Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Approximation and complexity of multi-target graph search and the Canadian traveler problem

Martijn van Ee [a,*,1], René Sitters [a,b]

[a] Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands
[b] Centrum voor Wiskunde en Informatica (CWI), Science Park 123, 1098 XG Amsterdam, the Netherlands

A R T I C L E I N F O

Article history:
Received 29 September 2017
Received in revised form 9 April 2018
Accepted 11 April 2018
Available online xxxx
Communicated by J. Díaz

Keywords:
Canadian traveler problem
Graph search problem
Approximation algorithms
Computational complexity
Routing under uncertainty

A B S T R A C T

In the Canadian traveler problem, we are given an edge weighted graph with two specified vertices s and t and a probability distribution over the edges that tells which edges are present. The goal is to minimize the expected length of a walk from s to t. However, we only get to know whether an edge is active the moment we visit one of its incident vertices. Under the assumption that the edges are active independently, we show NP-hardness on series-parallel graphs and give results on the adaptivity gap. We further show that this problem is NP-hard on disjoint-path graphs and cactus graphs when the distribution is given by a list of scenarios. We also consider a special case called the multi-target graph search problem. In this problem, we are given a probability distribution over subsets of vertices. The distribution specifies which set of vertices has targets. The goal is to minimize the expected length of the walk until finding a target. For the independent decision model, we show that the problem is NP-hard on trees and give a $(3.59+\epsilon)$-approximation for trees and a $(14.4+\epsilon)$-approximation for general metrics. For the scenario model, we show NP-hardness on star graphs.

1. Introduction

In this paper, we consider the *Canadian traveler problem* (CTP). Here, we are given an edge-weighted graph $G=(V,E)$ with two specified vertices s and t, and a probability distribution $p: 2^E \to [0,1] \cap \mathbb{Q}$. This function gives for each $A \subseteq E$ the probability that A is the active set of edges. Only the active edges are present. We need to construct a walk from s to t. However, we only know whether an edge is present when we visit one if its incident vertices. The problem is to find a policy that minimizes the expected length of our walk. Here, a policy may use the observed realizations as input to decide where to go next. We consider the problem in the independent decision model and in the scenario model. In the independent decision model, each edge has a probability of being present and the event of an edge being present is independent of the other edges. In the scenario model, we are given an explicit list of scenarios $\mathcal{S}$, where each scenario describes which edges are present. The probability distribution p is known in advance. We assume that the edge weights are integer.

It was shown in [20] that the decision version of CTP is NP-complete in the scenario model. They also showed that the problem is polynomially solvable if the number of scenarios is bounded by a constant. In the independent decision model,

* Corresponding author.
E-mail address: M.v.Ee.01@mindef.nl (M. van Ee).
[1] Present address: Netherlands Defence Academy, Het Nieuwe Diep 8, 1781 AC Den Helder, The Netherlands.

https://doi.org/10.1016/j.tcs.2018.04.022

the decision version of CTP is PSPACE-complete [10] and it is #P-hard to compute the expected length [19,20]. It is even not possible to describe an optimal policy, unless PSPACE $\subseteq$ P/poly [24]. On the other hand, the problem is polynomially solvable on disjoint-path graphs [5]. Until now, the computational complexity of the problem was still open on series-parallel graphs. The complexity of CTP on this class of graphs was mentioned as one of the major open problems in [18]. Fried [9] considered CTP on a subset of graphs which become trees after deleting t. He conjectured that CTP is intractable in this case. It is easy to see that the class of graphs considered by Fried is a subset of the class of series-parallel graphs. A consequence of our work is that CTP is indeed NP-hard in both cases.

