5	Complexity of TP	8
6	Background Survey	10
6.1	Linear and Integer Programming	10
6.1.1	Integer Programming Formulation of TSP	11
6.1.2	The Assignment Problem Relaxation	11
6.2	Branch and Bound Algorithms	13
6.3	Constraint Programming	13
6.3.1	Local Search	13
6.3.2	Heuristics	13
7	Proposed Approach	13
7.1	CP Model	14
7.2	IP Model	14
7.3	Experimental Datasets	14
7.3.1	Skyscanner Flights API	14
7.3.2	Randomly Generated TP instances	14

Education Use Consent

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic format. **Please note that you are under no obligation to sign this declaration, but doing so would help future students.**

Name: _____ Signature: _____

1 Introduction

In this work we present the Traveller's Problem (TP), a computational task whose extensions and variations are often encountered by travellers around the world. The task is concerned with creating a valid travel schedule, using airplanes as a means of transportation and in accordance with certain constraints specified by the traveller.

2 Problem Formulation

Each instance of TP consists of:

1. A set of airports $A = \{A_0, \dots, A_n\}$ for $n > 0$. Each airport $A_i \in A$ represents a location the traveller can begin their commute in, visit as a desired destination, or connect in on the way to their destination.
2. The trip starts and ends at the same airport A_0 , which is referred to as the *home point*.
3. The total travel time T , within which the traveller must have visited all destinations and returned to the home point. **The first day is day 0.**
4. A set of flights $F = \{f_0, \dots, f_m\}$. Each flight f_j has:
 - departure airport A_j^d ,
 - arrival airport A_j^a ,
 - date t_j ,
 - duration Δ_j ,
 - cost c_j ,

for some non-negative integer j less than or equal to n . The date t_j is a positive rational number less than or equal to T that shows at which day f_j leaves its departure airport. The duration Δ_j is a positive fraction that shows the amount of time that takes for flight f_j to go from A_j^d to A_j^a . The cost c_j is a positive number that denotes the number of units of some currency ϵ that the traveller pays in order to be able to board flight f_j .

5. Each airport A_i has a *connection time* C_{A_i} , that is the time that takes to switch from any selected flight f_p with $A_p^a = A_i$ to any selected flight f_q with $A_q^d = A_i$, **where f_q is immediately after f_p in a solution.**
6. A set of *destinations* $D = \{D_1, \dots, D_l\}$, $D \subseteq A$, $l \leq n$.

A solution to any instance of TP is a sequence s of k valid flights, $\langle f_{i_1}, f_{i_2}, \dots, f_{i_k} \rangle \subseteq F$. We say that s is valid if the flights in s have the following properties:

1. $A_{i_1}^d = A_{i_k}^a = A_0$
2. $A_{i_j}^a = A_{i_{j+1}}^d, \quad 0 < j < k \quad j$
3. $t_{i_j} + \Delta_{i_j} + C_r \leq t_{i_{j+1}}, \quad 0 < j \leq k, \quad \text{where } r = A_{i_{j+1}}^d$
4. $t_{i_k} + \Delta_{i_k} \leq T$
5. $\forall D_p \in D, \exists f_{i_j} \in s, \text{ such that } A_{i_j}^a = D_p$

Note that a valid sequence of flights may contain one or more flights to and from airports that are not destinations. Such airports are called *connections*.

The *optimization* version of TP (TPO) asks for an *optimal* solution s which minimizes the total sum of the prices of the flights in s , denoted by $c(s)$.

The *decision* version of TP (TPD) asks whether there exists a valid sequence of flights s , such that $c(s)$ is less than or equal to some given integer B . The solution of this problem is a ‘yes’ or ‘no’ answer.

There exist a variety of additional constraints and extensions that can be added to TP. Our problem formulation has only presented the hard constraints which every valid solution to a TP instance must satisfy. In real-world problems, travellers may have additional preferences (soft constraints) and requirements (hard constraints) with regards to their travel. These are discussed in the next two sections.

2.1 Hard Constraints

This section presents some additional constraints that might be imposed on a TP instance. If any of them is required, then a solution that does not satisfy the requirement is considered as invalid.

