AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **74(1)** (2019), Pages 33–45

Contractors' minimum spanning tree

Irith Ben-Arroyo Hartman

Caesarea Rothschild Institute
University of Haifa
Israel
irith.hartman@gmail.com

Ilan Newman

Computer Science Department
University of Haifa
Israel
ilan@cs.haifa.ac.il

Abstract

We consider the following problem. Given a graph $G = (V, E)$, a partition of E into k color classes, $E = \cup_{i=1}^{k} E_i$, and a cost function for each class $f_i : 2^{E_i} \mapsto \mathbb{R}^+$, find a spanning tree $T = (V, F)$ whose total cost is minimal, where the cost of T is defined as the sum of the costs of the color classes in T, namely $\sum_i f_i(F \cap E_i)$. We show that the general problem is NP-hard, even when the cost functions depend only on the number of edges and are discrete and concave. We also provide a characterization of when a tree, with a prescribed number of edges from each color class, exists, as well as an efficient algorithm for finding such a tree. Finally, we prove that the polytope of feasible solutions for cardinality cost functions values is integral.

1 Introduction

The *minimum spanning tree problem* is a very well known problem. Given a weighted graph $G = (V, E)$ with weight function $l : E \to R$, find a spanning tree of G of minimum weight. This problem has an efficient algorithm; see Borůvka [2], Jarník [6], Prim [11], and Kruskal [8]. Prasanna [10] proposed a variant of this problem which is very likely to occur in real life situations: suppose the graph is a map of a city, and each road (edge of the graph) belongs to some contractor. The mayor of the city has to pick a set of roads to be paved which span the graph and is of minimum cost; however, the cost of the roads is not directly proportional to the length of the

ISSN: 2202-3518

roads. Each contractor has a cost function which might allow discounts depending on the total lengths / number of the roads that she paves. In typical situations, the discount function will be non-decreasing and concave, but we need not make this assumption at this point.

More formally, the problem is as follows:

Problem 1 (General Contractors' MST) Let $G = (V, E)$ be a connected (not necessarily simple) graph, possibly with parallel edges. Let *color* : $E \mapsto \{1, \ldots, k\}$ be a partition of the edge set into k color classes. For each color class E_i there is a *cost function* defined by a function $f_i : 2^{E_i} \mapsto \mathbb{R}^+$, $i = 1, \ldots, k$. The cost of a forest is the sum of costs of all the color classes in the forest, i.e. $\sum_i f_i(F \cap E_i)$. The problem is to find a spanning tree T in G of minimum total cost.

In the problem above, the set of edges of color i, denoted by E_i, represents the edges which belong to contractor i.

Clearly, the algorithmic complexity of the general contractor's MST may depend on the cost functions, and the way they are represented as an input to the problem. We denote $|V(G)| = n$, and assume that $|E(G)| = poly(n)$ (which may not be true in general, since the graph is not simple). Obviously, in the case that each f_i is linear in the weights of the individual edges, this problem is just the well-know minimum spanning tree problem.

For non-linear functions, describing each f_i requires, in general, a description of exponential size in n (that is, its value on each forest). In that case, the problem is much less interesting, and we will not consider such functions. However, we will show that for general collection $f_i, i = 1, \ldots, k$ of polynomial description, even the problem of *one* contractor is in general non-tractable. Hence, we mostly restrict ourselves to cost functions f_i that depend only on the cardinality of the input set (for each i this could be quite different). In this case, the range of f_i is bounded to a set of cardinality at most n, and hence the total description of the cost function is polynomial in n and the length of the input. In such cases we assume that $f_i, i = 1, \ldots, k$ is explicitly given. We will call such function $f : 2^E \mapsto \mathbb{R}$, of which f is constant on all sets of the same cardinality, a *cardinality function*. A cardinality function f is naturally associated with a function $\tilde{f} : \{0, 1, \ldots, n-1\} \mapsto \mathbb{R}$, namely for which $f(E) = \tilde{f}(|E|)$. Abusing notation, we identify f with $\tilde{f}$ and refer to both as f when there is no risk of confusion.

One type of cardinality functions that will be of interest is what we call a *discrete concave*, or *discount* function, defined below. Other functions of interest may be sub-additive, sub-modular (semi) convex / concave, etc.

Definition 1 A function $f : \{0, 1, \ldots, n-1\} \mapsto \mathbb{R}$ is *discount (or discrete concave)*, if $g(i) = f(i) - f(i-1)$ is non-increasing with i.

A function $f : 2^E \mapsto \mathbb{R}$ is *submodular* if for every two subsets $A, B \subseteq E$,

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B).$$

Discount functions are interesting as they represent the natural situation when a contractor with a discount function, gives a reduced cost per additional edge, as one buys more edges. Submodular functions are similar in the above sense, but without the restriction of being constant on all subsets of the same cardinality.

