Highlights

On the Complexity of Degree-Constrained Network Design Problems with Nonlinear Objectives

Daniele Catanzaro, Gwenaël Joret, Brieuc Pierre, Francesco Pisanu

- New hardness results on network design problems are given for several objective functions that are not necessarily linear;
- hardness for the minimum spanning k-ary tree measured in average weighted path-length;
- hardness for the maximum and minimum spanning k-ary tree given the leaves' dissimilarities.

On the Complexity of Degree-Constrained Network Design Problems with Nonlinear Objectives

Daniele Catanzaro[a], Gwenaël Joret[b], Brieuc Pierre[a] and Francesco Pisanu[a,*]

[a]Center for Operations Research and Econometrics, Université Catholique de Louvain, Voie du Roman Pays 34, Louvain-la-Neuve, 1348, Belgium

[b]Computer Science Department, Université Libre de Bruxelles, Boulevard du Triomphe CP210/01, Bruxelles, 1050, Belgium

ARTICLE INFO

Keywords:
network design
k-ary trees
computational complexity

ABSTRACT

We prove the $\mathcal{NP}$-hardness of some versions of the network design problem that arise when strict degree constraints are imposed on the internal vertices of the spanning subgraph, and the objective function depends, possibly nonlinearly, on the length of the shortest path between pairs of vertices.

1. Introduction

The *Network Design Problem (NDP)* is a foundational problem in combinatorial optimization that consists of determining the optimal structure of a network while achieving specific objectives and simultaneously satisfying resources, capacity, demands, and budget constraints [1]. The NDP generalizes many classical optimization problems, such as the Steiner tree, the facility location, the multi-commodity flow, the shortest path, and the spanning tree problems, and serves as a unifying framework for modeling numerous network-based real-world applications [2, 3, 4]. Formally, consider a simple connected undirected graph $G = (V, E)$, with vertex-set V and edge-set E, and a weight function $w\colon E \to \mathbb{Q}_0^+$ that associates each edge in E with a nonnegative rational. For any two distinct vertices $u, v \in V$, let P_{uv}^G denote a shortest path between u and v in G and define the *weighted path-length* between them as $L_G(u, v) = \sum_{e \in P_{uv}^G} w_e$. Then, the NDP consists in finding a connected spanning subgraph $H = (V, E') \subseteq G$ that solves the following optimization problem:

$$\min \sum_{u,v \in V} L_H(u, v)$$
$$\text{s.t.} \sum_{e \in E'} w_e \leq B.$$

The $\mathcal{NP}$-hardness of the NDP was first established by Johnson et al [1]. Since then, numerous versions of the NDP have been proposed in the literature, the vast majority of which proved to be $\mathcal{NP}$-hard as well [5]. Several noteworthy examples that are particularly relevant to this work concern the search for a spanning tree subject to specific objectives and constraints. These include: the *minimum k-spanning tree problem* [6], which seeks a minimum-length tree spanning at least k vertices of a given graph; the *maximum leaf spanning tree problem* [7], which aims to find a spanning tree with the maximum possible number of leaves; and the *minimum degree spanning tree problem* [8], which consists of finding a spanning tree that minimizes the maximum vertex degree. Table 1 summarizes the key references for these problems, including their known complexity and approximability results.

*Corresponding author

daniele.catanzaro@uclouvain.be (D. Catanzaro); gwenael.joret@ulb.be (G. Joret); brieuc.pierre@uclouvain.be (B. Pierre); francesco.pisanu@uclouvain.be (F. Pisanu)

ORCID(s): 0000-0001-9427-1562 (D. Catanzaro); 0000-0002-7157-6694 (G. Joret); xxxx (B. Pierre); 0000-0003-0799-5760 (F. Pisanu)

Table 1
A summary of known $\mathcal{NP}$-complete versions of the NDP.

Problem	Approximability	Reference
SNDP		[1]
Minimum k-spanning tree	Approximable within 3	[6, 9]
Maximum leaf spanning tree	Approximable within 3	[7, 10]
Minimum degree spanning tree	Approximable within 1 Not approximable within $3/2-\varepsilon$, for some $\varepsilon > 0$	[8, 11]
Balanced Minimum Evolution Problem	Not approximable within c^n, for some $c > 1$ Approximable within 2 for metric instances	[12]
Fixed-tree Balanced Minimum Evolution Problem	Not approximable within c^n, for some $c > 1$	[13]

Contributions. This article adds to the above literature new complexity results about some versions of the NDP that arise when imposing strict degree constraints on the internal vertices of the spanning tree. In particular, we show in Section 3 that the following problem is strongly $\mathcal{NP}$-hard:

Problem 1 (The k-NDP)**.** *Given an undirected graph $G = (V, E)$, a weight function $w : E \to \mathbb{Q}^+$, and a positive integer k, find a spanning tree T of G, that minimizes $\sum_{u,v\in V} L_T(u, v)$ and whose internal vertices have degree k.*

We also study the complexity of two further versions of the NDP that arise when the objective function depends nonlinearly on the number of edges in the shortest path between pairs of leaves. To state them, consider a simple connected undirected graph $G_n^k = (\Gamma \cup V_I, E)$, for some integer $n > k \geq 3$, with $|\Gamma| = n$, $|V_I| = n - 2(k-2)$, and $E = \{uv : u, v \in V_I, u \neq v\} \cup \{uv : u \in \Gamma,\ v \in V_I\}$. We say that a spanning tree T of G_n^k is an *Unrooted k-Tree (UkT) of* Γ if Γ is the leaf-set of T and each internal vertex of T has degree k. When $k = 3$, we say that T is an *Unrooted Binary Tree (UBT) of* Γ. Figure 1 shows an example of this graph when assuming $n = 8$ and $k = 3$. For a given UkT T of Γ and a pair of distinct vertices $i, j \in \Gamma$, we define $\tau_{ij} = L_T(i, j) = |P_{ij}^T|$ as the *path-length* between i and j in T when assuming each $w_e = 1$; in other words, τ_{ij} is a positive integer that encodes the number of edges on the unique path in T having i and j as starting and ending vertices, respectively. Finally, let $\mathbf{D}$ denote an input square matrix of order n, whose generic entry $d_{ij} \in \mathbb{Q}_0^+$ is a measure of the dissimilarity associated with the vertices $i, j \in \Gamma$. Then, we consider the following two problems:

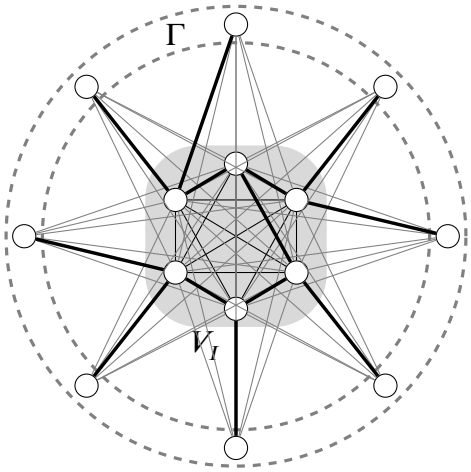

Figure 1: The graph $G_8^3 = (\Gamma \cup V_I, E)$. The vertices in Γ are arranged along the dashed annulus, while the vertices in V_I form a clique depicted inside the shaded region. Each vertex in Γ is adjacent to all vertices in V_I. Bold edges form a possible UBT of Γ.

Table 2
New $\mathcal{NP}$-hardness results for specific versions of the NDP proved in this article.

Problem	Reference	Problem	Reference
k-NDP	Theorem 1	Leaf-restricted k-NDP	Theorem 6
Rainbow degrees NDP	Theorem 7	Rainbow costs k-NDP	Theorem 8
Min-(k, β^τ)-NDP	Theorem 3	Fixed-topology Min-(k, β^τ)-NDP	Theorem 9
Max-(k, τ^α)-NDP	Theorem 5	Fixed-topology Max-(k, τ^α)-NDP	Theorem 10

Problem 2 (The Min-(k, β^τ)-NDP)**.** *Given an integer $n \geq 3$, a rational $\beta > 1$, and a $n \times n$ symmetric dissimilarity matrix $\mathbf{D} = \{d_{ij}\}$ associated with the vertices in Γ of G_n, find a UkT T of Γ that minimizes $\sum_{i,j\in\Gamma} d_{ij}\beta^{\tau_{ij}}$.*

Problem 3 (The Max-(k, τ^α)-NDP)**.** *Given an integer $n \geq 3$, a rational $\alpha \geq 1$, and a $n \times n$ symmetric dissimilarity matrix $\mathbf{D} = \{d_{ij}\}$ associated with the vertices in Γ of G_n, find a UkT T of Γ that maximizes $\sum_{i,j\in\Gamma} d_{ij}\tau_{ij}^\alpha$.*

Problems 2 and 3 originate from the context of hierarchical clustering [14, 15], where the goal is to identify a tree-like structure of nested clusters based on a given dissimilarity measure among input items. Both problems share strong similarities with the well-known *Balanced Minimum Evolution Problem (BMEP)*, a highly nonlinear $\mathcal{NP}$-hard network design problem extensively studied in molecular phylogenetics and consisting of finding an UBT over the item set Γ that minimizes the objective function $\sum_{i,j\in\Gamma} d_{ij}2^{-\tau_{ij}}$ [12, 13, 16, 17, 18, 19, 20]. We will prove in Sections 4 and 5 that also Problems 2 and 3 are $\mathcal{NP}$-hard. Moreover, we present in Section 6 additional complexity results that can be derived from the $\mathcal{NP}$-hardness proofs discussed in Sections 3-5. For ease of reading, we summarize in Table 2 the novel complexity results presented in this article.

Before turning to the formal proofs, we introduce in the next section some notation and definitions that will be used throughout the remainder of the paper.

2. Notation and definitions

All graphs introduced hereafter are assumed to be simple, connected, and undirected, unless stated otherwise. We say that two graphs $G' = (V', E')$ and $G'' = (V'', E'')$ have the same *topology* if they are isomorphic, i.e., structurally identical according to the standard notion of graph isomorphism described in [21]. With a slight abuse of language, we also say that a graph $G = (V, E)$ *has the same topology as* another graph, understood from context, to mean that the two are isomorphic.