- Travellers may wish to spend a certain amount of days at a given destination, specified by both upper and lower bounds. The days may additionally be constrained to be consecutive or not.
- Travellers may require to spend a given date at a given destination, for example due to an event occurring on that date in this destination.
- Travellers may require not to fly through a given airport more than once.

2.2 Soft Constraints

It may be desirable to search for a solution that satisfies some of the following requirements:

- Travellers may wish to spend a certain amount δ_i of days in each destination D_i , where δ_i may be specified as a lower or an upper bound.
- Travellers may wish to avoid taking connection flights. In such requirement, we wish to maximise the number of flights to and from destinations.
- Travellers may want to spend as little time on flying as possible. In such case, we wish to find a solution that minimises the sum of the durations of all flights.

Note that we may have an instance for which all soft constraints can not be satisfied simultaneously. In such case, the traveller may be required to rank his requirements in an order of preference. The instance becomes a lexicographic optimisation problem, where we first optimise the highest ranked objective, and subject to this we optimise the second ranked objective and so on. If all requirements are equally important for the traveller, we have to solve a multiobjective optimisation problem, where the objectives are all constraints required by the traveller. Each of the objectives is given a weight of importance. The problem is then to optimize the objective function, composed by the constraints, each of them multiplied by its weight. Multiobjective and lexicographic optimisation problems are discussed later in this work.

Note that most of the aforementioned constraints can be viewed as either hard or soft, depending on the user requirements. It is therefore suggested that any attempt at an investigation of TP assumes as an additional non-functional requirement that any proposed model to solve TP is flexible and can be easily extended by adding, removing and modifying the aforementioned constraints.

3 Worked Examples

We present an example instance of TP and comment on some of its solutions.

Example 1. A traveller wishes to visit 4 airports from a set of 7 airports available to travel to and from:

Glasgow (G), Berlin (B), Milan (M), Amsterdam (A), Paris (P), Frankfurt (F), London (L).

Airport G is the home point, F and L are connections, and B, M, A and P are the destinations. The travel time of the traveller is 15 days. All available flights are listed on Table 1. For simplicity, the duration of each flight is assumed to be 1 day. This means that if the traveller gets a flight at date x , they will reach the arrival airport at day $x + 1$.

Solution. A valid solution of the TP instance in the example above is the sequence of flights s , where the each flight is represented by its flight number, specified in the first column of Table 1:

$$s = \langle GA1, AP4, PM6, MF9, FB11, BG13 \rangle$$

Flight No	Departs	Arrives	Date	Price	
1	GA1	G	A	1	74
2	GF1	G	F	1	86
3	FB2	F	B	2	156
4	GL3	G	L	3	25
5	MF3	M	F	3	78
6	BP4	B	P	4	67
7	AP4	A	P	4	58
8	PM6	P	M	6	71
9	FM8	F	M	8	234
10	MF9	M	F	9	39
11	FA10	F	A	10	220
12	FB11	F	B	11	122
13	FM12	F	M	12	250
14	PL12	P	L	12	45
15	BG13	B	G	13	335
16	BL13	B	L	13	102
17	AG13	A	G	13	90
18	LG14	L	G	14	24

Table 1: List of flights with departure and arrival airports, flight date and price.

The total flights cost $c(s)$ is 699.

A valid solution with lower cost is the following sequence:

$$s' = \langle GA1, AP4, PM6, MF9, FB11, BL13, LG14 \rangle$$

Here $c(s')$ is 483 and hence s is not optimal.

Example 2. Given the same problem instance as in Example 1, suppose that the traveller has booked a ticket for a concert in B on day 3. The traveller requires to attend the concert.

Solution. In such case, neither s , nor s' from Example 1 are solutions, because both of them assign the traveller to be at a different location (airport A) at day 3. The following sequence is a solution:

$$s'' = \langle GF1, FB2, BP4, PM6, MF9, FA10, AG13 \rangle$$

The total cost $c(s'')$ is equal to 729, which is more expensive than s and s' .

4 List of Problems

This Section gives a list of known problems, referred to in this work when proving the NP-hardness of TP (Section 5) and when reviewing the existing work (Section 6). The problems, marked with an asterisk (*) next to their title, are known to be NP-complete [1].

1. Vehicle-Routing Problem * (VRP)

Instance. .

Question. .