Assume now that a given partition $\bar{p} = (n_1, \ldots, n_k)$ of $n-1$ gives the minimum of $\sum_i^k f_i(x_i)$. Does there exist a spanning tree F satisfying this partition, i.e does there exist a spanning tree in which the ith contractor is allocated n_i edges, $i = 1, \ldots, k$? Is there an efficient algorithm for constructing such a tree? This is formalised in the next definitions.

Definition 2 (Profile of a Forest) Given a graph $G = (V, E)$, a partition *color* : $E \mapsto \{1, \ldots, k\}$ of the edge set into k color classes $E_i = \{e \mid color(e) = i\}$, $i = 1, \ldots, k$, and some spanning forest F of G. The *profile* of F, denoted by $p(F)$, is a sequence of k non-negative integers $\bar{p} = (n_1, \ldots, n_k)$ such that $|F \cap E_i| = n_i$, $i = 1, \ldots, k$.

A sequence $\bar{p} = (n_1, \ldots, n_k)$ is called a *feasible profile*, or *profile* for short, if there exists a forest F such that $p(F) = \bar{p}$. Note that if $\bar{p}$ is feasible for a tree (rather than just a forest) then $\bar{p}$ is a partition of $n-1$ (i.e. $\sum_i n_i = n-1$).

The following is a statement of the existence problem.

Problem 2 (Contractors' Forest Feasibility Problem) Let $G = (V, E)$ be a connected (not necessarily simple) graph, possibly with parallel edges. Let *color* : $E \mapsto \{1, \ldots, k\}$ be a partition of the edge set into k color classes (corresponding to k contractors), and let E_i denote the set of edges of color i. Let $\bar{p} = (n_1, \ldots, n_k)$ be a sequence of natural numbers. Does there exist a spanning forest F of G with profile $p(F) = \bar{p}$?

Our results: We show that the most general contractors' MST problem, Problem 1, is NP-hard even for one contractor, when the cost function is given by a polynomial (in n) oracle. This is simply because the minimum spanning tree problem is NP-hard for general non-linear functions. We also show that the version with parallel edges is polynomially reducible to that on simple graphs. We then show that even in the case that each f_i is a discount function, Problem 1 remains NP-hard (but this is true when k is relatively large).

For the positive results, we show that the feasibility problem, Problem 2, can be solved in polynomial time for any k. In particular, this implies that for any constant k (that is independent of n), the minimum cost contractors' MST can be solved in polynomial time, by a simple search over all possible feasible trees. We further give an efficient characterization of all feasible solutions, and show that their convex hull is an integral polytope. We observe that this polytope admits a polytime separation oracle when $k = O(\log n)$.

The outline of this paper is as follows: We prove in Section 2 that Problem 1 is NP-hard both for simple graphs and for graphs with parallel edges. In Section 3, we discuss the case of cardinality functions. In this case, the cost of a tree depends only on the number of edges in each color class. We solve Problem 2, characterizing whether there exists a tree with a prescribed profile.In Section 3.3 we relate Problem 2 to the matroid intersection theorem, implying an efficient algorithm for any k. In Section 4 we characterize the set of all possible feasible profiles by their corresponding polytope. We prove that the polytope of feasible solutions for cardinality cost functions values is integral.

Related work: The most relevant related work is in [5]. This work considers the min-cost contractor MST, and other problems in which the solution is 'split' between several 'agents' (called there 'multi agent' problems). They consider the case where the cost functions are submodular and show a hardness result for approximating the min-cost-contractor spanning tree, in this general case.

2 Hardness for general enough functions

We begin by noting that even for one contractor, and $f : 2^E \mapsto \{0, 1\}$ given by a polytime oracle, the contractor MST is NP-hard. To see this we reduce the decision whether G contains a Hamilton path to this problem. Indeed, let G be a graph on which one wishes to decide whether there is a Hamilton path. Let $f : 2^{E(G)} \mapsto \{0, 1\}$ be defined by: $f(E') = 0$ if E' spans a connected subgraph of G with degree bounded by 2, and $f(E') = 1$ otherwise. Obviously $f(E')$ can be computed in time $O(|V(G)| + |E(G)|)$ for every subset E'. Further $f(E') = 0$ if an only if E' is a Hamilton path or cycle in G. In particular, G has a Hamilton path if and only if the contractor's MST value is 0, with respect to the function f.

We note that the function showing the hardness above, is clearly not a cardinality function. Neither is it monotone, additive, submodular or discount.

We next show that even if we restrict ourselves to discount functions then the problem is NP-hard.