Let $G = (V, E)$ be a graph and H be a subgraph of G. Then, $V(H)$ and $E(H)$ denote the vertex-set and edge-set of H, respectively. If $U \subseteq V$, $G[U]$ is the subgraph of G having U as a vertex-set and whose edges have both extremities in U, while $G \setminus U$ is the subgraph of G having $V \setminus U$ as vertex-set and whose edges have both extremities in the complement of U. We denote by $\delta(U)$ the *cut of* U, i.e., the subset of edges $uv \in E$ such that either u or v belongs to U. For the sake of notation, when $U = \{u\}$, we write $\delta(u)$ instead of $\delta(\{u\})$, and we say that $|\delta(u)|$ is the *degree* of u. If $U, W \subset V$ are disjoint, then $\delta(U, W)$ denotes the *cut of U and W* i.e., the subset of edges $uv \in E$ such that $u \in U$ and $v \in W$. Given a weight function $w\colon E \to Q \subseteq \mathbb{Q}_0^+$ and a subset $F \subseteq E$, we define the *weight of F* as $w(F) = \sum_{e\in F} w(e)$. Finally, we denote by $\delta(G)$ the minimum among the degrees of the vertices of G.

Let $T = (V, E)$ be a tree. We say that T is *k-ary*, for some integer $k \geq 3$, if every internal vertex has degree at most k. If $k = 3$, we say that T is *binary*. If every internal vertex of T has exactly degree k then we say that T is *unrooted*. Otherwise, there exists a nonempty subset $\mathcal{I}$ of internal vertices of T whose degree is at most $k - 1$; once chosen any vertex $r \in \mathcal{I}$, we refer to r as the *root* of T and we say that T is *rooted in r*. When it is not necessary to declare the root, we simply say that T is *rooted*. The *(first) common ancestor* of two vertices u and v of V is the unique (internal) vertex x for which $\tau_{ux} = \min\{\tau_{ux'} : \tau_{ux'} = \tau_{vx'}$ and $x' \in V\}$. The *depth* of a rooted k-ary tree is the maximum among the path-lengths between the root and every leaf. We denote the set of leaves of T as $\Gamma(T)$. To highlight the fact that

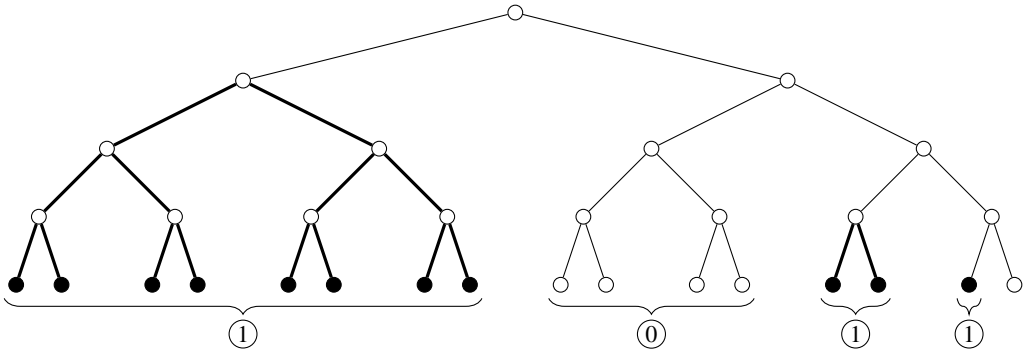

Figure 2: A binary distribution associated with a set Γ' of 11 vertices over a perfect binary tree with 16 leaves. The binary expression of 11 is '1011' and it is represented by three congruent perfect binary trees (highlighted with bold edges and black leaves) having 2^3, 2^1, 2^0 leaves.

we are considering a leaf instead of a generic vertex $u \in V$ (respectively $v \in V$) we denote it as i (respectively j). A *cherry* is a pair of leaves that are adjacent to their common ancestor.

If T is k-ary, then T is *full* if it is rooted and every internal vertex that differs from the root has degree k, and *perfect* if the path-length between the root and any leaf equals the depth. By definition, any perfect k-ary tree of depth ℓ is full and has $(k-1)^\ell$ leaves. Perfect binary trees have the smallest topology from among all full binary trees with the same number of leaves [22]. If T is k-ary, then T is a *caterpillar* if there are exactly two disjoint subsets, each of which has $k-1$ leaves, and such that every pair of elements of each of them forms a cherry. A folklore result shows that a k-ary tree with n leaves is a caterpillar if and only if there are two leaves whose path-length is $\frac{n-2}{k-2}+1$.

A tree $T' = (V', E')$ is a *congruent subtree of* T, if $V' \subseteq V$, $E' = E(V')$, and $\Gamma(T') \subseteq \Gamma(T)$. Two congruent subtrees of T are *disjoint*, if their vertex-sets do not intersect. If T is perfect, and $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$ are two congruent perfect subtrees of T of depth h_1 and h_2 respectively, then the common ancestor of every pair of leaves of $\Gamma(T_1) \cup \Gamma(T_2)$ is unique. In the case x is this common ancestor, T_1 and T_2 are *adjacent* if the (unique) congruent perfect subtree of T rooted in x has depth $\max\{h_1, h_2\} + 1$.

Suppose that T is a caterpillar. We say that T' is a *comb of* T if T' is a congruent subtree of T and for every $i \in \Gamma(T')$ there is $j \in \Gamma(T')$ for which $\tau_{ij} = 3$ holds. Two leaves i and j of T' are *adjacent* if $\tau_{ij} = 3$. If a leaf i of T' is adjacent to only one leaf, we say that i is an *extremal* of T'.

Let $\Gamma' \subseteq \Gamma(T)$. If T is a perfect k-ary tree, a congruent perfect subtree T' of T is *maximal in* Γ', if $\Gamma(T') \subseteq \Gamma'$ and no other congruent perfect subtree T'' of T is such that $\Gamma(T') \subset \Gamma(T'') \subseteq \Gamma'$. Similarly, if T is a caterpillar, a comb T' of T is *maximal in* Γ' if $\Gamma(T') \subseteq \Gamma'$ and no other comb T'' of T is such that $\Gamma(T') \subset \Gamma(T'') \subseteq \Gamma'$.

A *binary distribution* of T is a family of pairwise disjoint congruent perfect subtrees $T^1, \dots, T^k$ of T of depth $0 < h_1 < h_2 < \dots < h_k$ such that T^i is adjacent to T^{i+1}, for all $i \in \{1, \dots, k-1\}$. We say that $\Gamma' \subseteq \Gamma(T)$ *forms* a binary distribution of T if there exists a binary distribution $\mathfrak{T}$ such that $\Gamma' = \cup_{T^i \in \mathfrak{T}} \Gamma(T^i)$. Note that, if c is the binary expression of $|\Gamma'|$, with k non-zero digits, a binary distribution $T^1, \dots, T^k$ of T associated with Γ' can be seen as associating to the leaves of T^i the i-th nonzero digit of c starting from the most significant digit. If c_i is such a digit, c_i is the binary representation of $|\Gamma(T^i)|$ (see, e.g., Figure 2).

3. On the $\mathcal{NP}$-hardness of the k-NDP

In the light of the above notation and definitions, we study in this section the complexity of the k-NDP, by establishing the following result:

Theorem 1. *The k-NDP is strongly $\mathcal{NP}$-hard for every integer $k \in [3, \delta(G)]$.*

We prove the statement by showing that the following decision version of the k-NDP is $\mathcal{NP}$-complete:

Problem 4 (Decision version of the k-NDP — (d-k-NDP))**.** *Given an undirected graph $G = (V, E)$, a weight function $w : E \to \mathbb{Q}_0^+$, and two constants $C \in \mathbb{Q}^+$ and an integer $k \in [3, \delta(G)]$, decide whether there exists a spanning UkT T of G such that $\sum_{u,v\in V} L_T(u, v) \leq C$.*

We prove Theorem 1 by reducing the following classical $\mathcal{NP}$-complete problem to the d-k-NDP:

Problem 5 (The Exact m-Cover Problem (ECP) [11])**.** *Given two positive integers $\ell \geq 1$ and $m \geq 3$, a finite set $M = \{\mu_1, \dots, \mu_{\ell m}\}$, and a family of subsets $S \subseteq 2^M$, decide whether there exists a subset $P \subseteq S$ such that P is a partition of M.*

Garey and Johnson [11] showed that the ECP is $\mathcal{NP}$-complete in the case $m = 3$. This result, however, can be easily generalized to any $m > 3$. In particular, the following statement – usually considered folklore but reported here for the sake of completeness – holds:

Proposition 1. *The ECP is $\mathcal{NP}$-complete for every integer $m \geq 3$.*

Proof. Denote by $M' = \{\mu_1, \dots, \mu_{3\ell}, \mu_{3\ell}^1, \dots, \mu_{3\ell}^q\}$ the multiset obtained by duplicating $q = \ell(m-3)$ times the last element of M (i.e., $\mu_{3\ell}^j = \mu_{3\ell}$ for all $j = \{1, \dots, q\}$). Then, given a family of subsets $S \subseteq 2^M$, we can define the set

$$S' = \{(\mu_{i_1}, \mu_{i_2}, \mu_{i_3}, \mu_{i_3}^1, \dots, \mu_{i_3}^q) : (\mu_{i_1}, \mu_{i_2}, \mu_{i_3}) \in S\} \subset 2^{M'}$$

as an instance of the ECP. Now observe that, by construction,

$$P' = \{(\mu_{j_1}, \mu_{j_2}, \mu_{j_3}, \mu_{j_3}^1, \dots, \mu_{j_3}^q) : (\mu_{j_1}, \mu_{j_2}, \mu_{j_3}) \in P \subseteq S\}$$

is a partition of M' if and only if P is a partition of M. As the Exact 3-Cover is $\mathcal{NP}$-complete [11], so is the ECP. Thus, the statement follows. □

We now prove Theorem 1 by assuming that $m = (k-1)^2$. To this end, consider a positive integer r and define $s = |S|$. Consider a connected graph $G = (V, E)$ defined as follows. The vertex-set V is partitioned into the vertex-sets V_r, V_s, V_R, V_S, and $M = \{\mu_1, \dots, \mu_{\ell(k-1)^2}\}$, where the first four sets have cardinality $r, s, (k-2)r + (k-3)s + 2$, and ks, respectively. In particular, we consider s elements of V_S indexed by S. The edge-set E is as follows: every vertex of $V_r \cup V_s$ has degree k, and $V_r \cup V_s$ induces a path with terminal vertices u' and u''. Every vertex in V_R has degree 1 and is adjacent to a unique vertex in $V_r \cup V_s$. Specifically, every vertex of V_r is adjacent to $k-2$ vertices of V_R. Similarly, every vertex of V_s is adjacent to $k-3$ vertices of V_R. Exceptionally, u' and u'' also have another neighbor in V_R each, so that they have degree k in G. Every vertex indexed by S is adjacent to k vertices: one in V_s, and $k-1$ in $V_S \setminus S$ such that $V_S \setminus S$ respects the given degree constraint and can be partitioned with respect to the adjacencies of S. Finally, every $\sigma \in S$ is adjacent to k vertices. Because every $u \in V_S \setminus S$ is adjacent to precisely one element σ of S, then u is adjacent to $\mu_i \in M$ only if $\mu_i \in \sigma$. Figure 3 represents G for some $r > 4$, $s = 3$, and $k = 3$.