2. Travelling Salesman Problem * (TSP) ¹

Instance. Set A of n cities, distance $d(A_i, A_j)$ between each pair of cities $A_i, A_j \in A$, positive integer B .

Question. Is there a tour of A having length B or less, i.e., a permutation of cities $\gamma = \langle A_{\pi(1)}, A_{\pi(2)}, \dots, A_{\pi(n)} \rangle$ of A such that the total travel distance L_γ :

$$L_\gamma = \left(\sum_{i=1}^{n-1} d(A_{\pi(i)}, A_{\pi(i+1)}) \right) + d(A_{\pi(n)}, A_{\pi(1)}) \leq B \quad ?$$

3. Travelling Salesman Problem Under the Triangle Inequality * (TSP- Δ) ²

¹Note that TSP is a special case of VRP when only one vehicle is allowed.

²Note that this problem is a special case of TSP.

Instance. Set A of n cities, distance $d(A_i, A_j)$ between each pair of cities $A_i, A_j \in A$ that satisfies the triangle inequality, positive integer B .

Question. Is there a tour of A having length B or less, i.e., a permutation of cities $\gamma = \langle A_{\pi(1)}, A_{\pi(2)}, \dots, A_{\pi(n)} \rangle$ of A such that the total travel distance L_γ :

$$L_\gamma = \left(\sum_{i=1}^{n-1} d(A_{\pi(i)}, A_{\pi(i+1)}) \right) + d(A_{\pi(n)}, A_{\pi(1)}) \leq B \quad ?$$

4. Time-Constrained TSP* (TCTSP)³

Instance. Set A of n cities, distance $d(A_i, A_j)$ between each pair of cities $A_i, A_j \in A$, positive integer B , lower and upper bounds l_i and u_i respectively for each city A_i that specify its time window.

Question. Is there a permutation of cities $\gamma = \langle A_{\pi(1)}, A_{\pi(2)}, \dots, A_{\pi(n)} \rangle$ of A , such that each city A_{π_j} is visited at time t_j , where $l_j \leq t_j \leq u_j$, $t_j < t_{j+1}$ for $(1 \leq j \leq n-1)$ and

$$L_\gamma = \left(\sum_{i=1}^{n-1} d(A_{\pi(i)}, A_{\pi(i+1)}) \right) + d(A_{\pi(n)}, A_{\pi(1)}) \leq B \quad ?$$

5. The Assignment Problem (AP)

Instance. Set A and set B with equal size, cost $c(a, b)$ of matching $a \in A$ to $b \in B$.

Question. Find a bijection $f : A \leftarrow B$ such that $\sum_{a \in A} c(a, f(a))$ is minimised.

6. Job-Shop Scheduling Problem (JSSP)

7. Longest Path Problem

5 Complexity of TP

We state a theorem about the complexity of TP and prove it.

Theorem 1. *TPD is NP-complete.*

Proof. This proof first shows the membership of TPD in the NP class of problems. Second, we prove the NP-hardness of TP by constructing a polynomial-time reduction from a known NP-complete problem Π to TPD, where Π is chosen to be TSP, defined in Section 4. Its NP-hardness follows by a reduction from the Hamiltonian Cycle problem. The proof is presented by Garey and Johnson [1].

Given an instance of TPD and s , which is a sequence of flights from F , we can write an algorithm that checks in polynomial time whether s is a solution. To accept or reject validity, the algorithm only needs to traverse s and check that it satisfies all required properties. Therefore, TP is in NP.

Let π be an instance of TSP. Let π' be an instance of TPD with the following properties:

³Note that this problem is a generalisation of TSP.

- The set of airports in π' is identical to the set of cities in π and it is similarly denoted as A (a city in π is called an airport in π'). Airport A_1 is the home point.
- Each airport in A is also a destination.
- The connection time C_{A_i} for each airport A_i is equal to 0.
- T is equal to n .
- Let C be the Cartesian product of the airports in A with itself, that is $C = A \times A = \{(A_i, A_j) : A_i \in A, A_j \in A, i \neq j\}$. Then F is a set of flights, such that for every $(A_i, A_j) \in C$, there exists a flight f_k in F , such that $A_k^d = A_i$ and $A_k^a = A_j$ for every date $0 \leq t < T$.
- For every $f_k \in F$, c_k is equal to $d(A_k^d, A_k^a)$ in π . Therefore, the flight costs also satisfy the triangle inequality.
- For every $f_k \in F$, $\Delta_k = 1$.
- B is the upper bound on the allowed total cost.