Theorem 3 *The contractors' MST problem is NP-hard for monotone discount functions. The same is true for submodular functions.*

Proof. We do a reduction from set cover. The set-cover decision problem has an input $S_1, \ldots, S_m \subseteq [n]$, a collection of subsets of $[n]$, and in addition an integer r. The decision to be made is whether there is a sub-collection of at most r subsets that cover $[n]$. Namely, whether there is $I \subseteq [m]$ with $|I| \leq r$, such that $\cup_{i \in I} S_i = [n]$.

The problem is one of Karp's fundamental NP-Complete problems [7].

Let $\{S_1, \ldots, S_m\}$, r be an instance to the set-cover problem. We reduce it to the minimum contractors' MST with $k = m + 1$ and $G = (V, E)$ as follows: let $V = \{v_1, \ldots, v_n\} \cup \{u_1, \ldots, u_n\} \cup \{a\}$. For each $S_i, i = 1, \ldots, m$ we set $E_i = \{(u_j, v_j) | j \in S_i\}$. $E_i, i = 1, \ldots, m$, are the sets of edges corresponding to the contractors $1, \ldots, m$. Note that G may have parallel edges (of different contractors). Further we add $E_{m+1} = \{(a, u_i)| \ i = 1, \ldots, n\}$ as the set of edges of the $m + 1$ contractor.

The costs are defined as follows: For the $m+1$ contractor, $f_{m+1} \equiv 0$. Namely, for every subset it is allocated, it is paid 0. For every other contractor, the discount function is the same and defined as: $f(0) = 0$, and $f(i) = 1$ for $i \geq 1$. Namely, it charges a unit cost no matter how many edges it is allocated, provided it is allocated at least one edge.

Note that in any tree, the $m+1$ contractor will be allocated with the edge set of E_{m+1} to ensure connectivity. In addition, the cost functions are cardinality discount functions. Finally note that a cost of a tree is just the number of contractors among $\{1, \ldots, m\}$ that have an edge in that tree.

It is easy to see that if there is a solution for the set cover with r sets that cover $[n]$, and a corresponding index set I, $|I| = r$, then $E_{m+1} \cup (\cup_{i \in I} E_i)$ spans a connected subgraph of G of cost r and hence any tree in this subgraph has cost at most r. On the other hand, if there is a contractor MST of cost r, then the corresponding r subsets cover $[n]$.

Finally, the contractor cost functions are not only discount, but also monotone non-decreasing (as cardinality functions) and monotone submodular as functions of subsets of edges. ■

The reduction above heavily uses the fact that G may have parallel edges, arising from the fact that the sets in the set-cover problem are not necessarily disjoint. Indeed the set cover problem is trivial for disjoint sets. We next show that the contractor' MST problem on simple graph is no easier then the general one, for any set of cost functions.

Theorem 4 *The contractors' MST for k contractors on a graph G can be reduced in polytime to the $k+1$ contractors' MST problem on a simple graph G'.*

Proof. Let $G = (V, E)$, $E = \cup_1^k E_i$ be an instance for the contractors' MST with k contractors and functions f_i, $i = 1, \ldots, k$.

The graph G' is defined by subdividing each edge of G, $e = (u, v) \in E_i$ by inserting a new vertex x_e into it. In particular, G' is simple. The two edges resulting from e are split arbitrary between two contractors: the ith contractor if $e \in E_i$ and the other goes to the $k + 1^{st}$ contractor.

Note that the set E'_i of edges belonging to the ith contractor in G' corresponds via a $1 - 1$ mapping to the set E_i of the ith contractor in G. In addition, the $k+1$ contractor represents every edge in G. Hence we may identify every set of edges $A \subseteq \cup_1^k E'_i$ with the corresponding set in G which we also call A.

As for the costs, we define $f'_i(A \cap E_i) = f(A \cap E_i)$ for any set of edges $A \subseteq \cup_1^k E'_i$ via the mapping above. The cost for the $k+1$ contractor is identically 0.

It is obvious that if $T \subseteq \cup_1^k E_i$ is a spanning tree for G of cost α, then $T \cup E_{k+1}$ defines a connected spanning subgraph of G' of the same cost. Hence every MST in it has cost at most α. On the other hand, if $A \cup B$ is a contractor MST in G' with cost α, and $A \subseteq \cup_1^k E'_i$, while $B \subseteq E'_{k+1}$, then A defines a connected subgraph of G of cost α. ■

3 Solution to Problem 2

Recall that in this problem we are given a partition of the edge set into k contractors, as well as a sequence $\bar{p} = (n_1, \ldots, n_k)$. Does there exist a contractor spanning forest / tree F of G with $p(F) = \bar{p}$?

We give below a necessary and sufficient condition for the existence of a spanning forest / tree satisfying this partition, and an efficient algorithm for constructing such a forest, if it exists. We note that an implicit, and different characterization is due to the fact that this decision problem is in polytime, as it is an instance of the matroid intersection problem. This will be detailed in what follows.