We complete the construction of the instance of Problem 4 by defining the edge-weight function w such that $w(e) = 1$ if e covers one of the vertices in $V_R \cup M$ or exactly one of the terminals of e belongs to V_s, and $w(e) = 0$ otherwise. Moreover, by a slight abuse of notation, we will call *feasible solution* every spanning UkT of G.

For a feasible solution T and any couple of vertex-sets $U, U' \subseteq V$, we define *cost of the pair* (U, U') as

$$L_T(U, U') = 2 \sum_{(u,v)\in U\times U'} L_T(u, v).$$

When $U = U'$, we simply write $L_T(U)$ and we refer to this quantity as the *cost of* U. In particular, when $U = V$, we refer to $L_T(V)$ as the *cost of* T.

By construction, the edge-set of every feasible solution contains $E(G \setminus (V_S \setminus S))$. Moreover, for every feasible solution, every $u' \in M$ and $r' \in V_R$ (respectively, $v' \in V_s \cup V_r$) we have that $L_T(r', u') = 3$ (respectively, $L_T(r', u') = 2$). Therefore, the cost

$$C_0 = L_T(R) + L_T(R, V_r) + L_T(R, V_s) + L_T(R, S) + L_T(V_r) + L_T(V_r, V_s) + L_T(V_r, S) + L_T(V_s) + L_T(V_s, S) + L_T(S),$$

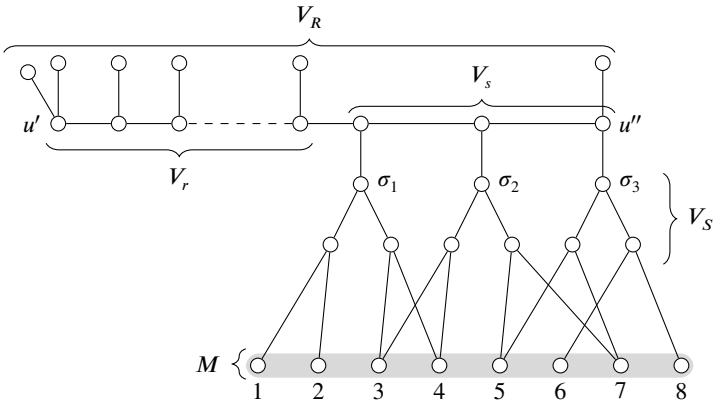

Figure 3: An example of an instance of Problem 4 when assuming $k = 3$, $r > 4$, $M = \{1, \dots, 8\}$, and $S = \{\sigma_1 = (1, 2, 3, 4), \sigma_2 = (3, 4, 5, 7), \sigma_3 = (5, 6, 7, 8)\}$. Note that the order imposed on the vertices of the path joining u' and u'' is not required.

is a constant for every feasible solution, i.e., this cost does not depend on the choice of T. Finally, denote by $\mathcal{T}_k$ the set of feasible solutions whose edge-set contains $E(V_S)$. Then, in the light of the above notation, the following proposition holds:

Proposition 2. *For $r > 8k^3\ell^2 s$, the cost of any feasible solution in $\mathcal{T}_k$ is strictly smaller than the cost of every feasible solution not in $\mathcal{T}_k$.*

Proof. Let $T \in \mathcal{T}_k$ and T' be any feasible solution not in $\mathcal{T}_k$. First, observe that

$$L_T(V) = C_0 + L_T(V_R, V_S \setminus S) + L_T(M) + L_T(M).$$

Since, $L_T(V_R, V_S \setminus S) = 4(k-1)s|R|$ by definition of G, $L_T(M) \leq 6\ell^2(k-1)^2$ as $L_T(u, v) \leq 3$ holds true for any couple of vertices u, v in M, and $L_T(M, V_S) \leq 8\ell sk(k-1)^2$ as $L_T(u, v) \leq 4$ holds true for any couple of vertices (u, v) in $M \times V_S$, we have that

$$L_T(V) \leq C' = C_0 + 4(k-1)s|R| + 6\ell^2(k-1)^4 + 8\ell sk(k-1)^2.$$

Also observe that, since T' is k-ary, all neighbors of some $\sigma \in S$ in $G[V_S]$ are leaves of T', contributing $8(k-1)|R|$ to $L_{T'}(V)$. Hence, the $L_{T'}(V) > C'' = C_0 + 4(k-1)s(|R|-1) + 8(k-1)|R|$. Then, if $r > 8k^3\ell^2 s$, it is straightforward to see that $C' < C''$, i.e., that $L_T(V) < L_{T'}(V)$. Thus, the statement follows. □

Now, assume that $r > 8k^3\ell^2 s$. By Proposition 2, every feasible solution T whose cost $L_T(V)$ is smaller than some appropriate constant belongs to $\mathcal{T}_k$. Moreover, always by Proposition 2, the cost $L_T(V_R, V_S \setminus S)$ is constant for every $T \in \mathcal{T}_k$, and every vertex in M has degree 1. This fact implies that the cost $L_T(M, V_S)$ is constant for every $T \in \mathcal{T}_k$, since for every vertex μ_i in M there is precisely one vertex s_j in V_S such that the $L_T(\mu_i, s_j) = 1$, and $L_T(\mu_i, s_{j'}) = 3$ for all $s_{j'} \in V_S$, with $j' \neq j$. Therefore, the cost $C_1 = L_T(V_R, V_S \setminus S) + L_T(M, V_S)$, is constant with respect to any $T \in \mathcal{T}_k$.

We now observe that the following proposition holds:

Proposition 3. *There exists $T^* \in \mathcal{T}_k$ such that $L_{T^*}(V) = C_0 + C_1 + 2(k-1)^2(2(k-1)^2 - \ell - 1)$ if and only if there exists a partition $P \subseteq S$ of M.*

Proof. Let $T \in \mathcal{T}_k$. Denote by S_σ the set of vertices of V_S that are adjacent in T to $\sigma \in S$. Since T is k-ary we can define the number s_h of sets S_σ, for all $\sigma \in S$ such that the number of vertices of M adjacent in T to S_σ is exactly h,

where $h \in H = \{0, k-1, 2(k-1), \dots, (k-1)^2\}$. Moreover, let $S(T)$ be the set of couples (u, v) of vertices of M such that $L_T(u, v) = 2$. Then,

$$L_T(M) = 4(k-1)^2((k-1)^2 - 1) - 4|S(T)| = 4(k-1)^2((k-1)^2 - 1) - 2\sum_{h \in H} h \cdot s_h.$$

Note that the cost $L_T(M)$ decreases as $2\sum_{h \in H} h \cdot s_h$ increases. In particular, since every h is a multiple of $k-1$, this sum is as small as possible when s_h is as big as possible. By definition, $s_{(k-1)^2} \leq \ell$, and $s_{(k-1)^2} = \ell$ if and only if all the corresponding S_σ's counted by $s_{(k-1)^2}$ are such that there is a partition P of M for which $\sigma \in P$. Moreover, if $s_{(k-1)^2} = \ell$, every other $s_h = 0$ for any $T \in \mathcal{T}_k$ by construction, and hence $L_T(M) = 2(k-1)^2(2(k-1)^2 - \ell - 1)$.

Therefore, assume that $s_{(k-1)^2} = \ell$ and set $C = C_0 + C_1 + 2(k-1)^2(2(k-1)^2 - \ell - 1)$. Then, $\sum_{u,v \in V} L_T(u, v) = C$ if and only if there is a partition $P \subseteq S$ of M. □

From the proof of Proposition 3 we can deduce that $L_{T^*}(V) \geq C$. Thus, if $r > 8k^3\ell^2 s$, answering "yes" or "no" to Problem 4 is equivalent to deciding whether S contains a partition of M or not in the ECP. As the ECP is $\mathcal{NP}$-complete, so is the d-k-NDP. □

4. On the $\mathcal{NP}$-hardness of the Min-(k, β^τ)-NDP

In this section, we prove the $\mathcal{NP}$-hardness of the Min-(k, β^τ)-NDP. We will first show that this result holds for $k = 3$ and $\beta = 2$, i.e., that

Theorem 2. *The Min-$(3, 2^\tau)$-NDP is strongly $\mathcal{NP}$-hard.*

Subsequently, we will generalize this result to any arbitrary integer $k > 3$ and rational $\beta > 1$.

We start by considering the following problem:

Problem 6. *Given a finite set Γ' and a positive integer n such that $2^n > |\Gamma'|$, find a perfect binary tree T with 2^n leaves that minimizes $\sum_{i,j \in \Gamma'} \phi(\tau_{ij})$, where $\phi(\cdot)$ is an arbitrary strictly increasing and non-negative function of the path-length τ_{ij}.*

We first observe that the following proposition holds:

Proposition 4. *Let T^* be an optimal solution to Problem 6 with input Γ' and n. Then, no two maximal congruent perfect binary subtrees of T^* whose leaves are indexed by Γ' have the same depth.*

Proof. Suppose, by contradiction, that T^* is an optimal solution to Problem 6. Denote by T_1 and T_2 two congruent perfect binary disjoint subtrees of T^*, (i) having the same depth h, (ii) being maximal in Γ', and (iii) whose leaves are indexed by Γ'. Denote by u the common ancestor of every couple of leaves $i \in \Gamma(T_1)$ and $j \in \Gamma(T_2)$, and by C_1, C_2, and C_3 the three connected components of $T \setminus \{u\}$. Without loss of generality, assume that $T_1 \subset C_1$ and $T_2 \subset C_2$. By maximality of T_1 and T_2, there exist two congruent perfect binary trees T_3 and T_4 of depth h that are contained in C_1 and C_2, and that are disjoint from and adjacent to T_1 and T_2, respectively. Moreover, always by maximality of T_1 and T_2, both $\Gamma(T_3)$ and $\Gamma(T_4)$ are not subsets of Γ'.