Suppose that $\gamma = \langle A_{i_1}, A_{i_2}, \dots, A_{i_n} \rangle$ is a solution to π , where $\langle i_1, \dots, i_n \rangle$ is a permutation of $\langle 1, \dots, n \rangle$ and the total travel distance $L_\gamma \leq B$. **Without loss of generality, assume that $i_1 = 1$.** In π' , γ is equivalent to the order of visited airports by some sequence of flights $s = \langle f_{j_1}, f_{j_2}, \dots, f_{j_n} \rangle$, such that for each p ($1 \leq p \leq n$) there exists q ($1 \leq q \leq n$) such that $A_{j_q}^d = A_{i_p}$ and $A_{j_q}^a = A_{i_{p+1}}$, where subscripts are taken modulo n . **Therefore, s satisfies the first property of a valid solution.** For each q ($1 \leq q \leq n$), $A_{j_q}^a = A_{j_{q+1}}^d$ and $t_{j_q} = q - 1$. We know that such flights exist in F by the construction of the set F .

From the construction of s , it follows that property 2 also holds. Properties 3 and 4 also hold, since we have chosen flights from F such that for every $f_{j_q} \in s$, $t_{j_q} = q - 1$ ($1 \leq q \leq n$). Property 5 is satisfied, since all airports in A are destinations.

Since the cost of every flight in F is equal to the distance between the two cities in π that correspond to its departure and arrival airport, it follows that $c(s) = L_\gamma \leq B$.

The sequence s satisfies all requirements for a valid solution to π' . Therefore, a solution of π is also a solution to π' .

Conversely, suppose that $s = \langle f_{j_1}, \dots, f_{j_k} \rangle$ is a solution to π' , where the flights in s visit destinations in the sequence $\gamma' = \langle A_{i_1}, A_{i_2}, \dots, A_{i_m} \rangle$. We will prove that γ' is a solution of π .

By construction of π' , all airports in A are also destinations. Therefore, γ' contains all cities in A , that is $m \geq n$. Suppose that $m > n$ and an arbitrary airport A_p ($1 \leq p < n$) is included more than once in γ' . Then s must contain more than one flight with arrival airport equal to A_p . The duration of each flight in π' is one day. Therefore, for every q ($1 \leq q < n - 1$), $t_{j_{q+1}} = q$. Since the traveller has only n days of total travel time, and s is restricted to contain exactly n flights, that is $k = n$. The only way to visit n distinct destinations, given n flights is that all flights in s have unique arrival airports. Assuming that A_p is visited more than once means that there is

more than one flight with arrival airport equal to A_p , which is a contradiction. Therefore, $m = n$ and each airport in A is visited exactly once.

From the properties of s it follows that $A_{j_1}^d = A_{j_n}^a = A_1$. Therefore, γ' is a cycle of size n . We know that $c(s) \leq B$. We assigned a cost of each flight f_k in F to be equal to $d(A_k^d, A_k^a)$ in π . Therefore, the total travel distance $L_{\gamma'} = c(s)$ which is less than or equal to B .

According to the specifications of π , γ' is a solution to π . Therefore, a solution to π' is also a solution to π .

The transformation from a TSP instance to an instance of TPD can be done in polynomial time. For each of the $n(n-1)/2$ distances $d(A_i, A_j)$ that must be specified in π , it is sufficient to check that the same cost is assigned to the flights from A_i to A_j for all dates.

Therefore, TP is in NP and the decision version of TSP can be reduced to TPD in polynomial time, from which it follows that TPD is NP-complete.

□

6 Background Survey

We present some existing approaches to dealing with NP-hard problems, namely branch and bound (Section 6.2), constraint programming (Section 6.3) and integer programming (Section 6.1). The way each of these methods tackles NP-hardness is illustrated with a specific problem as an example, defined in Section 4.

6.1 Linear and Integer Programming

Linear Programming (LP) is a way to solve problems by modeling their requirements as a system of linear inequalities and equations and subject to them finding an optimal solution to the problem, expressed as a function that has to be either minimised or maximised. This function is called the *objective function*.

