Approximation Algorithms for Union and Intersection Covering Problems

Marek Cygan[1], Fabrizio Grandoni[2], Stefano Leonardi[3], Marcin Mucha[1], Marcin Pilipczuk[1], and Piotr Sankowski[1]

1 **Institute of Informatics, University of Warsaw, Poland.** {cygan,mucha,malcin,sank}@mimuw.edu.pl
2 **Computer Science Department, University of Rome Tor Vergata, Roma, Italy.** grandoni@disp.uniroma2.it
3 **Department of Computer and System Science, Sapienza University of Rome, Italy.** stefano.leonardi@dis.uniroma1.it

Abstract

In a classical covering problem, we are given a set of *requests* that we need to satisfy (fully or partially), by buying a subset of *items* at minimum cost. For example, in the k-MST problem we want to find the cheapest tree spanning at least k nodes of an edge-weighted graph. Here, nodes represent requests whereas edges correspond to items.

In this paper, we initiate the study of a new family of *multi-layer* covering problems. Each such problem consists of a collection of h distinct instances of a standard covering problem (*layers*), with the constraint that all layers share the same set of requests. We identify two main subfamilies of these problems:

- in an UNION multi-layer problem, a request is satisfied if it is satisfied *in at least one* layer;
- in an INTERSECTION multi-layer problem, a request is satisfied if it is satisfied *in all* layers.

To see some natural applications, consider both generalizations of k-MST. UNION k-MST can model a problem where we are asked to connect a set of users to at least one of two communication networks, e.g., a wireless and a wired network. On the other hand, INTERSECTION k-MST can formalize the problem of providing both electricity and water to at least k users.

1998 ACM Subject Classification F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases Approximation algorithms, Partial covering problems

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2011.28

1 Introduction

In the fundamental MINIMUM SPANNING TREE problem (MST), the goal is to compute the cheapest tree which spans all the n nodes of a given edge-weighted graph $G = (V, E)$. To handle the subtleties of real-life applications, several natural generalisations and variants of the problem have been considered. For example, in the STEINER TREE problem we need to connect only a given subset W of k *terminal* nodes. In the k-MST problem instead, the goal is to connect at least k (arbitrary) nodes. One common feature of these generalizations is that we need to design a single network. However, this is often not the case in the applications. For example, suppose we want to provide at least k out of n users with both electricity and water. In this case, we cannot design the water and electricity infrastructures independently: our decisions on which users to reach have to be synchronized.

Consider now another classic problem, the TRAVELLING SALESMAN problem (TSP): here we are given a complete weighted graph, and the goal is to compute the minimum-length

31st Int'l Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2011).
Editors: Supratik Chakraborty, Amit Kumar; pp. 28–40
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

tour traversing all the nodes. Again, several natural generalizations and variants of the problem have been considered in the literature. Still, all of them deal only with the case where there is a single network. However, there are natural applications which do not fit in this framework. For example, suppose you want to visit a set of places (bank, post office, etc.), and you can use your bike and your car. Of course, you cannot just reach a place by bike, and then suddenly switch to your car (that you left at home). Your trip must consist of a tour by bike and another tour by car, which together touch all the places that you need to visit.

The above examples show the need for a new framework, which is able to capture coordinated decision-making over multiple optimization problems. In this paper we initiate the study of such *multi-layer covering* problems. These problems are characterized by a set of h instances of a standard covering problem (*layers*), sharing a common set of n *requests*. The goal is to *satisfy*, possibly partially, the requests by buying *items* in each layer at minimum total cost. We identify two main families of such problems:

- Intersection problems. Here, as in the water-electricity example, a request is satisfied if it is satisfied *in all* the layers.
- Union problems. Here, as in the car-bike example, a request is satisfied if it is satisfied *in at least* one layer.

1.1 Our results.

We provide hardness and approximation results for the union and intersection versions of several classical covering problems: MST, Steiner Tree, (Nonmetric and Metric) Facility Location, TSP, and Set Cover. (Formal definitions are given at the end of this section). We focus on the partial covering variant of these problems, i.e., k-MST, k-Steiner Tree, etc.: here wet need to satisfy a *target* number k of the n requests. This allows us to handle a wider spectrum of interesting problems. In fact, for intersection problems, if $k = n$ it is sufficient to compute an independent solution for each layer. On the other hand, some of the union problems above are interesting also for the case $k = n$. However, the results that we achieve for that case are qualitatively the same as for $k < n$.

For Intersection versions of k-MST, k-Steiner Tree, k-TSP, k-Set Cover, k-Metric Facility Location, and k-Nonmetric Facility Location, we show that:

- Even for two layers, a polylogarithmic approximation for these problems would imply a polylogarithmic approximation for k-Densest Subgraph. We recall that the best approximation for the latter problem is $O(n^{\frac{1}{4}+\varepsilon})$ [6] and finding a polylogarithmic approximation is a major open problem. Indeed, many researchers believe that a polylogarithmic approximation does not exist, and exploit this assumption in their hardness reductions (see, e.g., [2, 3]).
- On the positive side, we give $\tilde{O}(k^{1-1/h})$-approximation algorithms[1] for these problems.

Note that, in the single-layer case, the above problems can be approximated within a constant or logarithmic factor. Hence, our results show that the complexity of natural intersection problems changes drastically from one to two layers.