Denote by T' the perfect binary tree obtained by exchanging the leaves of T_1 with those of T_4. Similarly, denote by T'' the perfect binary tree obtained by exchanging the leaves of T_2 with those of T_3. Then, it is possible to show that the cost of either T' or T'' is strictly smaller than the cost of T^*. Specifically, for every i in $\Gamma(C_1)$ and j in $\Gamma(C_2)$, $\tau_{is} = \tau_{js}$ for all $s \in C_3$. Thus, by exchanging i with j, any possible change in the cost the objective function of the problem depends only on: the pairwise distances of vertices in $\Gamma(C_1) \cap \Gamma'$, those in $\Gamma(C_2) \cap \Gamma'$, and the distances between vertices in $\Gamma(C_1) \cap \Gamma'$ and $\Gamma(C_2) \cap \Gamma'$. Similarly, when we swap the positions indexed by $\Gamma(T_1)$ with those indexed by $\Gamma(T_4)$, $\sum_{i \in \Gamma(T_1), j \in \Gamma(T_4)} \phi(\tau_{ij})$ is invariant. Denote by C_0 these invariant costs. Also observe that any variation in the cost of T' and T'' depends on the following contributions:

- $l_i = \sum_{i' \in \Gamma(T_3)} \phi(\tau_{ii'})$ and $r_i = \sum_{i' \in \Gamma(T_2)} \phi(\tau_{ii'})$, where $i \in \Gamma(T_1)$;
- $l_j = \sum_{j' \in \Gamma(T_3)} \phi(\tau_{jj'})$ and $r_j = \sum_{j' \in \Gamma(T_2)} \phi(\tau_{jj'})$, where $j \in \Gamma(T_4)$.

Since T_1 and T_3 are perfect, the values l_i, r_i, l_j, and r_j do not depend on the choice of i and j in $\Gamma(T_1)$ and $\Gamma(T_4) \cap \Gamma'$, respectively, and $\sum_{i\in\Gamma(T_1)}(l_i + r_i) = |\Gamma(T_1)|(l_i + r_i)$, $\sum_{j\in\Gamma(T_3)}(l_j + r_j) = |\Gamma(T_4) \cap \Gamma'|(l_j + r_j)$.

Now, suppose that $\sum_{i\in\Gamma(T_1)}(l_i + r_i) > \sum_{j\in\Gamma(T_3)}(l_j + r_j)$. Then, the cost of T^* corresponding to $\Gamma(T_1)$ and $\Gamma(T_4) \cap \Gamma'$ is $c = |\Gamma(T_1)|(l_i+r_i)+|\Gamma(T_4)\cap\Gamma'|(l_j+r_j)$, while the corresponding one of T' is $c' = |\Gamma(T_4)\cap\Gamma'|(l_i+r_i)+(l_j+r_j)|\Gamma(T_1)|$. Due to the maximality of T_1, we have that $|\Gamma(T_4) \cap \Gamma'| < |\Gamma(T_1)|$, which implies that c' is strictly smaller than c. With similar arguments, it is easy to see that if $\sum_{i\in\Gamma(T_1)}(l_i + r_i) < \sum_{j\in\Gamma(T_3)}(l_j + r_j)$, then the cost of T'' is smaller than the one of T^*. Finally, note that the condition $\sum_{i\in\Gamma(T_1)}(l_i + r_i) = \sum_{j\in\Gamma(T_3)}(l_j + r_j)$ cannot hold true as, by maximality of T_1 and T_2, both $|\Gamma'(T_3)|$ and $|\Gamma'(T_4)|$ are strictly smaller than $|\Gamma'(T_1)|$ and $|\Gamma'(T_2)|$. Thus, the cost of either T' or T'' is strictly smaller than the corresponding one of T^*, contradicting the hypothesis of optimality of T^*. □

Now, let T^* denote an optimal solution to Problem 6 with input Γ' and n. Then, the following proposition holds:

Proposition 5. *The subset of leaves Γ' forms a binary distribution on T^* if and only if T^* is an optimal solution to Problem 6.*

Proof. We prove the statement by showing that every feasible solution to Problem 6 that does not form a binary distribution is not optimal. On the contrary, it is easy to see that every binary distribution has the same cost.

By Proposition 4, no two maximal congruent perfect binary subtrees of T^* have the same depth. Thus, there is a sequence $T^1, \dots, T^k$ of maximal congruent perfect binary trees of T^* having depth $h_1 < h_2 < \dots < h_k$. To conclude that Γ' forms a binary distribution over T^*, it remains to prove that T^i and T^{i+1} are adjacent, for all $i = 1, \dots, k-1$.

Let T^i be the first tree in the collection with respect to i such that T^i and T^{i+1} are not adjacent. Since T^i is maximal, there exists a (unique) congruent perfect binary subtree $\tilde{T}$ of T^* adjacent to T^i and having depth h_i and such that $\Gamma'(\tilde{T})$ is strictly contained in $\Gamma(\tilde{T})$. As in Proposition 4, to reduce the total cost it suffices to swap the leaves of T^{i+1} with those of $\tilde{T}'$, since $|\Gamma'(\tilde{T})| < |\Gamma'(T^{i+1})|$ by assumption. □

In the light of the above propositions, we now prove Theorem 2 by reducing the following classical $\mathcal{NP}$-hard problem to the Min-$(3, 2^\tau)$-NDP:

Problem 7 (The Graph Bisection Problem (GBP) [23])**.** *Given a simple connected undirected graph $G = (V, E)$ having $2p$ vertices, for some $p \in \mathbb{Z}^+$, find a partition of V into two equal-sized subsets U^* and W^* that minimizes $|\delta(U^*, W^*)|$.*

We will show that, given an optimal solution to an instance of the Min-$(3, 2^\tau)$-NDP, appropriately built from an instance of the GBP, it is possible to compute in polynomial time a sub-partition of the leaf-set Γ that induces an optimal solution to the given instance of the GBP. The $\mathcal{NP}$-hardness of the Min-$(3, 2^\tau)$-NDP, then, will immediately derive from the $\mathcal{NP}$-hardness of the GBP. We start by considering an input graph $G = (V, E)$ of the GBP, for some fixed $p \in \mathbb{Z}^+$. Denoted m as a positive integer such that $m > 2p^3$ and $p + m$ is the smallest possible power of two, we consider two further graphs, say $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, that are isomorphic to the complete graph on m vertices and such that V, V_1, and V_2 are mutually disjoint. We observe that, by definition, $m + p \leq 2(2p^3 + p)$; hence, the size of the graph $G \cup G_1 \cup G_2$ is polynomially bounded. Finally, we denote by M and γ two positive integers such that $M > 4p^2(p+m)^2$ and $\gamma > (2p + 2m)(2mM + 2p)^2$, respectively. Then, we build an instance of the Min-$(3, 2^\tau)$-NDP by setting

- $\Gamma = V \cup V_1 \cup V_2 \cup \{\hat{w}\}$, where $\hat{w}$ is a new vertex such that $\hat{w} \notin V \cup V_1 \cup V_2$;
- $d_{ij} = 1$ for all $ij \in E$;
- $d_{ij} = M$ for all $ij \in E_1 \cup E_2$;
- $d_{i\hat{w}} = \gamma$ for all $i \in V \cup V_1 \cup V_2$;
- and $d_{ij} = 0$ otherwise.

Now, let T^* denote the optimal solution to the Min-$(3, 2^\tau)$-NDP and let $\mathcal{T}_{(3,2^\tau)}$ denote the subset of feasible solutions to the Min-$(3, 2^\tau)$-NDP such that the removal of the vertex $\hat{w}$ from each of them (together with the relative incident edge) results in a perfect binary tree. Then, the following proposition holds:

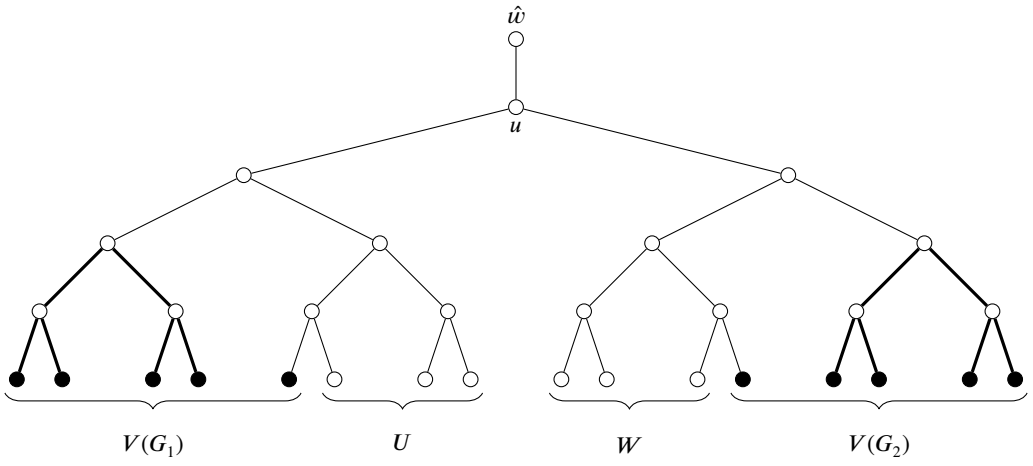

Figure 4: An example of the UBT in Proposition 6 when assuming $p = 3$ and $m = 5$. The UBT partitions the vertex-set V into two equal-sized vertex-subsets U, W as a consequence of the two binary distributions (highlighted with bold edges and black leaves) formed by the vertices in $V(G_1)$ and $V(G_2)$, respectively.

Proposition 6. *The optimal solution T^* belongs to $\mathcal{T}_{(3,2^\tau)}$.*

Proof. We prove the statement by showing that any feasible solution to the Min-$(3, 2^\tau)$-NDP that does not belong to $\mathcal{T}_{(3,2^\tau)}$ cannot be optimal. We start by observing that, there exists a feasible solution T to the Min-$(3, 2^\tau)$-NDP that belongs to $\mathcal{T}$ by definition of $\mathcal{T}_{(3,2^\tau)}$ and the choice of m. Indeed, $|V| + |V_1| + |V_2| = 2p + 2m$ is a power of two; hence, there exists a perfect binary tree whose leaves are indexed by $\Gamma \setminus \{\hat{w}\} = V \cup V_1 \cup V_2$. Now, denote by T' any feasible solution to the Min-$(3, 2^\tau)$-NDP that is not in $\mathcal{T}_{(3,2^\tau)}$, i.e., such that $T' \setminus \{\hat{w}\}$ is not a perfect binary tree, and let z and z' denote the values of the objective function of the Min-$(3, 2^\tau)$-NDP for T and T', respectively. Then, we have that $z < 2(\gamma(2^h + 1)(2p + 2m) + 2^h(2mM + 2p)^2)$, where $h = \log_2(2(p + m))$ is the depth of the perfect binary tree $T \setminus \{\hat{w}\}$, and $z' > 2\gamma\left(\sum_{\Gamma\setminus\{\hat{w}\}} 2^{\tau'_{i\hat{w}}}\right)$. Because perfect binary trees have the smallest topology from among all full binary trees with the same number of leaves, we have that $\sum_{\Gamma\setminus\{\hat{w}\}} 2^{\tau'_{i\hat{w}}} - (2^h + 1)(2p + 2m) \geq 1$. As $\gamma > (2p + 2m)(2mM + 2p)^2$, we have that $z' > z$. Since no particular assumption has been made on T', the cost of every feasible solution to the Min-$(3, 2^\tau)$-NDP that does not belong to $\mathcal{T}_{(3,2^\tau)}$ is strictly greater than z. Thus, by definition of optimality, T^* belongs to $\mathcal{T}_{(3,2^\tau)}$. □

