

Special Cases of the Minimum Spanning Tree Problem under Explorable Edge and Vertex Uncertainty

Corinna Mathwieser
RWTH Aachen University, Germany
mathwieser@combi.rwth-aachen.de

Eranda Çela
TU Graz, Austria
cela@math.tugraz.at

November 29, 2022

Abstract

This article studies the Minimum Spanning Tree Problem under Explorable Uncertainty as well as a related vertex uncertainty version of the problem. We particularly consider special instance types, including cactus graphs, for which we provide randomized algorithms. We introduce the problem of finding a minimum weight spanning star under uncertainty for which we show that no algorithm can achieve constant competitive ratio.

1 Introduction

Many real world problems do not allow to work with precise data as parts of the input are uncertain or only known approximately. Different approaches to deal with uncertainty include stochastic optimization where the input data is known to follow a specific probability distribution, robust optimization which aims to find good solutions for all possible inputs and explorable uncertainty. In the latter setting, it is possible to obtain more precise or even exact data by making queries. However, any query causes exploration cost. In an applied scenario this might be time, money or other resources which are needed for further measurements.

In this paper, we consider the *Minimum Spanning Tree Problem under Explorable Uncertainty* (MST-U) where the goal is to find a minimum spanning tree (MST) in an edge weighted graph where the weights are initially not known but can be revealed upon request. In an instance of MST-U, each edge is equipped with an uncertainty set and a query cost. The uncertainty set, usually an interval, is guaranteed to contain the edge's weight. An edge query reveals the edge's true weight. The goal is to find a set Q of queries of minimum cost which allows to find a minimum spanning tree with certainty, i.e. given the weights of

all edges in Q , there exists an edge set T such that T is a minimum spanning tree for all possible weights with respect to the remaining uncertainty sets. The queries may be chosen adaptively which means that we are allowed to choose the next update based on the previous outcomes of edge queries. Note that while MST-U requires the specification of an edge set which corresponds to an MST, it is not necessary to compute the MST weight.

1.1 Related work

The first work on problems where parts of the input are uncertain and can be queried is due to Kahan (see [10]) who studied the problem of finding the maximum, the median and the minimum of a set of real values, each of which is known to lie in a given interval. Since then, explorable uncertainty has been considered for different combinatorial problems, e.g. shortest paths (see [6]), scheduling (see [2]) and the Knapsack Problem (see [9]). Erlebach et al. were the first to introduce the Minimum Spanning Tree Problem under Explorable Uncertainty in [5]. They showed that without restrictions made on the uncertainty sets, no algorithm can achieve constant competitive ratio. This is why subsequently it is assumed that uncertainty sets are either singletons or do not contain neither their supremum nor their infimum (e.g. open intervals). [5] also presented the deterministic algorithm U-RED for MST-U with uniform query costs which achieves competitive ratio 2 and proved that no deterministic algorithm can have a smaller competitive ratio. Moreover, they introduced a different version of this problem, the *Minimum Spanning Tree Problem under Vertex Uncertainty* (V-MST-U) where vertices are points with uncertain locations in the plane and the edge weights correspond to the distances between the respective end vertices. They showed that U-RED can be adapted to work for the vertex uncertainty setting as well if uncertainty sets are (topologically) open. An important question left open was the effect of randomization in the setting of MST-U. This question was subsequently answered by Megow et al. in [11] where they provided the randomized algorithm RANDOM with competitive ratio $1 + \frac{1}{\sqrt{2}}$. The best known bound for the performance of randomized algorithms is 1.5 which holds true even for triangles and was observed by Erlebach and Hoffmann in [4]. [11] transformed their randomized algorithm into the deterministic algorithm BALANCE which achieves a competitive ratio of 2 on instances with general query cost. [11] also considered the problem of finding the weight of an MST under uncertainty, the problem of finding an α -approximate MST under uncertainty as well as a version of MST-U where queries return subintervals instead of the precise edge weights. Test results for MST-U are provided by Focke et al. in [7]. In [3], Erlebach and Hoffmann deal with the *verification problem* for MST-U, i.e. the problem of computing an optimal query set if the uncertainty sets as well as the exact edge weights are given. They show that the verification problem for MST-U is solvable in polynomial time while the verification problem for the vertex uncertainty problem V-MST-U is NP-hard.

1.2 Our Contribution

One of the main difficulties in solving MST-U stems from dealing with edges that are part of several cycles. Thus, it is a natural question to consider the special case of cactus graphs where every edge belongs to at most one cycle. However, the algorithm RANDOM presented in [11] reaches its worst case competitive ratio of $1 + \frac{1}{\sqrt{2}} \approx 1.71$ even for instances where the input graph is a cycle. We introduce an algorithm which achieves competitive ratio 1.5 on instances with cactus graphs. V-MST-U on the contrary, has only been considered in the setting of deterministic algorithms so far. We prove that there exists no randomized algorithm for V-MST-U with a competitive ratio better than 2.5 even if the input graph is a cycle. Unfortunately, structural differences between the edge uncertainty setting and the vertex uncertainty setting impede the straight forward adaption of RANDOM to the vertex uncertainty setting. Instead, we consider the special case of cactus-like instances where no two cycles share a non-trivial vertex and introduce the algorithm V-RANDOM_C for which we prove a competitive ratio of at most 2.5. While deterministic and randomized algorithms with reasonable performance guarantee exist for MST-U, this is not necessarily the case when aiming for more restricted graph classes than spanning trees. We demonstrate this by introducing the *Minimum Spanning Star Problem under Explorable Uncertainty* (MSS-U). MSS-U is defined analogously to MST-U except that we want to identify a spanning star of minimum weight rather than a general spanning tree. For MSS-U we derive a negative result with respect to competitive analysis, i.e we show that no algorithm for MSS-U can achieve constant competitive ratio.

The remainder of this paper is organized as follows: In Section 2 we provide the precise definition of MST-U and further basic definitions and concepts used throughout the rest of the paper. Section 3 deals with a randomized algorithm for MST-U on cactus graphs. In Section 4 we prove a lower bound for the performance of randomized algorithms for V-MST-U and introduce an optimal randomized algorithm for V-MST-U on cactus-like graphs. In Section 5 we study the Minimum Spanning Star Problem under Explorable Uncertainty. We conclude with a brief summary and some open questions in Section 6.

2 Preliminaries

In this section we will introduce definitions, notation and structural results used throughout this paper.

2.1 Problem definition MST-U and notation

Definition An *(edge-)uncertainty graph* is a tuple $\mathcal{G} = (G, (A_e)_e)$ where $G = (V, E)$ is an undirected, connected graph and for each edge $e \in E$, $A_e \subset \mathbb{R}$ is either a singleton set or a set which neither contains its infimum nor its supremum. The sets A_e with $e \in E$ are called *uncertainty sets*.