3.1 Notation

Let $G = (V, E)$ be an undirected not necessarily simple graph and let $color : E \mapsto [k]$ be a partition of E into k classes called *colors*. (Here $[k] := \{1, 2, \ldots, k\}$). We denote by E_i the set of all edges of color i. We also refer to members of E_i as i-edges (these that belong to the ith contractor in the contractors' MST problem).

For any subset of edges $F \subseteq E$, we denote by (F) =$\mathrm{rank}_G(F)$ the size of the largest forest spanned by the subgraph (V, F), and $\mathrm{rank}(G)$ =$\mathrm{rank}(E(G))$. A set of edges is *independent* if it contains no cycle. We denote by $G \setminus F = (V, E \setminus F)$ the subgraph obtained by deleting F, and by G/F the graph obtained by contracting F.

3.2 Characterization of possible profiles

Given graph G, a coloring $color : E \mapsto [k]$ of its edges. The decision if a sequence $\bar{p} = (n_1, \ldots, n_k)$ is feasible, (namely, there is a forest F of G such that $F \cap E_i = n_i,\ i = 1, \ldots, k$) is decidable by the Matroid-Intersection algorithm, as every feasible profile is an independent set in the intersection of two matroids. This is described in what follows. In particular, this gives an algorithm to decide if there is an MST of a given profile.

For the next subsection where we describe this, we assume basic knowledge with matroids and matroid intersection. However, we deduce a stand alone characterization of all possible profiles, and present a stand alone proof of it, as well as a proof by the matroid intersection theorem. The stand alone proof is somewhat more instructive. In particular, we deduce from it a theorem about the convex hull of all feasible profiles. This in turn, will imply polytime algorithm for the contractors' MST when the functions are discount but k is relatively small.

Our main result is the following theorem.

Theorem 5 *Let G be a colored graph as above, and $\bar{p} = (n_1, \ldots, n_k)$. Then $\bar{p}$ is a feasible profile if and only if for every $S \subseteq [k]$,*

$$\textstyle\sum_{i \in S} n_i \leq \mathit{rank}(\cup_{i \in S} E_i). \tag{1}$$

In addition $\bar{p}$ is a feasible profile of a tree *if and only if it is feasible and $\sum_i n_i = n-1$.*

Note that the "only if" direction is clear. Namely, if there is a forest F with profile $\bar{p}$, then $|F \cap (\cup_{i\in S} E_i)|$ cannot possibly have cardinality greater than $\text{rank}(\cup_{i\in S} E_i)$. Moreover, the condition for trees immediately follows from that of a forest.

3.3 Feasible profiles and matroid intersection

We will use here some basic matroid theory. We refer to [12] for a general reference on the relevant parts of this theory. Let $M_1 = (E, \mathcal{I}_1),\ M_2 = (E, \mathcal{I}_2)$ be two matroids on the same ground sets where $\mathcal{I}_1, \mathcal{I}_2$ are the corresponding families of independent sets. Let r_1, r_2 be the rank functions of M_1, M_2 respectively.

The matroid intersection theorem [3] states that for two such matroids the maximum size set in $\mathcal{I}_1 \cap \mathcal{I}_2$ is of size

$$\min_{U\subseteq E}\{r_1(U) + r_2(E \setminus U)\}. \tag{2}$$

To see how the profile feasibility problem is cast in this framework, let $G = (V, E)$ be a graph, and *color* : $E \mapsto [k]$ be a coloring of the edges by k colors. Let $\bar{p} = (n_1, \ldots, n_k)$ be a sequence of natural numbers. Let $M_1 = (E, \mathcal{F})$ be the graphic matroid, namely, $\mathcal{F}$ is the set of forests of G, each viewed as an edge-set. Thus $r_1(E)$ =rank(E) as defined in Section 3.1. Let $M_2 = (E, \mathcal{I})$ be the matroid in which $A \in \mathcal{I}$ if and only if $|A \cap E_i| \le n_i$ for all $1 \le i \le k$. It is known, and easy to verify, that M_2 is a matroid, also known as the *partition matroid*. Now, a set F is in $\mathcal{F} \cap \mathcal{I}$, namely in both matroids if and only if F is a forest, and $|F \cap E_i| \le n_i, i = 1, \ldots, k$. Hence, $\bar{p}$ is feasible if and only if there is a maximal set in $\mathcal{F} \cap \mathcal{I}$ which corresponds to an edge set of a forest whose profile is $\bar{p}$.

Since there is a polytime algorithm to find a maximum cardinality set in $\mathcal{F} \cap \mathcal{I}$ (for any two matroids), [9], this gives an efficient algorithm for the profile feasibility problem. Moreover, Equation (2) provides a CO-NP condition for a profile to be feasible. Note, however, that Equation (2) by itself is not an efficient characterisation as it is exponential in $|E|$. Theorem 5 provides a better characterization, since when $k = O(\log n)$ it is polynomially verifiable.