The problem also has an adversarial version, i.e., there is no probability distribution but there is an adversary that chooses the edges that are present. In this problem, we compare the length of the walk with the offline optimum, i.e., the optimal solution if we had full knowledge of the edges that are present. The worst-case ratio between these values is called the competitive ratio. It is reasonable to consider the restriction that at most k edges fail. This variant is called *k-CTP* and was introduced by Bar-Noy and Schieber [3]. It was shown in [25] that the Backtrack-algorithm that repeatedly chooses the shortest path and returns when the path is blocked, is $2k+1$-competitive. Westphal [25] also showed that no algorithm can beat this bound. In our setting, the Backtrack-algorithm is an $O(n)$-approximation in the independent decision and in the scenario model, since we have full information about the graph after at most n turns. The Backtrack-algorithm is also an $O(|\mathcal{S}|)$-approximation in the scenario model, since we know which scenario is active after at most $|\mathcal{S}|$ turns. The main open problem is to improve these approximability results.

An important modeling issue, when considering approximation algorithms, is how to deal with an st-disconnected graph. In this case, no walk can reach t. In the independent decision model, this is usually solved by adding an edge from s to t that is present with probability one and has an arbitrary large length. This modeling choice does not influence the computational complexity of the problem. However, it does influence the analysis of approximation algorithms. To get a sensible model from this perspective, we choose to minimize the conditional expectation of the length given that G contains an st-path. Equivalently, the value of a walk is zero whenever the active set of edges induces an st-disconnected graph. This way, we get an objective value equal to the conditional expected length times the probability of having an st-connected graph. We will use this objective function in this paper. For the independent decision model this is a stronger formulation in the sense that if we have an α-approximation in this formulation, we also have an α-approximation in the former one, but not vice versa. In the scenario model we can simply avoid this issue by deleting scenarios that induce an st-disconnected graph and normalize the remaining probabilities.

Before giving results on CTP, we consider its relation with the *multi-target graph search problem* (multi-target GSP), a generalization of the *graph search problem* (GSP) [16,2]. The GSP is formally defined as follows.

Definition 1. Suppose we are given an edge-weighted graph $G=(V,E)$ with root s and probability p_v for each vertex v, such that $\sum_v p_v = 1$. For a given path starting at s, the latency of v, C_v, is defined as the length of the subpath from s to v. In the graph search problem (GSP), the goal is to find a path on all vertices of V that starts in s and minimizes $\sum_v p_v C_v$, i.e., the total weighted latency.

The probability distribution in the GSP specifies the probability that the target is at the corresponding vertex. Hence, the goal in the GSP is to find a walk along the vertices that minimizes the expected length until finding the target. When all probabilities are equal (or polynomially bounded), the problem reduces to the *traveling repairman problem* (TRP), also known as the *minimum latency problem*. The TRP is formally defined as follows.

Definition 2. Suppose we are given an edge-weighted graph $G=(V,E)$ with root s. For a given path starting at s, the latency of v, C_v, is defined as the length of the subpath from s to v. In the traveling repairman problem (TRP), the goal is to find a path on all vertices of V that starts in s and minimizes $\sum_v C_v$, i.e., the total latency.

On general metrics, the current best approximation guarantee for TRP is 3.59 [6]. The problem is even NP-complete on weighted trees [22], but admits a PTAS in the Euclidean plane and on weighted trees [23]. In [2], the authors gave a 40-approximation for GSP by using similar techniques.

Here, we generalize the GSP to the case where there can be multiple targets. For this, we again consider the independent decision model and the scenario model. In the independent decision model, each vertex has a probability of having a target and the event of having a target at a vertex is independent of the presence of targets at other vertices. In the scenario model, we are given an explicit list of scenarios $\mathcal{S}$, where each scenario describes at what vertices a target is present. Now, we want to minimize the expected length of the walk until a target has been found, given that there is at least one target. The following theorem states how the multi-target GSP is related to the CTP. Recall that our objective value is equal to the conditional expected length times the probability of having an st-connected graph.

Theorem 1. *The multi-target GSP is equivalent to a special case of the CTP.*

Proof. We prove the theorem for the independent decision model. The proof for the scenario model is similar and omitted here. Given an instance of multi-target GSP, i.e., an edge-weighted graph $G=(V,E)$ with root s and a probability p_v for