An instance of the Minimum Spanning Tree Problem under Explorable Uncertainty (MST-U) is specified in terms of an uncertainty graph $\mathcal{G} = (G, (A_e)_e)$ as well as an a priori unknown edge weight $w_e \in A_e$ and a query cost $q_e > 0$ for each $e \in E$. We denote by $n = |V|$ the number of vertices and by $m = |E|$ the number of edges of G . If A_e contains a single element only we say that A_e is *trivial*. An edge e is trivial if A_e is trivial. For an edge $e \in E$, we denote by $L_e := \inf A_e$ the infimum and by $U_e := \sup A_e$ the supremum of the uncertainty set. We will also refer to L_e and U_e as *the lower and the upper limit* of A_e (or of e) respectively. We can query an edge e to determine its weight w_e . If we query e , the set A_e is updated to a singleton set containing only w_e . The cost of querying an edge e is $q_e > 0$. The goal is to find an edge set T of a minimum spanning tree (MST) in G with respect to edge weights w_e while minimizing the total cost of queries needed to find T . More precisely, a feasible query set is defined as follows:

Definition Given an uncertainty graph \mathcal{G} with graph $G = (V, E)$ and uncertainty sets $A_e, e \in E$, as well as edge weights $w_e \in A_e$ for $e \in E$, a set $Q \subset E$ is called a *feasible query set* if there exists a spanning tree T in G such that T has minimum weight with respect to any weight function $\bar{w} : E \rightarrow \mathbb{R}$ which fulfills that $\bar{w}(e) = w_e$ if $e \in Q$ and $\bar{w}(e) \in A_e$ if $e \in E - Q$. We say that Q *verifies* T .

MST-U thus consists in finding a feasible query set Q of minimum query cost $\sum_{e \in Q} q_e$. Moreover, we will use the following definition: Given an uncertainty graph $\mathcal{G} = (G, (A_e)_e)$ and a cycle C in G , we say that an edge f is *always maximal* in C if $L_f \geq U_e$ for all e in $C - f$.

2.2 Vertex uncertainty problem

In an instance of the Minimum Spanning Tree Problem under Explorable Vertex Uncertainty (V-MST-U) we are given an undirected, connected graph G where each vertex corresponds to a point in the Euclidean plane. The weight of an edge is determined by the distance between its end vertices. Instead of the precise location of a vertex v we are given an uncertainty set $A_v \subset \mathbb{R}^2$. A_v can either be a singleton set (in which case we refer to the vertex and the uncertainty set as trivial) or an open subset of \mathbb{R}^2 . An algorithm can query a vertex v at query cost q_v to reveal its exact location. We define a *(vertex-)uncertainty graph* as well as a *feasible (vertex) query set* in the same way as in the setting of MST-U. Then V-MST-U is defined analogously to MST-U, i.e. we want to identify a feasible vertex query set of minimum query cost.

In order to apply methods developed for MST-U to V-MST-U it is common to transform vertex uncertainty sets into edge uncertainty sets by computing all possible distances between vertices. The resulting instance is referred to as associated edge instance:

Definition Given an instance \mathcal{S} of V-MST-U with graph $G = (V, E)$ and uncertainty sets $A_v, v \in V$, the *associated edge instance* \mathcal{S}' is an instance of MST-U

with graph G and uncertainty sets $A_{\{u,v\}} = \{d(u', v') | u' \in A_u, v' \in A_v\}$.

2.3 Performance analysis

To analyze the quality of a solution found by an algorithm we compute the competitive ratio between the query cost of the algorithm's solution and the cost of an optimal query set. An optimal query set is an optimal solution to the offline problem where all edge weights (or vertex positions) are known a priori.

Definition Let \mathcal{I} be an instance of MST-U (or V-MST-U). By $OPT(\mathcal{I})$ we denote the cost of an optimal query set for instance \mathcal{I} . For an algorithm ALG we denote by $ALG(\mathcal{I})$ the cost of the query set which the algorithm outputs when applied to \mathcal{I} . We say that ALG achieves *competitive ratio* $c \geq 1$ or is *c-competitive* if

$$\frac{ALG(\mathcal{I})}{OPT(\mathcal{I})} \leq c$$

for all instances \mathcal{I} . A randomized algorithm is said to achieve competitive ratio $c \geq 1$ if the ratio between the expected query cost of the algorithm's solution and the cost of an optimal solution is at most c , i.e. if

$$\frac{\mathbb{E}(ALG(\mathcal{I}))}{OPT(\mathcal{I})} \leq c$$

for all instances \mathcal{I} .

2.4 Structural aspects of MST-U and V-MST-U

In the following we will recall some structural insights into MST-U which were provided by Megow et al in [11] and will be used throughout the remainder of the paper. Given an instance of MST-U, a *lower limit tree* is the edge set of a minimum spanning tree in G where all edge weights are set equal to the lower limit of the edge's uncertainty set. An *upper limit tree* is defined analogously.

Lemma 1 ([11]) *Let T_L, T_U be a lower and an upper limit tree respectively. All edges in $T_L \setminus T_U$ with non-trivial uncertainty sets lie in any feasible query set.*

Hence, the instance can be preprocessed by querying all edges in $T_L \setminus T_U$ and we may thus assume that $T_L = T_U$. Lemma 1 can be translated to the setting of vertex uncertainty in the following:

Lemma 2 *Given a V-MST-U instance \mathcal{I} , let T_L, T_U be a lower and an upper limit tree of the associated edge instance \mathcal{I}' . Then every feasible query set for \mathcal{I} contains a vertex cover of $\{e \in T_L \setminus T_U | e \text{ contains a non-trivial vertex}\}$ which consists of non-trivial vertices only.*

Now recall the following well-known properties of minimum spanning trees: Given a cycle C in a weighted graph G , we have that if an edge e is such

that $w_e > w_{e'}$ for all $e' \in C - e$ then e is not contained in any MST. If $w_e \geq w_{e'}$ for all $e' \in C$ then there exists an MST of G which does not contain e . In the context of MST-U this means that if we encounter a cycle C with an always maximal edge e then e can be discarded in the search for an MST. Conversely, given a cycle C and a non-trivial edge f which has largest upper limit U_f in C , it is impossible to verify an MST which contains f without querying f because f is a candidate for an edge with strictly largest weight in C . Possible candidates for a largest weight edge in C are f and all edges e in C with possibly larger weight than f , i.e. with $U_e > L_f$. We will refer to these edges as *neighbors* of f and denote the set of all neighbors by $X(f) := \{e \in C - f | U_e > L_f\}$. [11] prove the following with respect to f and its neighbor set:

Lemma 3 ([11]) *Let T_U be an upper limit tree. Let f be an edge in $G - T_U$ and let C be the cycle in $T_U + f$. If no edge in C is always maximal then any feasible query set contains f or $X(f)$. Moreover, if there is an edge $g \neq f$ in C such that $L_g \geq L_f$ and $U_g \neq L_f$ then f lies in any feasible query set. If $T_U = T_L$ is a lower limit tree too and no edge is known to have maximum weight even after querying f or all edges in $X(f)$ then an edge g in C with maximum upper limit U_g lies in any feasible query set.*

In the context of V-MST-U, we can derive the following result from Lemma 3:

Lemma 4 *Consider an instance \mathcal{S} of V-MST-U and let T_U be an upper limit tree for the associated edge instance \mathcal{S}' . Let f be an edge in $G - T_U$ and let C be the cycle in $T_U + f$. If no edge in C is always maximal in \mathcal{S}' then any feasible query set for \mathcal{S} contains a non-trivial vertex in f or a vertex cover of $\{e \in X(f) | e \text{ contains a non-trivial vertex}\}$ which consists of non-trivial vertices only. Moreover, if there is an edge $g \neq f$ in C such that $L_g \geq L_f$ and $U_g \neq L_f$ then any feasible query set of \mathcal{S} contains a non-trivial vertex in f .*

We omit a formal proof of Lemma 2 and of Lemma 4 as they follow directly from Lemma 1 and Lemma 3 and the fact that an MST for a vertex-uncertainty graph can only be identified if the edge-uncertainty graph of the associated edge instance allows to identify an MST.