We now give the first proof of Theorem 5 based on the matroid intersection theorem.

Proof. [**of Theorem 5**] As necessity of the condition in the theorem has already been discussed, we assume that the condition in Equation (2) is met for $\bar{p} = (n_1, \ldots, n_k)$, and show that the matroid intersection theorem implies that the maximum set in $\mathcal{F} \cap \mathcal{I}$ is of size $\sum_{i=1}^{k} n_i$.

In other words, we need to verify that for every $U \subseteq E$, $r_1(U) + r_2(E - U) \ge \sum_{i=1}^{k} n_i$. Here $r_1(U)$ =rank(U) is $n - c$ if U induces c connected components, namely, it is the rank of the forest matroid. Then by Equation (2) the matroid-intersection theorem will imply the existence of a forest of size $\sum_1^k n_i$, but such a forest must have profile $\bar{p}$ by the constrains of M_2.

A set $U \subseteq E$ is called *closed* with respect to a matroid M if for any $x \in E \setminus U,\ r_M(U \cup \{x\}) = r_M(U) + 1$, where r_M is the rank function of M.

Claim 3.1 *The minimum over all $U \subseteq E$ of $r_1(U) + r_2(E \setminus U)$ is obtained for some closed set U, with respect to M_1, such that $E \setminus U$ meets each color class E_i in either zero or n_i edges.*

Proof. Among all subsets $U \subseteq E$ that minimize $r_1(U) + r_2(E \setminus U)$, let U be such that $r_1(U)$ is maximized. If U is not closed, then there is some edge e whose removal from $E \setminus U$ and addition to U, means $r_1(U)$ is not changed while $r_2(E \setminus U)$ may decrease. Hence either U is closed or we get a larger U. Therefore, by induction we end with U being closed.

If $E \setminus U$ meets some color class E_i in $0 < a_i < n_i$ edges, then by removing any edge e in $(E \setminus U) \cap E_i$ and adding it to U we increase $r_1(U)$ by 1 (since U is closed!) and decrease $r_2(E \setminus U)$ by 1, contradicting our choice of U. ∎

From the claim above we conclude that $r_2(E \setminus U) = \sum_{i \notin S} n_i$, where S is the set of colors which meet $E \setminus U$ in zero edges. We also have $r_1(U) \geq r_1(\cup_{i \in S} E_i)$, since $\cup_{i \in S} E_i \subseteq U$, and finally

$$r_1(U) + r_2(E \setminus U) \geq r_1(\cup_{i \in S} E_i) + \sum_{i \notin S} n_i \geq \sum_{i \in S} n_i + \sum_{i \notin S} n_i = \sum_{i=1}^{k} n_i.$$

The last inequality follows from the assumptions in Equation (1). We conclude that for every $U \subseteq E$, $r_1(U) + r_2(E \setminus U) \geq \sum_1^k n_i$, which implies, by the matroid intersection theorem, the existence of a forest with $|E_i| = n_i$ for $i = 1, \ldots, k$, namely, a forest whose profile is $\bar{p}$. ∎

3.4 An alternative proof of Theorem 5, and the polytope of feasible profiles

We need the following definition and claims.

Definition 3 For a k-edge-colored graph as above, a set $S \subset [k]$, $|S| \neq [k]$ of colors is called *critical* if Equation (1) is met for S with equality.

Claim 3.2 *Suppose that G meets Equation (1) with respect to $\bar{p} = (n_1, \ldots, n_k)$ for every $S \subseteq [k]$. Assume further that rank$(E_1) = n_1$ (namely $\{1\}$ is critical). Then for any independent edge $e \in E_1$, Equation (1) is met for $G' = G/e$, with respect to $\bar{p}' = (n_1 - 1, n_2, \ldots, n_k)$ and every set $S \subseteq [k]$.*

Proof. Contracting an edge may cause the rank of any set to decrease by at most 1. Thus Equation (1) may cease to hold on S, with respect to G' only if S is critical with respect to G.

Let $S \subseteq [k]$. If $1 \in S$, then $\text{rank}_{G'}(\cup_{i \in S} E_i) \geq \text{rank}_G(\cup_{i \in S} E_i) - 1$, as we have contracted a unique edge. On the other hand, the corresponding sum of n_i' with respect to $\bar{p}'$ is $(\sum_{i\in S} n_i) - 1$ (as n_1 is decreased by 1). Hence Equation (1) for S in G implies the condition for S in G'.

If $1 \notin S$, Equation (1) may fail to hold only if S is critical in G, and $\text{rank}(\cup_{i\in S} E_i)$ has dropped by 1. But this may occur only if e is spanned by $\cup_{i\in S} E_i$. This implies, however, that $\text{rank}_G(\cup_{i\in S} E_i \cup E_1) \leq \text{rank}(\cup_{i\in S} E_i) + \text{rank}(E_1) - 1 = \sum_{i\in S\cup\{1\}} n_i - 1$ contradicting the fact that Equation (1) holds for $S \cup \{1\}$. ■

An immediate conclusion from the above is the following.