Now, given the optimal solution T^* to the Min-$(3, 2^\tau)$-NDP, let u denote the root of $\hat{T} = T^* \setminus \{\hat{w}\}$ and let C_1 and C_2 denote two connected components of $\hat{T} \setminus \{u\}$. Then, the following proposition holds:

Proposition 7. *The vertex-sets V_1 and V_2 form a binary distribution over $\hat{T}$. Moreover, $\Gamma(C_1) = V_1 \cup U$ and $\Gamma(C_2) = V_2 \cup W$, where (U, W) is a partition of V such that $|U| = |W|$.*

Proof. First note that, as $M > 4p^2(p+m)^2$ and $|E| < p^2$, it holds that $\sum_{ij\in E} 2^{\tau_{ij}} < 2^h|E| < 2^h(2p)^2 < 2^h p^2(2p+2m) < M$, i.e., augmenting by one the distance between two leaves indexed by V_1 or two indexed by V_2 improves the cost of the objective function of the Min-$(3, 2^\tau)$-NDP more than any assignment of the indexes of V among the leaves of $\hat{T}$. Therefore, any optimal solution necessarily minimizes the partial costs $\sum_{ij\in E_1} 2^{\tau_{ij}}$ and $\sum_{ij\in E_2} 2^{\tau_{ij}}$. By Proposition 5, any optimal solution with respect to V_1 and V_2 forms two binary distributions on $T^* \setminus \{\hat{w}\}$.

As $|V_1| = |V_2| = m > p$, and $|\Gamma(C_1)| = |\Gamma(C_2)| = p + m$, the leaves indexed by V_1 and V_2 can form binary distributions in $\hat{T}$ if and only if $V_1 \subset \Gamma(C_1)$ and $V_2 \subset \Gamma(C_2)$. As a consequence, there are exactly p leaves of C_1 and p leaves of C_2 that are not labeled by V_1 or V_2. Thus, there is a partition (U, W) of V, such that $\Gamma(C_1) = V_1 \cup U$ and $\Gamma(C_2) = V_2 \cup W$, and $|U| = |V| = p$, which completes the proof (see Figure 4). □

We now show that the partition of G provided by Proposition 7 constitutes an optimal solution for the given instance of the GBP. To this end, let z^* denote the cost of any solution to the Min-$(3, 2^\tau)$-NDP such that the cut induced by the partition of Proposition 7 is optimal for the corresponding GBP instance. Let δ^* denote the cardinality of such a solution. Similarly, let z' denote the cost of any feasible solution to the Min-$(3, 2^\tau)$-NDP for which the cut induced by the partition of Proposition 7 is not optimal for the GBP instance. Denote by δ' the cardinality of such a solution and observe that

$$C_0 = 2\left(\sum_{ij\in E_1} 2^{\tau_{ij}} + \sum_{ij\in E_2} 2^{\tau_{ij}} + \sum_{i\in\Gamma\setminus\{\hat{w}\}} 2^{\tau_{i\hat{w}}}\right)$$

is constant by Propositions 6 and 7. Then, $z^* < \phi_1 = C_0 + 4\delta^*(2p + 2m) + 4p(|E| - \delta^*)$, while $z' > \phi_2 = C_0 + 4\delta'(2p + 2m) + 4(|E| - \delta')$. As $m > 2p^3$ (by construction) and $p|E| + \delta' < 2p^3$, we have that $\phi_1 < \phi_2$. Thus, the cost of any optimal solution to the Min-$(3, 2^\tau)$-NDP is at most z^* and the equal-sized partition of G provided by $(V \cap \Gamma(C_1), V \cap \Gamma(C_2))$ constitutes an optimal solution to the considered instance of the GBP. □

Generalization. The concatenation of the propositions presented in this section allowed us to prove Theorem 2. We briefly observe here, however, that such a theorem can be easily generalized to any integer $k > 3$ and rational $\beta > 1$. Specifically, to account for generic $k > 3$ in the construction of an instance of the Min-(k, β^τ)-NDP, it is sufficient (i) to assume that $m + p$ is a power of $k - 1$ (instead of 2); (ii) to add $H_1, \dots, H_{k-3}$ complete graphs of size $m + p$ each, such that G, G_1, G_2, and $H_1, \dots, H_{k-3}$ are all mutually disjoint; (iii) to note that, for γ large enough, $T^* \setminus \{\hat{w}\}$ is a perfect k-ary subtree containing exactly $k - 1$ congruent perfect k-ary subtrees $T_1, \dots, T_{k-1}$ of depth $m + p$; and (iv) to note that, for M large enough, the leaves of $T_1, \dots, T_{k-3}$ are indexed by the vertices of $H_1, \dots, H_{k-3}$, respectively, while the leaves of T_{k-2} and T_{k-1} are indexed by the vertices of G, G_1, and G_2. Concerning β, instead, it is easy to see that the $\mathcal{NP}$-hardness of the Min-$(k, 2^\tau)$-NDP persists when replacing 2 with a generic $\beta > 1$ due to the logarithmic identity $\log_{k-1} \beta^{\tau_{ij}} = \left(\log_\beta(k-1)\right)^{-1} \tau_{ij}$, where $\log_{k-1}(\cdot)$ arises from the depth of perfect k-ary trees. The scalar factor $\left(\log_\beta(k-1)\right)^{-1}$ is a constant that appears uniformly across all terms in the reduction, and can thus be factored out without impacting the $\mathcal{NP}$-hardness argument. Thus, it is easy to see that the following more general result holds:

Theorem 3. *Problem 2 is strongly $\mathcal{NP}$-hard for every integer $k \geq 3$ and rational $\beta > 1$.*

5. On the $\mathcal{NP}$-hardness of Max-(k, τ^α)-NDP

In this section, we prove the $\mathcal{NP}$-hardness of the Max-(k, τ^α)-NDP. As for the Min-$(3, 2^\tau)$-NDP, we will first show that this result holds for $k = 3$ and every rational $\alpha \geq 1$, i.e., that

Theorem 4. *The Max-$(3, \tau^\alpha)$-NDP is strongly $\mathcal{NP}$-hard for every $\alpha \geq 1$.*

We will then discuss how to generalize this result to arbitrary integers $k > 3$. We start by observing that the following result holds:

Proposition 8. *Let $n, m \geq 1$ denote two integers, and let $\alpha \geq 1$ be a rational constant. Let $\{a_i\}_{i=1}^m$ and $\{b_i\}_{i=1}^m$ denote two sets of strictly positive integers such that $2 < a_1 < \dots < a_n$ and $2 < b_1 < \dots < b_m$. Then, precisely one of the following two inequalities holds:*

$$\sum_{i=1}^n a_i^\alpha + \sum_{i=1}^m b_i^\alpha < \sum_{i=1}^n (a_i - 1)^\alpha + \sum_{i=1}^m (b_i + 1)^\alpha; \tag{1}$$

$$\sum_{i=1}^n a_i^\alpha + \sum_{i=1}^m b_i^\alpha \leq \sum_{i=1}^n (a_i + 1)^\alpha + \sum_{i=1}^m (b_i - 1)^\alpha. \tag{2}$$

Proof. For a fixed $\epsilon > 0$, consider the following polynomial real functions defined on the interval $[0, 1 + \epsilon)$:

$$f_\alpha(x) = \sum_{i=1}^n (-x + a_i)^\alpha + \sum_{i=1}^m (x + b_i)^\alpha - A_\alpha - B_\alpha \quad \text{and} \quad g_\alpha(x) = \sum_{i=1}^n (x + a_i)^\alpha + \sum_{i=1}^m (-x + b_i)^\alpha - A_\alpha - B_\alpha,$$

where $A_\alpha = \sum_{i=1}^{n} a_i^\alpha$ and $B_\alpha = \sum_{i=1}^{m} b_i^\alpha$. We shall analyze now the behavior of these functions as α, n, and m vary, to determine when the inequalities under consideration hold. We begin by observing that when $\alpha = 1$, $f_1(x) = -nx + mx = (m-n)x$ and $g_1(x) = nx - mx = (n-m)x$. Hence, (1) holds when $m < n$, while (2) holds when $m \geq n$.

Now, assume that $\alpha > 1$ and consider the derivatives of both functions with respect to x, namely

$$f'_\alpha(x) = \alpha\left(-\sum_{i=1}^{n}(-x+a_i)^{\alpha-1} + \sum_{i=1}^{m}(x+b_i)^{\alpha-1}\right) \quad \textit{and} \quad g'_\alpha(x) = \alpha\left(\sum_{i=1}^{n}(x+a_i)^{\alpha-1} - \sum_{i=1}^{m}(-x+b_i)^{\alpha-1}\right).$$

If $f'_\alpha(x_0) > 0$, for some $x_0 \in (0, 1+\epsilon)$, then $\sum_{i=1}^{m}(x_0+b_i)^{\alpha-1} > \sum_{i=1}^{n}(-x_0+a_i)^{\alpha-1}$. Alternatively, if $f'_\alpha(x_0) \leq 0$, then we have that

$$\sum_{i=1}^{n}(a_i+x_0)^{\alpha-1} > \sum_{i=1}^{n}(a_i-x_0)^{\alpha-1} \geq \sum_{i=1}^{m}(b_i+x_0)^{\alpha-1} > \sum_{i=1}^{m}(b_i-x_0)^{\alpha-1},$$

i.e., $g'_\alpha(x_0) > 0$, for some $x_0 \in (0, 1+\epsilon)$.

Moreover, as $a_1 > 2$ and $b_1 > 2$, the second derivatives of both functions, i.e.,

$$f''_\alpha(x) = \alpha(\alpha-1)(f_{\alpha-2}(x) + A_{\alpha-2} + B_{\alpha-2}) \quad \textit{and} \quad g''_\alpha(x) = \alpha(\alpha-1)(g_{\alpha-2}(x) + A_{\alpha-2} + B_{\alpha-2})$$

are always strictly positive in $(0, 1+\epsilon)$. Therefore, the convexity of $f_\alpha(x)$ implies that, if $f'_\alpha(x_0) > 0$ for some $x_0 \in (0, 1+\epsilon)$, then $f_\alpha(x)$ is strictly increasing in the whole interval, while $g_\alpha(x)$ is weakly decreasing. Conversely, the convexity of $g_\alpha(x)$ implies that, if $g'_\alpha(x_0) > 0$ for some $x_0 \in (0, 1+\epsilon)$, then $g_\alpha(x)$ is strictly increasing in the whole interval, while $f_\alpha(x)$ is weakly decreasing.