3 A randomized algorithm for MST-U on cactus graphs

In this section we consider a randomized algorithm for MST-U on cactus graphs. A *cactus graph* is a connected graph in which any two cycles share at most one vertex. A feasible query set needs to allow for the detection of a maximum weight edge in each cycle. It can be easily seen that an optimal query set for a cactus graph consists of the disjoint union of the optimal query sets for each of the graph's cycles. Thus we can first consider MST-U on a cycle and then extend the result to the case of an edge uncertainty graph which is a cactus. Once we are able to treat cycles separately, it is possible to achieve an optimal competitive ratio of 1.5 using the following observation: Assume that we have

preprocessed the instance such that $T_L = T_U$ and let f be an edge in $G - T_L$ and let C be the cycle in $T_L + f$. Assume moreover that no edge in C is always maximal. Then Lemma 3 guarantees that if we start by querying f and continue to query edges in order of decreasing upper limit until an edge in C is always maximal, then we have queried at most one edge which is not in the optimal solution.

Theorem 5 *For cactus graphs there exists an algorithm RANDOM_C with competitive ratio at most 1.5, which is optimal. Moreover, if G is a cycle, RANDOM_C achieves competitive ratio $1 + \frac{q_{X(f)} \cdot q_f}{q_f^2 + q_{X(f)}^2}$ where $q_{X(f)}$ denotes the cumulative query cost of edges in $X(f)$.*

Proof We first consider a graph C which consists of a single cycle. Assume again that we have preprocessed the instance such that $T_L = T_U$ and that no always maximal edge in C is known. Let f be the edge in $E \setminus T_L$ and let $q_{X(f)} := \sum_{e \in X(f)} q_e$ be the query cost of all neighbors of f in C . With probability p , our algorithm starts by querying all edges in $X(f)$. With probability $1 - p$, its first step is to query f . Once it has queried $X(f)$ or f , it queries edges in order of decreasing upper limit until an always maximal edge in C can be identified.

We distinguish two cases: either an optimal solution queries f or it does not. If an optimal solution does not query f , it must by Lemma 3 query all of the neighbors in $X(f)$ and thus makes queries at cost $q_{X(f)}$. In this case with probability p , we query the same edges as the optimal solution and achieve competitive ratio 1. With probability $1 - p$, RANDOM_C queries f and possibly all edges in the neighbor set $X(f)$ such that the competitive ratio is at most $\frac{q_{X(f)} + q_f}{q_{X(f)}}$. Thus, in this case the overall competitive ratio is at most $1 + (1 - p) \frac{q_f}{q_{X(f)}}$.

If an optimal solution queries f , RANDOM_C queries the same edges as the optimal solution if it starts by querying f , i.e. with probability $1 - p$. With probability p , RANDOM_C starts by querying all neighbors in $X(f)$ and might have to query f too, while an optimal query set might contain f only. Summing up, the competitive ratio is bounded by $1 + p \cdot \frac{q_{X(f)}}{q_f}$ in this case. By setting $p = \frac{q_f^2}{q_f^2 + q_{X(f)}^2}$, the obtained bounds for both cases coincide and equal

$$1 + \frac{q_{X(f)} q_f}{q_f^2 + q_{X(f)}^2}.$$

Note that the following equivalences hold:

$$\begin{aligned} \frac{q_{X(f)} \cdot q_f}{q_f^2 + q_{X(f)}^2} &\leq \frac{1}{2} && \Leftrightarrow \\ 2q_{X(f)} q_f &\leq q_f^2 + q_{X(f)}^2 && \Leftrightarrow \\ 0 &\leq (q_f - q_{X(f)})^2. \end{aligned}$$

Thus, RANDOM_C is 1.5-competitive on instances where the uncertainty graph is a cycle if p is chosen as above.

For an instance \mathcal{I} where the uncertainty graph is a general cactus G , we again preprocess the instance such that $T_L = T_U$ is a lower and an upper limit tree. Denote the edges in $G - T_L$ by f_1, \dots, f_{m-n+1} , the order is arbitrary. Let C_i denote the cycle in $T_L + f_i$. We apply RANDOM_C for cycles as described above to each of the cycles separately. Let Q_i^* denote an optimal query set for an instance where the uncertainty graph consists only of C_i and the uncertainty sets and weights are as in \mathcal{I} . We set $OPT_i := |Q_i^*|$, $i = 1, \dots, m - n + 1$. As any two cycles in G do not share common edges and edges that lie in no cycle need to be part of any spanning tree, the disjoint union $Q^* := \dot{\bigcup}_{i=1}^{m-n+1} Q_i^*$ is an optimal solution for \mathcal{I} and thus $OPT(\mathcal{I}) = \sum_{i=1}^{m-n+1} OPT_i$. Moreover, the structure of cactus graphs guarantees that for any $i = 1, \dots, m - n + 1$, $X(f_i)$ is independent of the choice of queries that RANDOM_C makes when applied to C_j with $j \neq i$ as well as the queries' outcome. Let q_{prep} denote the cost of queries made in the preprocessing. We denote by ALG_i the cost of the queries RANDOM_C makes when applied to C_i , $i \in \{1, 2, \dots, m - n + 1\}$. As we expect at most $1.5 \cdot OPT_i$ queries when applying RANDOM_C to C_i , we obtain that

$$\begin{aligned} \frac{\mathbb{E}[\text{RANDOM}_C(\mathcal{I})]}{OPT(\mathcal{I})} &= \frac{q_{prep} + \sum_{i=1}^{m-n+1} \mathbb{E}[ALG_i(\mathcal{I})]}{OPT(\mathcal{I})} \\ &\leq \frac{q_{prep} + \sum_{i=1}^{m-n+1} 1.5 \cdot OPT_i}{q_{prep} + \sum_{i=1}^{m-n+1} OPT_i} \\ &\leq 1.5. \end{aligned}$$

This completes the proof of the competitiveness. Note that the correctness of RANDOM_C is immediate as for each cycle the algorithm proceeds to query edges until an edge becomes always maximal. The optimality follows from the fact that no randomized algorithm can achieve a competitive ratio less than 1.5 even on triangles, see [4]. A precise description of RANDOM_C for general query costs is given in Algorithm 1.

4 V-MST-U and randomization

We will now turn to the vertex uncertainty version of MST-U which has not been considered in the context of randomized algorithms so far. We start by showing that no randomized algorithm for V-MST-U can achieve a better competitive ratio than 2.5. Our proof is based on the proof of the bound for the deterministic performance guarantee by [5].

Theorem 6 *No randomized algorithm for V-MST-U can achieve a competitive ratio less than 2.5. This remains true even for cycles and under the assumption of uniform query costs.*

Proof We will prove the theorem by applying a variant of Yao's Principle (see [1]) which allows to derive a lower bound for the performance of randomized

<p>input : An instance of MST-U with cactus graph $G = (V, E)$, uncertainty sets A_e and query costs $q_e, e \in E$</p> <p>output: A feasible query set Q</p> <ol style="list-style-type: none"> 1 Draw b uniformly at random from $[0, 1]$; 2 Preprocess the instance such that $T_L = T_U$; 3 Index the edges f_1, \dots, f_{m-n+1} in $R := E \setminus T_L$ arbitrarily ; 4 Initialize $Q = \emptyset$; 5 for $i \leftarrow 1$ to $m - n + 1$ do 6 Add f_i to T_L and let C_i be the unique cycle closed; 7 Let $X(f_i)$ be the set of edges $g \in T_L \cap C_i$ with $U_g > L_{f_i}$; 8 Let $q_i := \sum_{e \in X(f_i)} q_e$; 9 if $X(f_i) \neq \emptyset$ then 10 if $b \leq \frac{q_{f_i}^2}{q_{f_i}^2 + q_i^2}$ then 11 add all edges in $X(f_i)$ to Q and query them. 12 else 13 Add f_i to Q and query f_i. 14 while <i>no edge in the cycle C_i is always maximal</i> do 15 Query an unqueried edge $e \in C_i \setminus Q$ with maximum U_e and add it to Q.
--

Algorithm 1: The algorithm RANDOM_C for MST-U with general query costs in cactus graphs

algorithms from the best possible expected performance of a deterministic algorithm against a finite randomized family of instances. A *finite randomized family of instances* is a pair (\mathcal{R}, p) where \mathcal{R} is a finite set of instances for V-MST-U and p is a probability vector of length $|\mathcal{R}|$ which indicates for each instance $R \in \mathcal{R}$ the probability of its occurrence.