Claim 3.3 *Suppose that G meets Equation (1) with respect to $\bar{p} = (n_1, \ldots, n_k)$ and that rank$(E_1) = n_1$ (namely $\{1\}$ is critical). Then Equation (1) is met for $G' = G/E_1$, with respect to $\bar{p}' = (n_2, \ldots, n_k)$, and every set $S \subseteq \{2, \ldots, k\}$.*

Proof. We apply Claim 3.2 repeatedly for the sequence $e_1, \ldots, e_{n_1}$ of independent 1-edges. ■

Proof. [**of Theorem 5**] Again, we only prove sufficiency of the condition, namely, we assume that Equation (1) holds for all $S \subseteq [k]$ in G with respect to $\bar{p} = (n_1, \ldots, n_k)$. We will then show how to find a forest with profile $\bar{p}$.

The proof is by induction on $|E|$. We may assume, without loss of generality, that $n_i \geq 1$ for every i, as otherwise one may delete those E_i for which $n_i = 0$. We may also assume that $k \geq 2$, as for $k = 1$ the statement is trivial.

Suppose, first, that there is no critical set $S \subsetneq [k]$. Pick any non-self loop edge $e \in E_1$. Clearly, Equation (1) is met for $G' = G/e$, with respect to $\bar{p}' = (n_1 - 1, n_2, \ldots, n_k)$, for all $S \subsetneq [k]$. By the induction hypothesis, G' contains a spanning forest F' which corresponds to a forest F in G with the required conditions.

Assume now, that $S \subset [k]$, $S \neq [k]$ is critical. Let G' be the subgraph of G that contains only the edges $\cup_{i\in S} E_i$ and let $G'' = G/G'$. We will show that there exists a forest $F' \subseteq G'$ consistent with the sequence $\bar{p}' = (n_i)_{i\in S}$, and that there is a forest $F'' \subseteq G''$ consistent with the sequence $\bar{p}'' = (n_i)_{i\notin S}$. Clearly $F' \cup F''$ is a forest of G for the original sequence $\bar{p}$.

Indeed, Equation (1) for G directly implies that the condition is met for G' with respect to $\bar{p}'$. Hence by induction (as $S \subset [k], S \neq [k]$), the existence of F' follows.

We now show that Equation (1) holds for G''. Again, by induction, this will imply the existence of F''. Indeed, G'' is constructed from G by contracting F', a maximal independent set in $E' = \cup_{i\in S} E_i$. But this is just the premise of Claim 3.3, when all the edges in E' are viewed as having the same color (recalling that S is critical). Hence by Claim 3.3, Equation (1) holds for G''. ■

4 The contractors' polytope

Our aim in this section is to formulate the contractors' MST problem for discount cardinality functions as a concave (or piece-wise linear) optimization problem over an explicitly given polytope. For $k = O(\log n)$ this polytope is of $poly(n)$ faces, and hence our hope is that this will facilitate heuristics or possibly deterministic good algorithms for the contractors problem when $k = O(\log n)$ and the functions are concave. We note that for constant k, as there are only $O(n^k)$ possible profiles, one can exhaustively search them all and choose the one that gives the best cost. Hence, for constant k, the contractors' problem with any collection of cardinality functions (not necessarily discount) is reducible to the feasibility problem, and can be solved in polytime.

The discussion below holds for any k. Moving to a "geometric" jargon, Theorem 5 can be alternatively stated as follows:

Observation 4.1 (Theorem 5 restated) *Let $G = (V, E)$, $E = \cup_1^k E_i$ be an instance to the k contractors' problem. Then the feasible profiles for G are all integral points of*

$$P_F = \{(x_1, \ldots, x_k) \in \mathbb{R}_+^k, \text{ such that, } \forall \mathrm{S} \subseteq [\mathrm{k}], \sum_{\mathrm{i} \in \mathrm{S}} \mathrm{x_i} \leq \mathrm{rank}(\cup_{\mathrm{i} \in \mathrm{S}} \mathrm{E_i})\}.$$

Furthermore, the feasible profiles of trees are these integral points in P_F for which $\sum_1^k n_i = rank(G)$.

Our main goal in this section is to show that P_F is integral, namely, that the extremal points of P_F are integral, implying that the extremal points of the intersection of P_F with the hyperplane $\sum_1^k x_k = \text{rank}(G)$ are integral.