Since $f_\alpha(0) = g_\alpha(0) = 0$, we distinguish two cases depending on the monotonicity of f_α and g_α over the interval $(0, 1+\epsilon)$. If $f_\alpha(x)$ is strictly increasing on $(0, 1+\epsilon)$ and $g_\alpha(x)$ is weakly decreasing over the same interval, then $f_\alpha(x) > 0$ and $g_\alpha(x) \leq 0$ for every $x \in (0, 1+\epsilon)$. Conversely, if $g_\alpha(x)$ is strictly increasing and $f_\alpha(x)$ is weakly decreasing, then $f_\alpha(x) \leq 0$ and $g_\alpha(x) > 0$ for every $x \in (0, 1+\epsilon)$. In both cases, the statement follows for $x = 1$. □

We prove Theorem 4 by reducing the following classical $\mathcal{NP}$-hard problem to the Max-$(3, \tau^\alpha)$-NDP:

Problem 8 (The Max-Cut Problem (MCP) [23])**.** *Given a simple undirected graph $G = (V, E)$, with p vertices, for some $p \in \mathbb{Z}^+$, find a partition of V into two subsets U^* and W^* that maximizes $|\delta(U^*, W^*)|$.*

We will show that, given an optimal solution to an instance of the Max-$(3, \tau^\alpha)$-NDP, appropriately built from an instance of the MCP, it is possible to compute in polynomial time a sub-partition of the leaf-set Γ that induces an optimal solution to the given instance of the MCP. The $\mathcal{NP}$-hardness of the Max-$(3, \tau^\alpha)$-NDP, then, will immediately follows from the $\mathcal{NP}$-hardness of the MCP.

We start by considering an input graph $G = (V, E)$ of the MCP, for some fixed $p \in \mathbb{Z}^+$. Denoted by ℓ any positive integer such that $\ell > \max\{2p, \left(2p^{\alpha+2} + p^2(p+1)^\alpha\right)^{\frac{1}{\alpha}}\}$, consider a set of vertices L disjoint from V and having cardinality ℓ. We observe that, by definition, $p + \ell$ is polynomially bounded in p. Finally, denote by ω a positive integer such that $\omega > p^2(p+\ell+1)^\alpha$. Then, we build an instance of the Max-$(3, \tau^\alpha)$-NDP as follows. Set:

- $\Gamma = V \cup L \cup \{\hat{w}_1, \hat{w}_2\}$, where $\hat{w}_1$ and $\hat{w}_2$ are two distinct vertices such that $\hat{w}_1, \hat{w}_2 \notin V \cup L$;
- $d_{ij} = 1$ for all $ij \in E$;
- $d_{\hat{w}_1\hat{w}_2} = \omega$;
- and $d_{ij} = 0$ otherwise.

Let $\mathcal{T}_{(3,\tau^\alpha)}$ denote the set of caterpillars that are feasible solutions to the considered instance of the Max-$(3, \tau^\alpha)$-NDP and such that $\hat{w}_1$ indexes a leaf of one cherry and $\hat{w}_2$ indexes a leaf of the (unique) other cherry. Then, the following claim holds:

Proposition 9. *The optimal solution T^* belongs to $\mathcal{T}_{(3,\tau^\alpha)}$.*

Proof. Let T denote any feasible solution in $\mathcal{T}_{(3,\tau^\alpha)}$ and let T' denote any feasible solution not in $\mathcal{T}_{(3,\tau^\alpha)}$, i.e., T' is not a caterpillar, or it is a caterpillar but at least one between $\hat{w}_1$ and $\hat{w}_2$ does not index leaves of different cherries. Note that for a UBT on $|\Gamma|$ leaves, the maximal path-length between a pair of leaves is at most $|\Gamma|-1$. Moreover, the path-length $|\Gamma|-1$ is achieved by some pair of leaves if and only if the UBT is a caterpillar [17, 22]. Then, $\tau_{\hat{w}_1,\hat{w}_2} = p+\ell+1$ holds for T, while $\tau_{\hat{w}_1\hat{w}_2} < p+\ell+1$ holds for T', but possibly $\tau_{u,v} = p+\ell+1$, for some $u,v \in V$. Then, if z and z' are the costs of T and T' with respect to the objective function of the Max-$(3,\tau^\alpha)$-NDP, then we have that $z > 2w(p+\ell+1)^\alpha > \phi_1 = 2w[(p+\ell)^\alpha+1]$ and $z' < 2[w(p+\ell)^\alpha+|E|(p+\ell+1)^\alpha] < \phi_2 = 2[w(p+\ell)^\alpha+p^2(p+\ell+1)^\alpha]$. As $\omega > p^2(p+\ell+1)^\alpha$, $\phi_1 > \phi_2$, hence, $z > z'$. Thus, the cost of every feasible solution not belonging to $\mathcal{T}_{(3,\tau^\alpha)}$ is smaller than z and the statement follows. □

The following proposition holds:

Proposition 10. *There exists an optimal solution T^* to the Max-$(3,\tau^\alpha)$-NDP that can be partitioned into three connected components C_L, C_U, and C_W, where the leaves of C_L are indexed by L, those of C_U by $U \cup \{\hat{w}_1\}$, and those of C_W by $W \cup \{\hat{w}_2\}$, for some partition (U,W) of V.*

Proof. Denote by $z(T) = \sum_{i,j} d_{ij}\tau_{ij}^\alpha$, the cost of any feasible solution T to Max-$(3,\tau^\alpha)$-NDP, and by $\Gamma_E(\tilde{T},i)) = \{j \in \Gamma(\tilde{T}) : ij \in E\}$ for any congruent subtree $\tilde{T}$ of T.

Suppose, by contradiction, that any optimal solution $T^* \in \mathcal{T}_{(3,\tau^\alpha)}$ does not respect the statement. Then, we will show that either T^* cannot be optimal or that it is possible to generate in polynomial time a sequence of UBTs that starts from T^* and whose last element is an optimal solution to the Max-$(3,\tau^\alpha)$-NDP that respects the statement, by leading in both cases to a contradiction.

First, assume that there is a comb T_0 of T^* such that $\Gamma(T_0) = V$. Because ℓ is strictly bigger than $2p$, then for every $i \in \Gamma \setminus \{\hat{w}_1,\hat{w}_2\}$, we must have either $\tau_{\hat{w}_1,i} \geq p$, or $\tau_{\hat{w}_2,i} \geq p$, or both. Assume the first case and denote by $\hat{i}$ the leaf of T such that $\hat{w}_1$ and $\hat{i}$ form a cherry. Let $j_0 \in \Gamma(T_0)$ be any leaf of T^* such that $j_0 = \arg\min_{j\in\Gamma(T_0)}\{\tau_{j\hat{i}}\}$, and denote by T' the feasible solution obtained by exchanging $\hat{i}$ with j_0. By construction, $\tau_{j_0\hat{w}_1} \leq p$, hence $z(T') \geq z(T^*) - \sum_{j\in\Gamma_E(T_0,j_0)}\tau_{jj_0}^\alpha + \sum_{j\in\Gamma_E(T_0,j_0)}(\tau_{jj_0}+p)^\alpha > z(T^*)$, contradicting the optimality of T^*. Thus, there are at least two disjoint combs of T^* that are maximal with respect to V.

We now proof that if there exist strictly more than two disjoint combs of T^* that are maximal with respect to the subset V, then one of the following two scenarios must hold: either T^* is not an optimal solution to the problem, or it is possible to transform in polynomial time T^* into a new feasible solution where the number of maximal combs with respect to V has been reduced by one while preserving the optimality of the solution. Specifically, let T_1 be one of the combs of T^* that are maximal in V, and suppose, without loss of generality, that both connected components T_2 and T_3 of $T^* \setminus T_1$ contain at least one leaf indexed by a vertex in V. Let $i_1,\ldots,i_{m_1}$ denote the leaves of T_1, for some integer $0 < m_1 < p-1$, and assume that $\tau_{i_ji_{j+1}} = 3$ for all $j = 1,\ldots,m_1-1$; i.e., the leaves of T_1 are pairwise adjacent in the given order. Since i_1 and i_{m_1} are the extremals of T_1 and the comb is maximal in V, there must exist two leaves $\lambda_1,\lambda_2 \in L$ such that $\tau_{i_1\lambda_1} = 3$ and $\tau_{i_{m_1}\lambda_2} = 3$. Now, recall that T^* is a caterpillar, hence, there exists a total order

$$2 < \tau_{i_1a_1} < \ldots < \tau_{i_1a_{m_2}}, \tag{3}$$

where $\{a_1,\ldots,a_{m_2}\} = \Gamma_E(T_2,i_1)$. Similarly, we have another total order

$$2 < \tau_{i_1b_1} < \ldots < \tau_{i_1b_{m_3}}, \tag{4}$$

where $\{b_1,\ldots,b_{m_3}\} = \Gamma_E(T_1,i_1) \cup \Gamma_E(T_3,i_1)$; see Figure 5.

Now, consider the effect of swapping i_1 with λ_1. In this case, all path-lengths $\tau_{i_1a_j}$ for $j = 1,\ldots,m_2$ either increase or decrease by one unit, depending on the structure, while an opposite change occurs for all path-lengths associated with the leaves in $(\Gamma(T_1)\setminus\{i_1\}) \cup (\Gamma(T_3)\cap V)$.