In [3], Erlebach and Hoffmann define a graph G with four non-trivial vertices A, B, C and D . For each such vertex $v \in \{A, B, C, D\}$ they provide an instance R_v where vertex positions are such that $\{v\}$ is the only optimal query set while after having queried any combination $S \subset \{A, B, C, D\} \setminus \{v\}$ of the other vertices, vertex positions can still turn out to be as in R_u with $u \in \{A, B, C, D\} \setminus S$. For more details, see Figure 1.

Consider now the randomized family of instances (\mathcal{R}, p) , where $\mathcal{R} = \{R_A, R_B, R_C, R_D\}$ and $p(R_v) = \mathbb{P}[R = R_v] = 0.25$ for $v \in \{A, B, C, D\}$. Then no deterministic algorithm ALG achieves a better expected competitive ratio $\mathbb{E}_{R \sim p} \left(\frac{ALG(R)}{OPT(R)} \right)$ than the algorithm ALG_1 which queries A, B, C, D (or less if an MST can already be identified) in this order independently from the queries' results. This is due to the fact that querying v only reveals whether or not we are facing instance R_v but if not, it is indistinguishable which of the remaining instances it might be. Then by Yao's Principle, no randomized

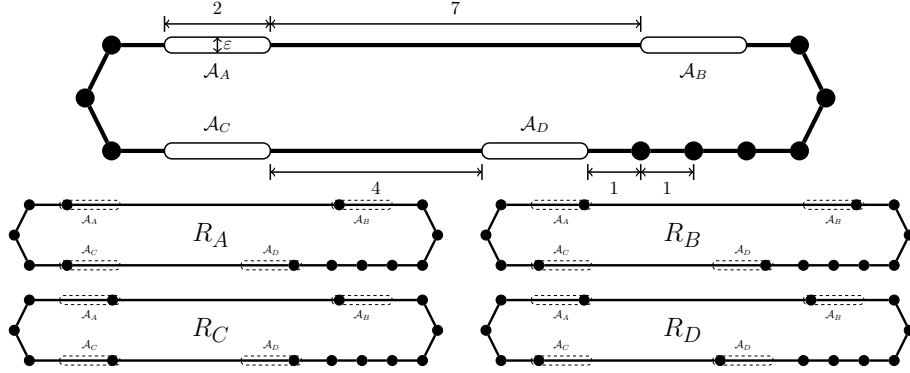


Figure 1: Instances R_A, R_B, R_C and R_D for the finite randomized family (\mathcal{R}, p) of instances

algorithm has a better performance than

$$\begin{aligned} \min_{ALG \in \mathcal{A}} \mathbb{E}_{R \sim p} \left(\frac{ALG(R)}{OPT(R)} \right) &= \mathbb{E}_{R \sim p} \left(\frac{ALG_1(R)}{OPT(R)} \right) = \\ &0.25 \cdot \left(ALG_1(R_A) + ALG_1(R_B) + ALG_1(R_C) + ALG_1(R_D) \right) = \\ &0.25 \cdot (1 + 2 + 3 + 4) = 2.5, \end{aligned}$$

where \mathcal{A} denotes the class of all deterministic algorithms.

4.1 A randomized algorithm for V-MST-U on cactus-like graphs

V-MST-U structurally differs from MST-U mainly due to the following two aspects:

- a) Querying a single end vertex of an edge already yields partial information about the edge's length.
- b) A vertex may be incident to several edges. Thus, knowing its precise position impacts on the possible lengths of all adjacent edges.

While for MST-U it is sometimes possible to identify edges which have to lie in any feasible query set, it becomes significantly harder to tell whether a vertex has to be queried due to aspect a). However, we will make use of the weaker statement that for certain edges at least one out of two end vertices needs to be queried.

Similarly to the special case of cactus graphs for MST-U, we will now consider instances of V-MST-U where no two cycles share a non-trivial vertex and query costs are uniform. This makes it easier to deal with aspect b), as vertex queries may only impact on the possible lengths of at most two adjacent edges. We will

refer to vertex-uncertainty graphs where cycles do not intersect in non-trivial vertices as *cactus-like*. Our algorithm considers cycles separately. For an instance \mathcal{I} with a cycle C , our algorithm works as follows: First, it computes the associated edge instance and preprocesses the instance s.t. $T_L = T_U$ holds. Then it identifies an edge f with largest upper limit U_f in C . Usually, our algorithm deterministically queries the end vertices of f and will then query the end vertices of edges in C in order of decreasing upper limit in the associated edge instance until a longest edge in C can be identified. Only if the neighbor set is small (where “small” is yet to be defined) we will make use of randomization. In case that the neighbor set is large and the algorithm performs deterministically, we will say that the algorithm follows the *deterministic procedure*.

Remark: To simplify notation, we will w.l.o.g. assume throughout this section that all uncertainty sets in associated edge instances have the form of intervals. However, the argumentation can be translated straightforwardly to the general case, e.g. by replacing A_f contains A_e by $L_f \leq L_e \leq U_e \leq U_f$ where one out of two consecutive inequalities is strict.

Preprocessing

If $T_L \neq T_U$ there is exactly one edge in $T_L \setminus T_U$ because C is a cycle. Our preprocessing consists in repeatedly computing T_L and T_U for the associated edge instance and querying the end vertices of the edge in $T_L \setminus T_U$ until $T_L = T_U$. However, the property $T_L = T_U$ does not necessarily remain true throughout the algorithm as vertex queries have impact on the uncertainty sets of all adjacent edges.

Deterministic procedure

We will now argue why the deterministic procedure makes at most $2 \cdot \text{OPT}(\mathcal{I}) + 2$ queries.