We start by observing that Claims 3.2 and 3.3 also hold for non-integral partitions of n, namely:

Claim 4.1 *Let $x_i \geq 0$, $i = 1, \ldots, k$, and suppose that G satisfies Equation (1) with respect to $\bar{x} = (x_1, \ldots, x_k) \in \mathbb{R}_+^k$. Assume further that $rank(E_1) = x_1$ (namely $\{1\}$ is critical), and let $e \in E_1$ be a non-loop edge. Then Equation (1) holds for $G' = G/e$, with respect to $\bar{x}' = (x_1 - 1, x_2, \ldots, x_k)$, and every set $S \subseteq \{2, \ldots, k\}$.*

Proof. This claim is the analog version of Claim 3.2 for non-integer vectors $\bar{x}$, and its proof is identical to the proof of Claim 3.2. ∎

Lemma 4.1 *Let $x_i \geq 0$, $i = 1, \ldots, k$ and suppose that G meets Equation (1) with respect to $\bar{x} = (x_1, \ldots, x_k) \in \mathbb{R}_+^k$. Assume further that $rank(E_1) = x_1$ (namely $\{1\}$ is critical). Then Equation (1) is met for $G' = G/E_1$, with respect to the colors $\{2, \ldots, k\}$ and $\bar{x}' = (x_2, \ldots, x_k)$, and every set $S \subseteq \{2, \ldots, k\}$. Moreover, if in addition $\sum_1^k x_i = rank(G)$ then $\sum_2^k x_i = rank(G/E_1)$.*

Proof. For the first part we just repeat contracting every non-loop edge in E_1 applying repeatedly Claim 4.1. For the second part, if $\sum_1^k x_i = \text{rank}(G)$, then $\text{rank}(G/E_1) = \sum_2^k x_i$ since $x_1 = \text{rank}(E_1)$. ∎

Definition 4 (Contractor Polytope) Let H be the hyperplane defined by

$$H = \{x = (x_1, \ldots, x_k) \in \mathbb{R}^k_+ \ ; \sum_1^k x_i = \text{rank(G)}\}.$$

Define the *contractor polytope*, $P_c \subseteq \mathbb{R}^k$, to be $P_c = P_F \cap H$.

Note that the contractor polytope P_c is a face of P_H and its integral points (if any) are the profiles of spanning trees in G.

Theorem 6 *The extremal points of P_c are all integral.*

Proof. To prove the statement it is enough to show that for every *linear* cost function $c \in \mathbb{R}^k$, $\min c^T x,\ x \in P_c$, is achieved on an integer point [4].

We prove the latter fact by induction on k. For $k = 1$ it is trivial (for any G), and for $k = 2$ it is quite immediate for any G.

Assume that the theorem is established for every G and $k' < k$.

Fix a cost function c and let $x \in P_c$. We show that there is an integral $\bar{p} \in P$ for which $c^T \bar{p} \leq c^T x$, or that x is not optimal with respect to c. We may assume, without loss of generality, that c is non constant, that is, $c_1 < c_k$, as otherwise every point in P_c has the same cost. In particular so does any maximal spanning forest.

Assume first that there is no critical set with respect to x. Namely, for every $T \subset [k],\ T \neq [k]$, Equation (1) holds with strict inequality, and $x_i > 0, i = 1, \ldots k$.

Let δ be the minimum slack in Equation (1), over all $S \neq [k]$; namely:

$$\delta = \min\{\text{rank}(\bigcup_{i\in S} E_i) - \sum_{i \in S} x_i \ ; \quad S \subseteq [k],\ S \neq [k], S \neq \emptyset\}.$$

Let $\epsilon = \min\{\delta, x_k\}$, and $x^* = (x_1 + \epsilon, x_2, \ldots, x_{k-1}, x_k - \epsilon)$. It is easy to see that $x^* \in P_c$ and with $c^T \cdot x^* < c^T \cdot x$ (on account of $c_1 < c_k$).

Hence we may assume that there is a non-empty subset $S \subset [k],\ S \neq [k]$, that is critical. We proceed similarly as in the proof of Theorem 5. Let G' be the subgraph of G that contains only the edges $\cup_{i \in S} E_i$ and let $G'' = G/G'$. Let $x' = (x_i)_{i \in S} \in \mathbb{R}^{|S|}$. Obviously x' is in the corresponding polytope P' for G' and by induction, there is a corresponding profile $\bar{p}' = (p_i)_{i \in S}$ and a corresponding forest F' of G' attaining the cost $\sum_{i \in S} c_i p_i \leq \sum_{i \in S} c_i x_i$.

Now, G'' corresponds to the graph obtained by contracting the forest F'. In particular considering $\cup_{i \in S} E_i$ as corresponding to one color that is critical, this is just the premise of Lemma 4.1. Hence the first part of Lemma 4.1 implies that $x'' = (x_i)_{i \notin S}$ is a feasible point in P''_F, the polytope corresponding to G''. Further the second part asserts that $[k] \backslash S$ is critical with respect to x'', namely x'' is feasible in P'' the corresponding polytope for G''. Therefore, by induction, there is a profile, namely, an integral point $\bar{p}'' = (p_i)_{i \notin S}$ in P'', that achieves the cost $\sum_{i \notin S} c_i p_i \leq \sum_{i \notin S} c_i x_i$.