The field of LP is well-studied [? ?] and it has many applications in the industry for resource optimisation [? ?]. However, [?] shows that there are many important optimisation problems that can not be modeled as a linear program, because their variables can take only integer values. Such problems can be modeled with the tools of integer programming (IP). Integer programming (IP) is an extension of linear programming (LP). The difference between LP and IP is that in the IP model variables are restricted to take only integer values.

The rest of this section presents different methods to solve problems with the tools offered by IP and comments on their performance by using TSP as an example problem.

6.1.1 Integer Programming Formulation of TSP

we formulate TSP as the following IP problem. Let $x(i, j)$ be a variable that denotes whether city j succeeds city i in a TSP tour. The value of $x(i, j)$ is equal to 1 if this is true, or 0 if it is false. If $x(i, j)$ is 1, then i is called the *outgoing* city and j - the *incoming* city. The objective function of TSP is to minimise the total length of the tour, that is:

$$\min \sum_{i \in A} \sum_{j \in A} d(i, j) x(i, j) \quad (1)$$

In each TSP tour every city is visited once. Moreover, each city has to be incoming and outgoing exactly once. Therefore, the value of $x(i, j)$ for every i will be 1 for only one j and 0 for the rest. This can be enforced with the following constraint:

$$\begin{aligned} \sum_{j \in A} x(i, j) &= 1, & i \in A, \\ \sum_{i \in A} x(i, j) &= 1, & j \in A, \end{aligned} \quad (2)$$

Constraint (2) ensures that each city in A is picked exactly once as an incoming and an outgoing city. However, it allows for the existence of one or more cycles of $n_1 < n$ cities, called *subtours*. For instance, consider Figure 1 and Figure 2. Both of them satisfy constraint (2). However, Figure 2 does not represent a valid tour, because it contains two subtours, each of size 3. To tackle this problem, we introduce the *subtour elimination* constraint, that is:

$$\sum_{S \subseteq A} x(i, j) \geq 2, \quad (i, j) \in (S \times \bar{S}) \cup (\bar{S} \times S) \text{ for all } S \subset A, \bar{S} = A \setminus S, \quad \bar{S}, S \neq \emptyset \quad (3)$$

Here, S is a subset of A and \bar{S} is the set of all cities that are in A and not in S . Constraint (3) enforces that at least two cities in S are connected with cities from \bar{S} . We give Figure 1 and Figure 2 as an example. Let $S = \{A_1, A_2, A_3\}$, then $\bar{S} = \{A_4, A_5, A_6\}$. Constraint (3), would detect the tour on Figure 2 as invalid, as there is no city in S that is connected to a city in \bar{S} . Figure 1 will be accepted, since A_1 and A_3 are connected with A_6 and A_4 respectively.

The formulation of the general TSP problem is given by the objective function (1), which has to be minimised, subject to constraints (2) and (3).

6.1.2 The Assignment Problem Relaxation

The assignment problem (AP) relaxation consists of removing the subtour elimination constraint from the TSP IP model and minimising the objective function (1), only subject to constraint (2).

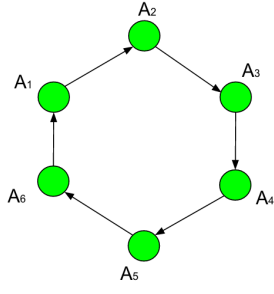


Figure 1: A valid TSP tour

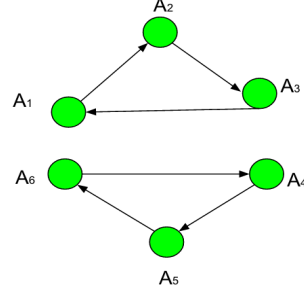


Figure 2: Two subtours of size 3

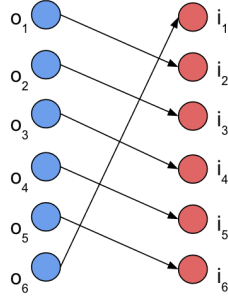


Figure 3: AP solution equivalent to Figure 1

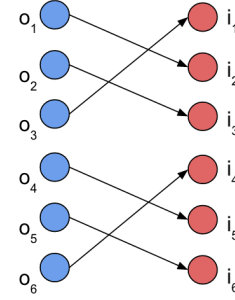


Figure 4: AP solution equivalent to Figure 2

It is well known and extensively used relaxation [? ? ? ? ?]. In this section we explain how this method is used for solving NP-hard problems, using TSP as an example.