Lemma 7 *Let \mathcal{I} be an instance of V-MST-U on a cycle C . Let ALG be an algorithm which computes the associated edge instance and in each subsequent step queries the end vertices of an edge with largest upper limit in the associated edge instance until a longest edge is found. Then ALG makes at most $2 \cdot \text{OPT}(\mathcal{I}) + 2$ queries when applied to \mathcal{I} .*

Moreover, if \mathcal{I} is such that $T_L = T_U$, $f_1 \in C - T_L$ is the first edge with largest upper limit considered by ALG and Q^ is an optimal query set with $f_1 \cap Q^* \neq \emptyset$ then ALG makes at most $2.5 \cdot \text{OPT}(\mathcal{I})$ queries when applied to \mathcal{I} .*

Proof We will denote by $\{u_i, v_i\}$ the set of vertices that ALG queries during the i 'th step. (Note that $u_i = v_i$ is possible if the edge considered in the i 'th step has one trivial end vertex.) Moreover, we will denote by V_i the set of vertices queried during the first i steps such that $|V_T| = \text{ALG}(\mathcal{I})$ where T is the last iteration. Let Q^* be an optimal query set and denote by A_e^i , U_e^i and L_e^i the uncertainty set, upper limit and lower limit of an edge e in iteration i . We will

argue that for each iteration i it either holds that at least one vertex in $\{u_i, v_i\}$ lies in Q^* or that for each $j < i$ at least one vertex in $\{u_j, v_j\}$ lies in Q^* . This trivially holds for the first iteration. Assume now that it is known to be true for iterations $1, \dots, i-1$ and consider the i -th iteration. Note that u_i and v_i are non-trivial end vertices of an edge f which has largest upper limit U_f^i in the uncertainty graph considered in iteration i . First assume that there exists an edge $e \in C$ with $A_e^i \subset A_f^i$. Then by Lemma 4 we need to know the position of at least one of f 's end vertices in order to be able to identify a minimum spanning tree. Now assume that none of the other uncertainty sets is a subset of A_f^i . By induction hypothesis we know that there is at most one $j > i$ for which we have not yet argued that $\{u_j, v_j\}$ intersects with Q^* . Let h be the edge with largest upper limit in iteration j such that u_j and v_j are end vertices of h . Then w_h can neither lie in A_f^i else $A_h^i = \{w_h\} \subset A_f^i$ nor can we have that $U_h^i = w_h > U_f^i$ as f has largest upper limit in the i -th iteration. Hence we know that $w_h < w_f$ and thus h lies in any MST. Let g be a longest edge in C . By the choice of h we have that $U_h^j \geq U_g^j \geq w_g$ and thus without querying neither u_j nor v_j (which are the non-trivial end vertices of h in iteration j) we will not be able to verify an MST which does not contain g .

From this the first part of the lemma follows: Let j be the last step for which we cannot guarantee that v_j or u_j lies in any feasible query set. Then we know that half of V_{j-1} lies in Q^* and that for each step $i > j$ we know that v_i or u_i lies in Q^* . Hence, $OPT(\mathcal{I}) = |Q^*| \geq \frac{|V_{j-1}|}{2} + \frac{|V_T| - |V_j|}{2} = \frac{|V_T| - |\{u_j, v_j\}|}{2} \geq \frac{ALG(\mathcal{I}) - 2}{2}$. Note that if $j \neq 1$ then $u_j = v_j$: By the choice of j , no uncertainty set is contained in A_e^j where e is the edge with largest upper limit in iteration j and non-trivial end vertices u_j and v_j . Thus $w_{f_1} \leq L_e$ which cannot be the case if both end vertices of e remain unqueried until iteration j due to our preprocessing. Hence, the first iteration is the only iteration where we might query two vertices none of which is in Q^* and thus if $Q^* \cap f_1 \neq \emptyset$ then ALG makes at most $2 \cdot OPT(\mathcal{I}) + 1$ queries, i.e. it achieves competitive ratio 2.5, unless $OPT(\mathcal{I}) = 1$. However, if $OPT(\mathcal{I}) = 1$ and Q^* intersects f_1 then Q^* is a subset of f_1 and thus the deterministic procedure finishes after querying only the end vertices of f_1 . This proves the second part of the claim.

Randomization

The above lemma guarantees that the deterministic procedure achieves a competitive ratio of $2 + \frac{2}{OPT(\mathcal{I})}$ which is at most 2.5 unless $OPT(\mathcal{I}) \leq 3$. We will first handle the exception where $OPT(\mathcal{I}) = 3$ by making a slight adaption: The algorithm starts with the first three iterations of the deterministic procedure. In the fourth iteration it queries a vertex in $\{u_4, v_4\}$ with probability 0.5 (or probability 1 if $u_4 = v_4$), then queries the other vertex if necessary and finally proceeds with the deterministic procedure if still no longest edge can be identified. See Algorithm 2 for a precise description of the algorithm RAND3 that is applied to the uncertainty graph that results from querying f_1 's end vertices.

Lemma 8 *Let \mathcal{I} be a preprocessed instance of V-MST-U on a cycle C . Let ALG*

be an algorithm which computes the associated edge instance, queries the end vertices of an edge f_1 with largest upper limit in the associated edge instance and applies RAND3 to the resulting uncertainty graph. Then ALG achieves competitive ratio 2.5 when applied to \mathcal{I} unless $OPT(\mathcal{I}) \leq 2$ and $f_1 \cap Q^* = \emptyset$ for all optimal query sets Q^* .

Proof Again we denote by $\{u_i, v_i\}$ the set of vertices that ALG queries during the i 'th step (i.e. $\{u_1, v_1\} = f_1$). The algorithm ALG makes at most as many queries as the deterministic procedure. Thus if $Q^* \cap f_1 \neq \emptyset$ or $OPT(\mathcal{I}) \geq 4$ then the claim follows from Lemma 7. Assume now that $Q^* \cap f_1 = \emptyset$ and $OPT(\mathcal{I}) = 3$. If the algorithm finishes after the first three iterations then it has made at most 6 queries and yields a competitive ratio of at most 2. If the algorithm has not finished after the first three iterations, then we know from the proof of Lemma 7 that two of the three sets $\{u_1, v_1\}$, $\{u_2, v_2\}$ and $\{u_3, v_3\}$ intersect with Q^* such that $\{u_4, v_4\}$ contains the last vertex in Q^* . Hence, we expect to make $0.5 \cdot (7 + 8)$ queries which yields a competitive ratio of at most 2.5.

input : An instance \mathcal{I} of V-MST-U with uniform query costs with a cycle $C = (V, E)$ and uncertainty sets A_v for $v \in V$
output: A feasible query set Q

- 1 Compute the associated edge instance and T_L ;
- 2 Initialize $Q = \emptyset$;
- 3 **for** $i \leftarrow 1$ **to** 2 **do**
- 4 **if** no edge in C is always maximal **then**
- 5 Let g be an edge with largest upper limit in C ;
- 6 Query the end vertices of g and add them to Q
- 7 **if** no edge in C is always maximal **then**
- 8 Let g be an edge with largest upper limit in C and let u, v be the non-trivial end vertices of g ;
- 9 Pick $b \in [0, 1]$ uniformly at random;
- 10 **if** $b \leq 0.5$ **then**
- 11 Query u and add it to Q ;
- 12 **if** no edge in C is always maximal **then**
- 13 Query v and add it to Q
- 14 **else**
- 15 Query v and add it to Q ;
- 16 **if** no edge in C is always maximal **then**
- 17 Query u and add it to Q
- 18 **while** no edge in C is always maximal **do**
- 19 Let g be an edge with largest upper limit in C ;
- 20 Query the end vertices of g and add them to Q

Algorithm 2: The subroutine RAND3 of V-RANDOM $_C$

Following from Lemma 8, we only need to deal with instances where $OPT(\mathcal{I}) \leq$

2 and $f_1 \cap Q^* = \emptyset$. Thus we first check whether the (original) uncertainty graph contains candidates for a feasible query set W of size at most 2 that does not contain an end vertex of f_1 . We do so by computing the size a of a smallest vertex cover of the non-trivial edges in $X(f_1)$ which does not intersect f_1 and consists of non-trivial vertices only. Note that if there is a feasible query set of size at most 2 which does not intersect the edge f_1 then $a \leq 2$ by Lemma 4. If $a = 1$ we apply a randomized procedure RAND1 as described below. If $a = 2$, we query the end vertices of f_1 and depending on the outcome either decide to stick to the strategy RAND3 or apply a randomized strategy as described below.