Analogously, if we swap i_1 with i_2, the second leaf in the ordering of T_1, all values $\tau_{i_1b_j}$ either increase or decrease by one, while the opposite change is observed for the leaves in $\Gamma(T_2)\cap V$ and $(\Gamma(T_1)\setminus\{i_1,i_2\}) \cup (\Gamma(T_3)\cap V)$. Then,

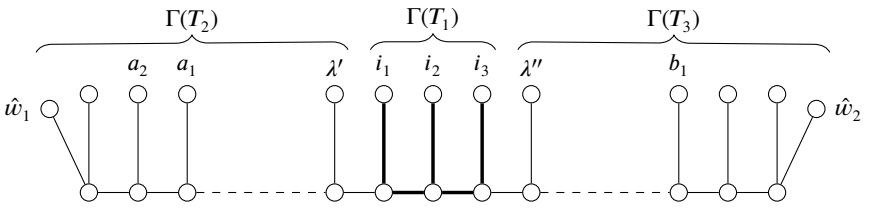

Figure 5: An example of a UBT from the family described in Proposition 10, partitioning the vertex-set V, with $|V| = 6$. In particular, $\Gamma_E(T_2, i_1) = \{a_1, a_2\}$, and $\Gamma_E(T_3, i_1) = \{b_1\}$, while bold edges represent the comb $\Gamma(T_1) = \{i_1, i_2, i_3\}$.

swapping i_1 with λ_1 yields a new feasible solution T'' in which the cost changes by an amount equal to

$$\sum_{a_j \in \Gamma_E(T_2, i_1)} (\tau_{i_1 a_j} - 1)^\alpha + \sum_{b_j \in \Gamma_E(T_1, i_1) \cup \Gamma_E(T_3, i_1)} (\tau_{i_1 b_j} + 1)^\alpha - \sum_{a_j \in \Gamma_E(T_2, i_1)} \tau^\alpha_{i_1 a_j} - \sum_{b_j \in \Gamma_E(T_1, i_1) \cup \Gamma_E(T_3, i_1)} \tau^\alpha_{i_1 b_j},$$

with respect to the cost of T^*. Similarly, swapping i_1 with i_2 yields a new feasible solution T''' in which the cost changes by an amount equal to

$$\begin{aligned}
&\sum_{a_j \in \Gamma_E(T_2, i_1)} (\tau_{i_1 a_j} + 1)^\alpha + \sum_{b_j \in \Gamma_E(T_1, i_1) \cup \Gamma_E(T_3, i_1)} (\tau_{i_1 b_j} - 1)^\alpha - \sum_{a_j \in \Gamma_E(T_2, i_1)} \tau^\alpha_{i_1 a_j} - \sum_{b_j \in \Gamma_E(T_1, i_1) \cup \Gamma_E(T_3, i_1)} \tau^\alpha_{i_1 b_j} + \\
&\sum_{a'_j \in \Gamma_E(T_2, i_2)} (\tau_{i_2 a'_j} - 1)^\alpha + \sum_{b'_j \in \Gamma_E(T_1, i_2) \cup \Gamma_E(T_3, i_2)} (\tau_{i_2 b'_j} + 1)^\alpha - \sum_{a'_j \in \Gamma_E(T_2, i_2)} \tau^\alpha_{i_2 a'_j} - \sum_{b'_j \in \Gamma_E(T_1, i_2) \cup \Gamma_E(T_3, i_2)} \tau^\alpha_{i_2 b'_j},
\end{aligned}$$

with respect to the cost of T^*, where $\{a'_1, \dots, a'_{m'_2}\} = \Gamma_E(T_2, i_2)$ and $\{b'_1, \dots, b_{m'_3}\} = \Gamma_E(T_1, i_2) \cup \Gamma_E(T_3, i_2)$. By applying Proposition 8 first to the sequences (3) and (4), and then to the sequences $a'ii = 1^{m'_2}$ and $b'ii = 1^{m'_3}$, one of the following three cases holds:

- $z(T'') > z(T^*)$;
- $z(T''') > z(T^*)$;
- $z(T'') \leq z(T^*)$ and $z(T''') = z(T^*)$.

The first two cases contradict the optimality of T^*. In the third case, fix i_1 in its current position and proceed by applying the same analysis to the swap between i_2 and i_3. If no improvement is found, we move to the next pair of vertices and we iterate this swapping approach until terminating in one of the following two scenarios:

- either the final swap between i_{m_1} and λ_2 provides an increment of the total cost, contradicting the optimality of T^*;
- or, none of the swaps modifies the total cost. In this case, we can cyclically permute the positions of these leaves without affecting the objective: indeed, each i_j takes the place of i_{j+1}, for $j = 1, \dots, m_1 - 1$, i_{m_1} takes the place of λ_2, and λ_2 fills the position originally occupied by i_1, which has been vacated by the shift. Because this operation preserves feasibility and cost, we get a new feasible solution to the problem where T_1 is still a comb, but differs from T^* for the path-lengths of the extremals with respect to the cherries.

After having performed a finite number of such swaps, we can conclude that either T^* was not optimal, or that the comb T_1 is no longer maximal in V, as its extremal i_{m_1} becomes adjacent to a leaf that differs from $i_{m_1 - 1}$ in a new comb T'_1 maximal in V, obtained by the iteration of the above cyclic permutations of the indices.

In conclusion, we can assume, without loss of generality, that the optimal tree T^* contains exactly two disjoint combs, denoted by T_U and T_W, both of which are maximal in V. Moreover, these two combs together cover all the vertices of V, i.e., $\Gamma(T_U) \cup \Gamma(T_W) = V$.

It is important to note that the optimality of T^* further implies that the connected components of the tree obtained by removing T_U and T_W from T^*, namely $T^* \setminus (T_U \cup T_W)$, consist precisely of the two singleton vertices $\{\hat{w}_1\}$ and

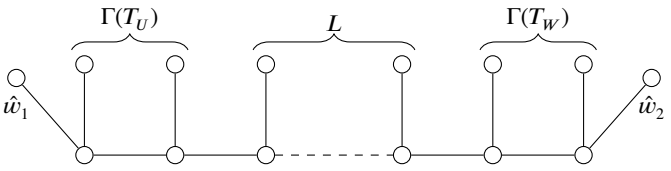

Figure 6: An example of a UBT from the family described in Proposition 10, partitioning the vertex-set V, with $|V| = 4$, into two subsets $\Gamma(T_U)$ and $\Gamma(T_W)$.

$\{\hat{w}_2\}$, as well as an additional comb component C_L, whose leaves are indexed by the set L (see Figure 6). Therefore, to finalize the construction, it suffices to define the connected components $C_U = T_U \cup \{\hat{w}_1\}$ and $C_W = T_W \cup \{\hat{w}_2\}$. These components, together with C_L, yield the required partition of T^* into three connected components with the desired leaf indexing properties. Thus, setting $U = \Gamma(T_U)$ and $W = \Gamma(T_W)$ completes the proof. □

We now show that the partition of G provided by Proposition 10 constitutes an optimal solution for the given instance of the MCP. To this end, let z^* be the cost of any solution to the Max-$(3, \tau^\alpha)$-NDP such that the cut induced by the partition of Proposition 10 is optimal for the corresponding MCP instance. Let δ^* be the cardinality of such a solution. Similarly, let z' denote the cost of any feasible solution to the $(3, \tau^\alpha)$-NDP for which the cut induced by the partition of Proposition 10 is not optimal for the MCP instance. Denote by δ' the cardinality of such a solution. Then, $z^* > \phi_1 = 2\tau^\alpha_{\hat{w}_1\hat{w}_2} + 2\delta^*\ell^\alpha$, since the contribution to the cost of $\hat{w}_1$ and $\hat{w}_2$ is constant for any feasible solution in $\mathcal{T}_{(3,\tau^\alpha)}$ by Proposition 9 and $\tau_{ij} > \ell$ for every $i \in U$ and $j \in W$ by Proposition 10. While, $z' < \phi_2 = 2\tau^\alpha_{\hat{w}_1\hat{w}_2} + 4p^{\alpha+2} + 2\delta'(\ell + p + 1)^\alpha$, where p^α is the maximal path length between any pair of leaves indexed by U or by W assuming that $G[U']$ and $G[W']$ are cliques such that $|U'| = |W'| = p - 1$. Both assumptions give trivial upper bounds for any instance in $\mathcal{T}_{(3,\tau^\alpha)}$ not corresponding to an optimal solution to the MCP for G. By definition of the input and maximum cut, $\ell > (2p^{\alpha+2} + p^2(p+1)^\alpha)^{\frac{1}{\alpha}}$ and $\delta' \leq \delta^* - 1 < p^2$, and hence $\phi_1 < \phi_2$.

To conclude, for any optimal solution z^* there is a partition of G splitting G with the most number of edges, i.e., an optimal solution to the MCP for G. Hence, solving Problem 3 is at least as hard as solving the MCP for G. □

Generalization. As in the previous section, we briefly observe here that Theorem 4 can be easily generalized to any integer $k > 3$. Specifically, in the construction of an instance of the Max-(k, τ^α)-NDP it is sufficient (i) to add the graphs $G_1, G_2, \ldots, G_{k-3}$ to G, where G_i is isomorphic to G, for every $i = 1, \ldots, k-3$; (ii) to note that for ω large enough, the optimal solution T^* is a k-ary caterpillar; and finally (iii) to note that any leaf of T^* that is indexed by a vertex i of G forms a cherry with any of the corresponding vertices $u_1, \ldots, u_{k-3}$ of $G_1, \ldots, G_{k-3}$. As a consequence, the above reduction must account for a scalar factor $k - 2$ in any term involved. This term can be factored out without impacting the $\mathcal{NP}$-hardness argument, by leading us to the following general result:

Theorem 5. *Problem 3 is strongly $\mathcal{NP}$-hard for every integer $k \geq 3$ and rational $\alpha \geq 1$.*

6. Further considerations

We present here additional complexity results for certain variants of the problems discussed in the previous sections. These results follow from straightforward adaptations of the earlier proof techniques.

On the complexity of the leaf-restricted k-NDP. The objective function of the k-NDP measures the weighted path-length between all pairs of vertices in the input graph G. In some practical applications, however (see, e.g., [24]), this function may be restricted just to a subset of vertices of G, which usually is requested to be the set of leaves of the spanning tree of G. It is straightforward to verify that the k-NDP remains $\mathcal{NP}$-hard even under this restriction.

Specifically, by reusing the same notation and input construction as in the proof for the decision version of the k-NDP, it is sufficient to observe that (i) by construction every spanning tree of G must include all edges in $E(V_R \cup V_S \cup V_r)$, hence the cost $L_T(R)$ is identical across all feasible solutions; (ii) for sufficiently large r, any optimal solution to the

problem must belong to $\mathcal{T}_k$, since for any feasible solution $T' \notin \mathcal{T}_k$, there exist vertices $u \in V_S$ and $v \in R$ such that $L_{T'}(u, v) = 4$; and (iii) for any $T \in \mathcal{T}_k$, the quantity $L_T(\Gamma(T) \setminus (R \cup M))$ is constant, as $|\Gamma(T)| = |R| + |S| - \ell + |M|$. It follows that the $\mathcal{NP}$-completeness result extends to the considered restriction of the k-NDP, and the following result holds:

Theorem 6. *Given a simple connected undirected graph $G = (V, E)$, a constant $k \in \mathbb{Z}^+$ and a weight function $w : E \to \mathbb{Q}_0^+$. Find a spanning UkT T of G that minimizes $\sum_{i,j \in \Gamma(T)} L_T(i, j)$ is strongly $\mathcal{NP}$-hard.*

On the complexity of the rainbow version of the k-NDP. The input graph G in the k-NDP can be slightly modified to accommodate additional constraints on the assignment of weights and/or vertex degrees within the spanning tree, requiring that these quantities range over and/or cover a prescribed set of values. Such constraints arise naturally in practical settings, e.g., when a contractor or a supplier imposes specific usage requirements on available components, or when all items must be utilized due to budgetary, regulatory, or logistical constraints. We refer to these constrained variants as the *rainbow versions* of the k-NDP, and we briefly discuss them below.