The forest $F' \cup F''$, which is a maximal spanning forest of G, achieves a cost that is at most $c^T x$ and this completes the proof. ■

4.1 Implication for small k

We have already observed that for constant k, the contractors' MST can be solved for any set of cardinality functions. Theorem 6 implies that for $k = O(\log n)$ finding the optimal contractor's MST is a minimization of a concave function over a polytope that is explicitly given by its $poly(n)$ facets. While concave optimization over a polytope is NP-hard in general, this setting opens a way for various heuristics, see e.g., [1] and citations within.

To give on example of usability, recall that the k functions are assumed to be cardinality discrete functions each. Namely, they can be extended to piece wise linear. Suppose that each is the combination of at most r linear functions. Then the MST global cost would be piece-wise linear (in dimension k) that is defined by at most r^k linear functions. If r is a constant independent of n then for $k = O(\log n)$, $r^k = poly(n)$. Hence, one could find the optimal point in the polytope for each of the linear functions, and then check the feasibility of each and choose the optimal. Finally, once the optimal value is found, finding the actual tree that achieves it is standard: an edge $e \in E_1$ may be deleted, and the optimal value for $G - e$ can be determined. If this is identical to that of G, it means that e is not essential to the solution, and the problem is reduced to $G - e$. Otherwise, e is contracted (and is forced to be in the solution), and again the problem is reduced to G/e. Hence, recursively, a solution attaining the optimal value can be constructed.

5 Conclusion

We have defined a new concept, a variant of the minimum spanning tree problem, where each edge in the graph belongs to some contractor, and each contractor has his/her own cost function for paving a subset of the edges. We proved that the general optimization problem for this concept is NP-hard, even when restricted to cardinality cost functions, i.e. cost functions which depend only on the number of edges, and the number of contractors is relatively large $\Omega(\log n)$. We gave a characterisation of when a tree, with a prescribed number of edges from each color class exists, and a new efficient algorithm for finding such a tree when $k = O(\log n)$. Finally, we proved that the polytope of feasible solutions for cardinality cost functions values is integral.

An interesting open problem is to find a good approximation for the optimal tree. It is not without hope that for the case of $k = O(\log n)$ contractors there is a polynomial time algorithm.

References

[1] H. P. Benson, Concave Minimization: Theory, Applications and Algorithms, In: *Handbook of Global Optimization* (eds. R. Horst and P. M. Pardalos), Springer 1995, 43–148.

[2] O. Borůvka, O jistém problému minimálním (About a certain minimal problem), *Práce Moravské Přírodovědecké Společnosti Brno* **3** (1926), 37–58.

[3] J. Edmonds, Submodular functions, matroids and certain polyhedra, In: *Combinatorial Structures and Their Applications*, Proc. Calgary (1969), Gordon and Breach (1970), 69–87.

[4] J. Edmonds and R. Giles, In: A. Schrijver, *Combinatorial Optimization, Polyhedra and Efficiency*, Springer, Vol **A**, 2003, Ch. 5, p. 74.

[5] G. Goel, C. Karande, P. Tripathi and L. Wang, *Approximability of Combinatorial Problems with Multi-agent Submodular Cost Functions*, Proc. 50th Annual IEEE Symposium on Foundations of Comp. Sc., FOCS (2009), 755–764.

[6] V. Jarník, O jistém problému minimálním, *Práce Moravské Přírodovědecké Společnosti Brno* **6** (1930), 57–63.

[7] R. M. Karp, Reducibility among combinatorial problems, In: *Complexity of Computer Computations* (Eds. R. E. Miller, J. W. Thatcher and J. D. Bohlinger), The IBM Research Symposia Series, Springer, Boston, MA, 1972.

[8] J. B. Kruskal, On the shortest spanning subtree of a graph and the travelling salesman problem, *Proc. Amer. Math. Soc.* **7** (1956), 48–50.

[9] E. L. Lawler, Matroid intersection algorithms, *Math. Prog.* **9** (1975), 31–56.

[10] S. Prasanna, IIIT-Bangalore, personal communication.

[11] R. C. Prim, Shortest connection networks and some generalizations, *Bell Sys. Tech. J.* **36** (1957), 1389–1401.

[12] A. Schrijver, *Combinatorial Optimization, Polyhedra and Efficiency*, Vol. **B**, Matroids, trees, stable sets, Springer-Verlag, Berlin, 2003, Chapters 39–41.

(Received 15 Nov 2017; revised 3 July 2017, 30 Jan 2019)