input : An instance \mathcal{I} of V-MST-U with uniform query costs with a cycle $C = (V, E)$ and uncertainty sets $A_v, v \in V$

output: A feasible query set Q

- 1 Compute the associated edge instance;
- 2 Initialize $Q = \emptyset$;
- 3 Let f_1 be an edge with largest upper limit and let $X(f_1)$ be the set of edges $g \in T_L \cap C$ with $U_g > L_{f_1}$;
- 4 **if** no edge in C is always maximal **then**
- 5 index the vertices v_1, \dots, v_k in $\bigcap X(f_1) \cup f_1$ arbitrarily;
- 6 index the permutations $\sigma_1, \dots, \sigma_{k!}$ in \mathcal{S}_k arbitrarily;
- 7 draw b uniformly at random from $[0, 1]$ and pick permutation $\sigma = \sigma_i$ if $b \in [\frac{i-1}{k!}, \frac{i}{k!})$;
- 8 $j := 1$;
- 9 **while** no edge in C is always maximal and $j < k$ **do**
- 10 Query $v_{\sigma_i(j)}$ and add it to Q ;
- 11 $j := j + 1$;
- 12 **if** no edge in C is always maximal **then**
- 13 Query the remaining vertices in $X(f_1)$

Algorithm 3: The subroutine RAND1 of V-RANDOM_C

Case a = 1 : If $a = 1$ then $X(f_1)$ either consists of a single edge e or of two edges g and h which are incident to each other but not to f_1 (see Figure 2). The algorithm picks an order in which to query the three (or four) vertices in $(g \cap h) \cup f_1$ or $e \cup f_1$ uniformly at random and queries the vertices in the respective order up to the point where a longest edge in C can be identified. If necessary, it queries the remaining vertices in $X(f_1)$. See Algorithm 3 for a precise description.

Case a = 2 : The algorithm starts by querying the end vertices of f_1 . Let g be an edge in the resulting uncertainty graph with largest upper limit U_g . If now $w_{f_1} \leq L_g$ then we continue as in RAND3. If $w_{f_1} \in A_g$ then the algorithm picks a vertex cover W of $X(f_1)$ which contains $|W| = 2$ non-trivial vertices such that $W \cap f_1 = \emptyset$ uniformly at random. The algorithm queries the vertices in W but starts with a vertex in $W \cap g$. After that it queries the end vertices of edges in order of decreasing upper limit. For a precise description, see Algorithm 4.

input : An instance \mathcal{I} of V-MST-U with uniform query costs with a cycle $C = (V, E)$ and uncertainty sets $A_v, v \in V$

output: A feasible query set Q

- 1 Compute the associated edge instance and T_L ;
- 2 Initialize $Q = \emptyset$;
- 3 Let f_1 be the edge in $R := E \setminus T_L$ and let $X(f_1)$ be the set of edges $g \in T_L \cap C$ with $U_g > L_{f_1}$;
- 4 Query f_1 and add its end vertices to Q ;
- 5 Let g be an edge with largest upper limit in C ;
- 6 **if** no edge in C is always maximal **then**
- 7 **if** $w_{f_1} \leq L_g$ **then**
- 8 $Q = Q \cup \text{RAND3}(C_i, \{A_v | v \in V(C_i)\})$
- 9 **else**
- 10 Compute $\mathcal{W} = \{W \subset V(C) - f_1 | W \text{ is a smallest vertex cover of } X(f_1) \text{ which consists of non-trivial vertices only}\}$ and index the elements $W_1, \dots, W_{|\mathcal{W}|}$ arbitrarily;
- 11 draw b uniformly at random from $[0, 1]$ and pick a vertex cover $W = W_i$ if $b \in \left[\frac{i-1}{|\mathcal{W}|}, \frac{i}{|\mathcal{W}|}\right)$;
- 12 query a vertex v in $W \cap g$ and add it to Q ;
- 13 **if** no edge in C is always maximal **then**
- 14 query a vertex in $W - v$ and add it to Q
- 15 **while** no edge in C is always maximal **do**
- 16 Let h be an edge with largest upper limit in C ;
- 17 query the end vertices of h and add them to Q

Algorithm 4: The subroutine RAND2 of V-RANDOM $_C$

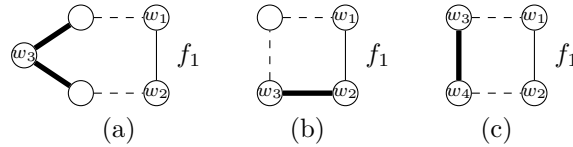


Figure 2: Sketches of cycles with $a = 1$ where edges in $X(f_1)$ are bold and dashed lines indicate parts of the cycle that are not in $X(f_1) \cup f_1$. In (a) and (b) RAND1 picks a vertex in $\{w_1, w_2, w_3\}$ with probability $\frac{1}{3}$ each. In (c) RAND1 picks a vertex in $\{w_1, w_2, w_3, w_4\}$ with probability $\frac{1}{4}$ each.

Lemma 9 *Let \mathcal{I} be a preprocessed instance of V-MST-U on a cycle C . Let f_1 be an edge with largest upper limit in the associated edge instance and denote by a the size of a smallest vertex cover of $X(f_1)$ which consists of non-trivial vertices only. If $a = 2$ then Algorithm 4 achieves competitive ratio 2.5 when applied to \mathcal{I} .*

Proof As $|f_1 \cup V(X(f_1))| \leq 8$ (see Figure 3) we only need to discuss instances

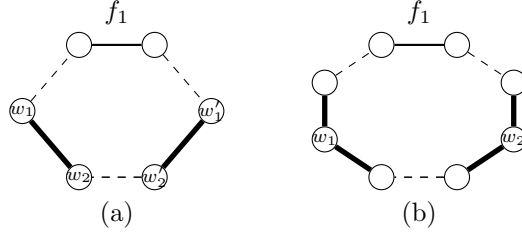


Figure 3: Sketches of cycles with $a = 2$ where edges in $X(f_1)$ are bold and dashed lines indicate parts of the cycle that are not in $X(f_1) \cup f_1$. In (a) RAND2 picks a cover $\{w_i, w'_j\}$, $i, j = 1, 2$ with probability $\frac{1}{4}$ each. In (b) RAND2 picks $\{w_1, w_2\}$ deterministically.

with $OPT(\mathcal{S}) = |Q^*| < 4$ where Q^* is an optimal query set. If $OPT(\mathcal{S}) = 1$ then $a = 2$ implies that $Q^* \subset f_1$ by Lemma 4 and the vertex in Q^* is queried within the first two queries. Say $OPT(\mathcal{S}) = 2$. Let g be an edge with largest upper limit in the associated edge instance after having queried the end vertices of f_1 . We will distinguish the cases where Q^* either intersects f_1 or not. First assume that $Q^* \cap f_1 \neq \emptyset$. If $Q^* = f_1$ then the claim trivially holds. We thus assume that $|Q^* \cap f_1| = 1$. If $w_{f_1} \leq L_g$ we proceed as in RAND3 and the claim follows from Lemma 8. Thus, consider the case where $w_{f_1} \in A_g$. RAND2 proceeds by querying a vertex cover of $X(f_1)$. We will now argue that if still no longest edge is found and h is an edge which subsequently has largest upper limit in C then the remaining vertex $v \in Q^*$ is the (unique) non-trivial end vertex of h . If Q^* verifies an MST which contains h then Q^* contains the end vertex of h otherwise h will remain a candidate for a strictly longest edge. Note that h is not incident to f_1 otherwise h would be trivial after querying f_1 and a vertex cover of $X(f_1)$. Thus after querying f_1 we have that $w_{f_1} \in A_h$ due to the preprocessing which proves that v needs to be an end vertex of h if Q^* verifies that h is a longest edge in C . Hence, if $OPT(\mathcal{S}) = 2$ and $Q^* \cap f_1 \neq \emptyset$ we can verify an MST after querying the end vertices of f_1 , a vertex cover of size 2 and an end vertex of h which proves a competitive ratio of at most 2.5.