Let $\Delta(T)$ denote the set of degrees of the internal vertices of a tree T, and consider the input construction as in the proof for the decision version of the k-NDP. Since $w(e) = 0$ for every edge $e \in E(V_r \cup V_s)$, it is possible to contract some of these edges without affecting the objective function value. As a result, we can assume, without loss of generality, that any spanning tree of G satisfies $\Delta(T) = N$, for any prescribed set $N \subseteq \mathbb{Z}^+$. We refer to this variant as the *rainbow degrees NDP*. This leads to the following result:

Theorem 7. *Given a simple connected undirected graph $G = (V, E)$, a finite set $N \subseteq \mathbb{Z}^+$, and a weight function $w : E \to \mathbb{Q}_0^+$, find a spanning tree T of G, with $\Delta(T) = N$, that minimizes $\sum_{u,v \in V} L_T(u, v)$ is strongly $\mathcal{NP}$-hard.*

The weight vector w, defined over the edges of the graph G in the proof of Theorem 1, takes values in $\{0, 1\}$. A natural question is whether this setting can be generalized by allowing edge weights to range over a given set $W = \{w_1, \dots, w_q\} \subseteq \mathbb{Q}^+$, with $w_1 < \cdots < w_q$, such that the function $w : E \to W$ is surjective, i.e., each value in W is assigned to at least one edge. We refer to this variant as the *rainbow costs NDP*. The output of this problem consists of a pair constituted by a spanning k-ary tree T of G, and a surjective weight function $w : E(T) \to W$. Let $W(T)$ denote the set of edge weights used in the tree T. Then, the following observations hold:

(i) All pairwise costs $L_T(u, v)$ are polynomially bounded in $\max W(T)$, regardless of the specific choice of T.

(ii) When minimizing the total cost $\sum_{u,v \in V} L_T(u, v)$, an optimal weight assignment w will always assign the smallest available weights to as many edges as possible.

(iii) Any optimal solution (T^*, w^*) admits a partition of the edge set $E(T^*)$ into subsets $(F_1, \dots, F_q)$ such that:

- $|F_1| = |E(T^*)| - q + 1$;
- $|F_i| = 1$ for each $i = 2, \dots, q$;
- $w^*(e) = w_1$ for all $e \in F_1$;
- $w^*(e) = w_i$ for $e \in F_i$, $i = 2, \dots, q$.

Moreover, the selection of edges in F_1 can be made arbitrarily, since the objective function is linear and each edge contributes to a constant number of vertex-to-vertex paths in any feasible tree. Therefore, for our specific instance graph G, we may assume without loss of generality that $w(e) = w_1$ for all edges except for $q - 1$ edges in $\delta(R)$, which can be assigned weights in $W \setminus \{w_1\}$ via a one-to-one correspondence.

Thus, the following result holds:

Theorem 8. *Given a simple connected undirected graph $G = (V, E)$, a constant $k \in \mathbb{Z}^+$, and a finite set $W \subseteq \mathbb{Q}^+$, find a spanning UkT T of G, with $W(T) = W$, that minimizes $\sum_{u,v \in V} L_T(u, v)$ is strongly $\mathcal{NP}$-hard.*

On the complexity of the fixed-topology Min-(k, β^τ)-NDP and Max-(k, τ^α)-NDP. In the proofs of Theorems 2 and 4, respectively, we were able to fix the topology of the trees that are candidates for being optimal solutions by assuming that γ and ω are sufficiently large. Thus, without invoking Propositions 6 and 9, respectively, the reductions used for the Min-(k, β^τ)-NDP and the Max-(k, τ^α)-NDP remain valid even when we restrict feasible solutions to share the same topology, by leading to the following results:

Theorem 9. *Consider two positive integers n and k, with $n > k \geq 3$, an unlabeled UkT T with n leaves, and a $n \times n$ symmetric distance matrix $\mathbf{D} = \{d_{ij}\}$ associated with the vertices in Γ of G_n^k. Then, finding an assignment of the vertices in Γ to the leaves of T so as to minimize $\sum_{i,j\in\Gamma} d_{ij}\beta^{\tau_{ij}}$ is strongly $\mathcal{NP}$-hard.*

Theorem 10. *Consider two positive integers n and k, with $n > k \geq 3$, an unlabeled UkT T with n leaves, and a $n \times n$ symmetric distance matrix $\mathbf{D} = \{d_{ij}\}$ associated with the vertices in Γ of G_n^k. Then, finding an assignment of the vertices in Γ to the leaves of T so as to maximize $\sum_{i,j\in\Gamma} d_{ij}\tau_{ij}^\alpha$ is strongly $\mathcal{NP}$-hard.*

On the complexity of the Min-(k, τ^α)-NDP The reduction presented in Section 4 does not extend to the case in which $\beta^{\tau_{ij}}$ is replaced by τ_{ij}^α for some rational constant $\alpha \geq 1$, i.e., to the Min-(k, τ^α)-NDP. In fact, the depth of a perfect k-ary tree grows only logarithmically with the input size, and a direct computation shows that the parameter m used in the proof of Theorem 2 must grow faster than any polynomial in order to guarantee the optimality of the partition (U, W) for the instance G of the GBP.

On the complexity of the Max-(k, β^τ)-NDP. The reduction described in Section 5 does not extend to the case in which τ_{ij}^α is replaced by $\beta^{\tau_{ij}}$ for some constant $\beta > 0$, i.e., to the Max-(k, β^τ)-NDP. In fact, in this setting the input size $n = \ell + p + 2$ becomes exponential, since ℓ must grow faster than any polynomial in order to ensure the optimality of the partition (U, W) in the corresponding instance of the reduction for the MCP with input graph G.

To the best of our knowledge, the bounds achievable in the two cases above are not sufficient to ensure that the size of the corresponding instances remains polynomial. Thus, the computational complexity of these last two problems remains an open question.

Acknowledgements

The first and the fourth authors acknowledge support from the Belgian National Fund for Scientific Research (FNRS) via the grant FNRS PDR 40007831. The first author also acknowledges support from the Fondation Louvain, via the grant "COALESCENS".

References

[1] D. S. Johnson, J. K. Lenstra, A. H. G. R. Kan, The complexity of the network design problem, Networks 8 (1978) 279–285.
[2] P. C. Pop, Generalized network design problems: Modeling and optimization, De Gruyter, Berlin, Germany, 2012.
[3] R. K. Ahuja, T. L. Magnanti, J. B. Orlin, Network Flows: Theory, Algorithms, and Applications, Prentice Hall, United States edition, 1993.
[4] D. Z. Du, X. Hu, Steiner tree problems in computer communication networks, World Scientific Publishing Company, Singapore, 2008.
[5] P. Crescenzi, V. Kann, A compendium of NP optimization problems, https://www.csc.kth.se/~viggo/wwwcompendium/node69.html (2000).
[6] R. Ravi, R. Sundaram, M. V. Marathe, D. J. Rosenkrantz, S. S. Ravi, Spanning trees short or small, SIAM Journal on Discrete Mathematics 9 (2) (1996) 178–200.
[7] G. Galbiati, F. Maffioli, A. Morzenti, A short note on the approximability of the maximum leaves spanning tree problem, Information Processing Letters 52 (1) (1994) 45–49.
[8] P. Crescenzi, V. Kann, R. Silvestri, l. L. Trevisan, Structure in approximation classes, SIAM Journal on Computing 28 (5) (1999) 1759–1782.
[9] N. Garg, A 3-approximation for the minimum tree spanning k vertices, in: Proceedings of 37th Conference on Foundations of Computer Science, Burlington, VT, 1996, pp. 302–309.
[10] H. Lu, R. Ravi, The power of local optimization: Approximation algorithms for maximum-leaf spanning tree, in: Proceedings of the 38th Annual Allerton Conference on Communication, Control and Computing, Pasadena, CA, 2000.
[11] M. R. Garey, D. S. Johnson, Computers and Intractability: A guide to the theory of NP-Completeness, Freeman, New York, NY, 2003.
[12] S. Fiorini, G. Joret, Approximating the balanced minimum evolution problem, Operations Research Letters 40 (1) (2012) 31–35.
[13] M. Frohn, On the approximability of the fixed-tree balanced minimum evolution problem, Optimization Letters 15 (6) (2021) 2321–2329.
[14] O. Gascuel, Mathematics of evolution and phylogeny, Oxford University Press, New York, NY, 2005.
[15] G. Gan, C. Ma, J. Wu, Data clustering: Theory, algorithms, and applications, SIAM, 2007.

[16] D. Catanzaro, M. Labbé, R. Pesenti, The balanced minimum evolution problem under uncertain data, Discrete Applied Mathematics 161 (13–14) (2013) 1789–1804.
[17] D. Catanzaro, R. Pesenti, L. A. Wolsey, On the Balanced Minimum Evolution polytope, Discrete Optimization 36 (2020) 1–33.
[18] D. Catanzaro, M. Frohn, R. Pesenti, An information theory perspective on the balanced minimum evolution problem, Operations Research Letters 48 (3) (2020) 362–367.
[19] D. Catanzaro, M. Frohn, O. Gascuel, R. Pesenti, A tutorial on the balanced minimum evolution problem, European Journal of Operational Research 300 (1) (2022) 1–19.
[20] D. Catanzaro, R. Pesenti, A. Sapucaia, L. Wolsey, Optimizing over path-length matrices of unrooted binary trees, Mathematical Programming, in press, (2025).
[21] D. Reinhard, Graph Theory, Springer-Verlag, New York, NY, 2005.
[22] D. Stott-Parker, P. Ram, The construction of Huffman codes is a submodular ("convex") optimization problem over a lattice of binary trees, SIAM Journal on Computing 28 (5) (1996) 1875–1905.
[23] M. R. Garey, D. S. Johnson, L. Stockmeyer, Some simplified NP-complete problems, in: Proceedings of the sixth annual ACM symposium on Theory of computing, 1974, pp. 47–63.
[24] D. Catanzaro, The minimum evolution problem: Overview and classification, Networks 53 (2) (2009) 112–125.