Let π' be the problem that results after performing AP relaxation on some instance π of TSP. We will show that π' is equivalent to the Assignment Problem (AP).

The model for π' can be expressed as a bipartite graph G like the graphs shown on Figure 4 and Figure 3. G has two sets of vertices: $O = o_1, \dots, o_n$ and $I = i_1, \dots, i_n$ and a set of edges $E = \{(o_p, i_q) \mid a_p, a_q \in A, p \neq q\}$ for $a_p, a_q \in A$, where I and O are both identical to A . Each edge $e = (o_p, i_q)$ in E has weight w_e equal to $d(a_p, a_q)$ and it represents a path from some city $o_p \in O$ to some city $i_q \in I$. Finding a solution to π' is equivalent to finding a perfect matching in G of minimum value of the sum of weight of the edge between every two matched vertices. Each vertex in O then represents an outgoing city and each vertex in I represents an incoming city.

Problem π' asks for assigning each vertex in O to a vertex in I , which is equivalent to an instance of AP with $A = O$, $B = I$ and $w_e = c(a_p, b_q)$, $e = (o_p, i_q)$, $p \neq q$ or $w_e = \infty$ if $p = q$.

Note that a solution to π' can be either a tour or a collection of subtours (since we relaxed the subtour elimination constraint). Figure 4 and Figure 3 are both solutions to AP obtained after performing an AP relaxation to a TSP instance with 6 cities. They are examples of how not all solutions to π' are solutions to π . Let the optimal solution to π' π'^* have total edge weight equal to α . Suppose that it is an invalid tour in π . Therefore, the tour in π^* is of length bigger than α . Conversely, suppose that π'^* is a valid tour in π . Therefore, the length of the tour in π^* is equal to α . Hence, the length of an optimal solution to π^* is bigger than or equal to the value of π'^* . In the existing work, this feature is used to derive a lower bound on the cost of the most optimal

tour [? ? ?]. This approach is frequently combined with a branch and bound technique [?]. We discuss this approach more in the next section.

The relation between TSP and AP is explained very nicely by ?]

The solution by ?] continues with explanation of the cutting planes method. Although this has been one of the most discussed topics in research about TSP, we skip a detailed discussion about it, because it is not directed to TP. For more information, we point the reader back to [?], (cite Dantzig, Flukerson, Johnson 1959 paper), where the method is explained with more detail using a 10 cities TSP instance.

6.2 Branch and Bound Algorithms

1. Bounds based on the Assignment Problem
2. Lin-Kernigham heuristic (2-opt)

6.3 Constraint Programming

Constraint Programming (CP) is like integer programming

what is a variable, variable domain?

domain wipeout?

search tree?

explain a model of the vehicle routing problem or jjsp with cp

what is the difference between cp and ip?

6.3.1 Local Search

6.3.2 Heuristics

7 Proposed Approach

We use some of the concepts of the existing work to construct a CP and IP models for TP.

7.1 CP Model

We have the following variables:

- An array f of m , where $m = |F|$. For each $0 < i < m$, $f[i]$ is 1 if flight $F[i]$ is taken in a TP tour, or 0 otherwise.
- an array s of size k , which is the sequence of flights taken during the trip. Each variable in s has domain $0, k$. If $s[i]$ is 0, then the trip is finished. Therefore, the value of every subsequent variable in s will also be 0. If no variables in s are equal to 0, then the solution is a trip consisting of all flights.
- an array c of size n which represents the number of times each city is visited during the trip. The value of each variable must be even.

We have the following constraints:

1. for every $i > 0$

7.2 IP Model

7.3 Experimental Datasets

7.3.1 Skyscanner Flights API

7.3.2 Randomly Generated TP instances

8 Work Plan

show how you plan to organize your work, identifying intermediate deliverables and dates.

References

- [1] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Bell Telephone Laboratories, Incorporated, United States of America, 1979.