Assume now that $Q^* \cap f_1 = \emptyset$. By Lemma 4, this implies that Q^* is a vertex cover of $X(f_1)$ (which verifies that f_1 is a longest edge in C). Say RAND2 picks a vertex cover which is not a feasible query set. During any of the following steps, let g be an edge which currently has largest weight in C . Then A_g contains w_{f_1} and thus Q^* contains the non-trivial vertex of g else g cannot be shown to have at most the weight of f_1 . Note that we can have at most four different smallest vertex covers of $X(f_1)$ of which do not intersect f_1 and at most two which are pairwise disjoint from each other (see Figure 3). Thus with probability at most $\frac{1}{4}$ we pick Q^* right away, with probability $\frac{1}{4}$ we pick a vertex cover that is disjoint from Q^* and need two additional queries while with probability $\frac{1}{2}$, we pick a vertex cover that intersects with Q^* in a single vertex after which we need to make one more query. This leads to a competitive ratio

of $\frac{0.25 \cdot 4 + 0.25 \cdot 6 + 0.5 \cdot 5}{2} = 2.5$.

To complete the proof, we need to discuss instances with $OPT(\mathcal{I}) = 3$. Note that the claim trivially holds if we have that $|f_1 \cup V(X(f_1))| > 7$ which is the case unless $X(f_1)$ consists of two paths which are neither incident to each other nor to f_1 (as in Figure 3 (b)). Assume that the latter is the case and that we have queried the end vertices of f_1 and the vertices in the vertex cover W of $X(f_1)$. Thus, all remaining vertices in $X(f_1)$ have at most one non-trivial end vertex. As $V(C) - W - f_1$ contains no such vertex cover of $X(f_1)$ of size 3, Lemma 4 guarantees that we have already queried one of the vertices in Q^* before we consider the edge h . Analogously to the proof of Lemma 7 for the deterministic procedure, we can argue that at least two out of the next three queries contain vertices in Q^* and thus we make at most seven queries in total which proves the claim for this case.

4.1.1 The algorithm V-Random_C

A precise description of the algorithm V-RANDOM_C is displayed in Algorithm 5

	<p>input : An instance \mathcal{I} of V-MST-U with uniform query costs, a graph $G = (V, E)$ and uncertainty sets $A_v, v \in V$ such that no two cycles share a non-trivial vertex</p> <p>output: A feasible query set Q</p> <ol style="list-style-type: none"> 1 Compute the associated edge instance; 2 Initialize $Q = \emptyset$; 3 Preprocess the instance such that $T_L = T_U$; 4 Index the edges f_1, \dots, f_{m-n+1} in $R := E \setminus T_L$ arbitrarily ; 5 for $i \leftarrow 1$ to $m - n + 1$ do <li style="padding-left: 1em;">6 Let C_i be the unique cycle in $T_L + f_i$; <li style="padding-left: 1em;">7 Let $X(f_i)$ be the set of edges $g \in T_L \cap C_i$ with $U_g > L_{f_i}$; <li style="padding-left: 1em;">8 Compute the size a of a smallest vertex cover of $X(f_1)$ which does not intersect f_1 and consists of non-trivial vertices only; <li style="padding-left: 1em;">9 if <i>no edge in C_i is always maximal</i> then <li style="padding-left: 2em;">10 if $a = 1$ then <li style="padding-left: 3em;">11 $Q = Q \cup \text{RAND1}(C_i, \{A_v v \in V(C_i)\})$ <li style="padding-left: 2em;">12 else <li style="padding-left: 3em;">13 Query the end vertices of f_i and add them to Q; <li style="padding-left: 3em;">14 Let g be an edge with largest upper limit in C_i; <li style="padding-left: 3em;">15 if $a \geq 3$ or $w_{f_i} \leq L_g$ then <li style="padding-left: 4em;">16 $Q = Q \cup \text{RAND3}(C_i, \{A_v v \in V(C_i)\})$ <li style="padding-left: 3em;">17 else <li style="padding-left: 4em;">18 $Q = Q \cup \text{RAND2}(C_i, \{A_v v \in V(C_i)\})$
--	---

Algorithm 5: The algorithm V-RANDOM_C for V-MST-U with uniform query costs in cactus-like graphs

Theorem 10 *For instances of V-MST-U with uniform query cost where no two cycles intersect in a non-trivial vertex the algorithm V-RANDOM_C achieves a competitive ratio of 2.5 and this is best possible.*

Proof Let \mathcal{I} be an instance with uncertainty graph G to which we apply V-RANDOM_C. For a cycle C in G we denote by Q_C^* an optimal query set for an instance of V-MST-U with graph C and uncertainty sets and weights as in \mathcal{I} . If we prove for each such cycle C that V-RANDOM_C achieves a competitive ratio of 2.5 when applied to C the claim follows because $Q^* = \bigcup_C Q_C^*$ is an optimal query set for \mathcal{I} . Moreover, Lemma 2 guarantees that each vertex pair queried during the preprocessing intersects with Q^* . Thus we assume that \mathcal{I} is a preprocessed instance of V-MST-U with uncertainty graph C where C is a cycle. Let Q^* be an optimal query set for \mathcal{I} . Now we prove that V-RANDOM_C queries at most $2.5 \cdot OPT(\mathcal{I})$ vertices when applied to \mathcal{I} .

First assume that $a \geq 3$, i.e. the algorithm queries the end vertices of f_1 and applies RAND3 to the resulting uncertainty graph. From $a \geq 3$ it follows by Lemma 4 that $Q^* \cap f_1 \neq \emptyset$ or $OPT(\mathcal{I}) \geq a \geq 4$ and thus the claim follows from Lemma 8.

if we assume that $a = 2$ then the claim follows from Lemma 9. Finally, assume that $a = 1$. Note that the maximum number of vertices in $f_1 \cup V(X(f_1))$ is 5. Thus we make at most five queries and only need to consider the case where $OPT(\mathcal{I}) = 1$. The neighbor set $X(f_1)$ consists either of a single edge e or of two intersecting edges g and h . In the first case, each non-trivial vertex in $e \cup f_1$ could be a feasible query set of size one and we find the vertex in Q^* within at most $\frac{1}{4}(1 + 2 + 3 + 4) = 2.5$ queries in expectation. In the second case each non-trivial vertex in $(g \cap h) \cup f_1$ could be a feasible query set of size one and we find the vertex in Q^* within $\frac{1}{3}(1 + 2 + 3) = 2$ queries in expectation. Note that if $a = 0$ then f_i has no neighbors and is thus always maximal right away. The optimality follows from Theorem 6 and the fact that all uncertainty graphs used throughout its proof are cycles.

5 Minimum Spanning Stars

Several online algorithms exist for MST-U and are shown to deliver a solution the cost of which is bounded by a constant multiple of the cost of an optimal solution. But what happens if we ask to find a minimum spanning tree of a specified type at minimum query cost? For many types of special spanning trees, it is already NP-hard to find such a tree in a graph, e.g. Hamiltonian paths or spiders (see Gargano et al. [8]). In the following we will thus consider spanning stars which are generally easy to find.

Definition The complete bipartite graph $K_{1,k}$ is called a *star* and is denoted by S_k . We will refer to the vertex with degree k as centre of S_k . If a graph G with n vertices contains S_{n-1} as a subgraph, then S_{n-1} is said to be a spanning star in G .

We define the *Minimum Spanning Star Problem under Explorable Uncertainty* (MSS-U) analogously to MST-U, i.e. we want to identify a minimum spanning star by making queries of as little weight as possible. However, it turns out that no algorithm for MSS-U can achieve constant competitive ratio.

Theorem 11 *There exists no algorithm for MSS-U which achieves constant competitive ratio.*

Proof We prove the theorem by defining for each $n \in \mathbb{N}$ a finite family $(\mathcal{I}_n^k)_k$ of instances with uniform query cost on a graph G with n vertices such that every algorithm needs to make at least $(n - 2)$ times as many queries as necessary on at least one instance in $(\mathcal{I}_n^k)_k$. For $n = 5$ the construction is illustrated in Figure 4.

Consider a graph $G = (V, E)$ with $V = \{v_1, \dots, v_n\}$ and $E = \{\{v_1, v_i\} | i = 2, \dots, n\} \cup \{\{v_n, v_i\} | i = 1, \dots, n - 1\}$. Then G has precisely two spanning stars, one with centre v_1 and the other one with centre v_n . (All other vertices have degree 2.) Let $q_e = 1$ for all $e \in E$. We define the uncertainty sets as follows:

- $A_{\{v_1, v_n\}} = \{1\}$,
- $A_{\{v_1, v_i\}} = (0, 1)$ for $i = 2, \dots, n - 1$ and
- $A_{\{v_n, v_i\}} = (0, n)$ for $i = 2, \dots, n - 1$.

For $j \in \{2, \dots, n - 1\}$, \mathcal{I}_n^j is defined such that $w_{\{v_j, v_n\}} = n - 1$ and $w_e = 0.5$ for all $e \in E \setminus \{\{v_j, v_n\}, \{v_1, v_n\}\}$. It is then sufficient to query only the edge $\{v_j, v_n\}$ to know that the spanning star with center v_n has larger weight. Conversely, without querying $\{v_j, v_n\}$, it is impossible to tell which spanning star has minimum weight. Note that for an algorithm all non-trivial edges adjacent to v_n are undistinguishable. It has thus to decide on an order in which these edges are queried. If j is such that $\{v_j, v_n\}$ is the last edge adjacent to v_n to be queried by an algorithm then the algorithm competitive ratio at least $n - 2$ when applied to instance \mathcal{I}_n^j .

6 Conclusion

In this paper we considered the Minimum Spanning Tree Problem Under Explorable Uncertainty (MST-U) and the related Minimum Spanning Tree Problem Under Explorable Vertex Uncertainty (V-MST-U) for specified instance types where cycles can be considered independently from each other. We provided a randomized algorithm for MST-U on cactus graphs and proved that it achieves a competitive ratio of 1.5 which is best possible. For V-MST-U instances where cycles do not share non-trivial vertices and query costs are uniform, we provided the algorithm V-RANDOM_C which achieves a competitive ratio of 2.5. We showed that 2.5 is a lower bound for the performance of randomized algorithms for V-MST-U which remains true even for instances with

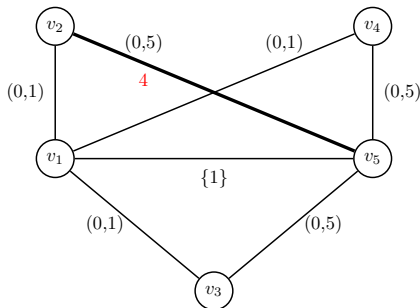


Figure 4: The weight of $\{v_2, v_5\}$ is 4. All missing edge weights equal 0.5. An optimal solution only needs to query $\{v_2, v_5\}$ which has larger weight than the star with center v_1 can have. A deterministic algorithm can not distinguish between the edges $\{v_2, v_5\}, \{v_3, v_5\}$ and $\{v_4, v_5\}$ and might have to query all three of them.

uniform query cost where the uncertainty graph is a cycle. Thus, the performance guarantee shown for $V\text{-RANDOM}_C$ is best possible. Finally, we introduced the Minimum Spanning Star Problem under Explorable Uncertainty (MSS-U) and proved that no algorithm for MSS-U can achieve constant competitive ratio. (V-)MST-U in itself is a problem which still deserves further investigation. A major open question in the setting of MST-U is whether there exists a randomized algorithm with a competitive ratio of 1.5 for general instances as well or whether the lower bound of 1.5 can be improved for instances where the graph is not necessarily a cactus.

As for V-MST-U, no randomized algorithm for general V-MST-U instances exists so far and even for the special case where cycles intersect in trivial vertices only, the performance guarantee of $V\text{-RANDOM}_C$ relies on the fact that query costs are uniform. Note that no deterministic algorithm with constant competitive ratio for V-MST-U with non-uniform query costs is known either.

Moreover, different models of uncertainty exploration could be subject of further research. Consider for instance a scenario where installing a camera in a certain location allows to measure the distance between itself and all surrounding objects. A setting like this could motivate a hybrid model between edge and vertex uncertainty where edge weights are uncertain but known to lie inside given uncertainty sets and can be revealed upon querying an adjacent vertex.

References

- [1] A. Borodin and R. El-Yaniv. *Online Computation and Competitive Analysis*. Cambridge University Press, 1998.
- [2] C. Dürr, T. Erlebach, N. Megow, and J. Meißner. Scheduling with Explorable Uncertainty. In *9th Innovations in Theoretical Computer Science Conference*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 30:1–30:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.
- [3] T. Erlebach and M. Hoffmann. Minimum spanning tree verification under uncertainty. In *Proceedings of the International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 164–175, 2014.
- [4] T. Erlebach and M. Hoffmann. Query-competitive algorithms for computing with uncertainty. *Bulletin of the European Association for Theoretical Computer Science*, 116, 2015.
- [5] T. Erlebach, M. Hoffmann, D. Krizanc, M. Mihalák, and R. Raman. Computing minimum spanning trees with uncertainty. *Proceedings of Symposium on Theoretical Aspects of Computer Science*, pages 277–288, 2008.
- [6] T. Feder, R. Motwani, L. O’Callaghan, C. Olston, and R. Panigrahy. Computing shortest paths with uncertainty. *Journal of Algorithms*, 62(1):1–18, 2007.
- [7] J. Focke, N. Megow, and J. Meißner. Minimum Spanning Tree under Explorable Uncertainty in Theory and Experiments. In *16th International Symposium on Experimental Algorithms*, volume 75 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.
- [8] L. Gargano, M. Hammar, P. Hell, L. Stacho, and U. Vaccaro. Spanning spiders and light-splitting switches. *Discrete Mathematics*, 285:83–95, 2004.
- [9] M. Goerigk, M. Gupta, J. Ide, A. Schöbel, and S. Sen. The robust knapsack problem with queries. *Computers and Operations Research*, 55:12–22, 2015.
- [10] S. Kahan. A model for data in motion. *23rd Annual ACM Symposium on Theory of Computing (STOC’91)*, pages 267–277, 1991.
- [11] N. Megow, J. Meißner, and M. Skutella. Randomization helps computing a minimum spanning tree under uncertainty. *SIAM Journal on Computing*, 46(4):1217–1240, 2017.