For Union versions of k-MST, k-Steiner Tree, k-TSP and k-Metric Facility Location we show that:

- The problems are $\Omega(\log k)$-hard to approximate for an unbounded number h of layers. Furthermore, there is a greedy $O(\log k)$-approximation algorithm. For the first three

[1] The $\tilde{O}$ notation suppresses polylogarithmic factors.

problems this only holds for the rooted version — the unrooted case is inapproximable (i.e., any bounded approximation factor implies $P = NP$).

- There is an LP-based algorithmic framework which provides $O(h)$-approximate solutions. Furthermore, the natural LPs involved have $\Omega(h)$ integrality gap.

We remark that Union k-Set Cover and Union k-Nonmetric Facility Location can be solved by collapsing all layers into one, and hence they are less interesting.

1.2 Related Work.

To the best of our knowledge, and somewhat surprisingly, approximation algorithms for union and intersection problems seem to not have been studied much in the literature. The notable exception is Matroid Intersection, which however is solvable in polynomial time [7]. Some team formation games in social networks can be seen as special cases of our intersection problems [1, 21]. Riaz et al. [25] observe that traditionally, wired and wireless infrastructures have been planned separately, but there is a need for complementary use of wired and wireless technologies in future networks. However, the authors do not consider the optimization aspects of such planning, which is captured by our Union k-MST.

The idea of introducing multiple cost functions into one optimization problem is the main theme of *multi-objective optimization*. Standard and multi-criteria approximation algorithms have been developed for the multi-objective version of several classical problems, such as Shortest Path, Spanning Tree, Matching, etc. (see, e.g., [4, 14, 15, 22, 23]). One could view these problems as having several layers with different costs. However, in contrary to our setting, solutions in different layers have to be *exactly the same*.

Partial covering problems (also known as problems with *outliers*), are well-studied in the literature: e.g., k-MST [12], k-TSP [12], k-Metric Facility Location [19], and k-Set Cover [26]. Their generalization on multiple layers is significantly harder, as our results show. Note that our Union k-Steiner Tree problem generalizes all of the following problems: k-Steiner Tree (and hence k-MST), Prize-Collecting Steiner Tree (see the proof of Theorem 7), and k-Set Cover (see the proof of Theorem 8).

Rent-or-buy and *buy-at-bulk* problems [8, 9, 13, 16] can be seen as multi-layer union problems where edge weights in different layers are related by a multiplicative factor. In contrast, weights of different layers are unrelated in our framework.

Recently Krishnaswamy et al. [20] considered a *matroid median* problem, where the set of open centers must form an independent set in a matroid. We can interpret this as a generalization of a union problem, where we add side constraints between bought items. However, the authors assume that all layers share the same metric space.

1.3 Preliminaries.

In *covering* problems we are given a set $\mathcal{U}$ of n *requests*, and a set $\mathcal{S}$ of *items*, with costs $w : \mathcal{S} \to \mathbb{R}_{\geq 0}$. The goal is to satisfy all requests by selecting a subset of items at minimum cost. We already defined MST, Steiner Tree, and TSP. Here, nodes and edges represent requests and items (with costs $w : E \to \mathbb{R}_{\geq 0}$), respectively. In the Set Cover problem, requests are the elements of a universe $\mathcal{U}$, and items $\mathcal{S}$ are subsets $S_1, \ldots, S_m$ of $\mathcal{U}$. Any S_i satisfies all the $v \in S_i$. Nonmetric Facility Location is a generalization of Set Cover, where we are given a set $\mathcal{F}$ of *facilities*, with opening costs $o : \mathcal{F} \to \mathbb{R}_{\geq 0}$, and a set $\mathcal{C}$ of *clients*, with connection costs $w : \mathcal{C} \times \mathcal{F} \to \mathbb{R}_{\geq 0}$. The goal is to compute a subset $\mathcal{A}$ of *open* facilities such that $\sum_{f \in \mathcal{A}} o(f) + \sum_{c \in \mathcal{C}} w(c, \mathcal{A})$ is minimized. Here, $w(c, \mathcal{A}) := \min_{f \in \mathcal{A}} w(c, f)$. We also say that c is connected to (or served by) $\mathcal{A}(c) := \arg\min_{f \in \mathcal{A}} w(c, f)$. If connection

costs satisfy triangle inequality, the problem is called Metric Facility Location. We can naturally define partial covering versions for the above problems, i.e., k-MST, k-Steiner Tree, k-TSP, k-Nonmetric Facility Location, and k-Metric Facility Location[2].

It is straightforward to define union and intersection versions of the above problems (more details in the corresponding sections). In the rest of this paper, the number of layers is denoted by h, and variables associated to layer i have an apex i (e.g., w^i, o^i, etc.), whereas OPT denotes the optimum solution, and opt its cost. By N we denote the total number of requests and items (in all layers).

By standard reductions, a ρ-approximate algorithm for the k-MST problem implies a 2ρ-approximate algorithm for k-Steiner Tree and k-TSP. Moreover, a ρ-approximation for k-TSP gives a 2ρ-approximation for k-MST. Essentially, the same reductions extend to the union and intersection versions of these problems. For this reason, in the rest of this paper we will consider the union and intersection version of k-MST only.

Proofs and details which are omitted due to lack of space will be given in the full version of the paper.

2 Intersection Problems

In this section we present our main results on the intersection problems. In Intersection k-Set Cover we are given h collections $\mathcal{S}^1, \mathcal{S}^2, \ldots, \mathcal{S}^h$ of subsets of a given universe $\mathcal{U}$, where $w^i : \mathcal{S}^i \to \mathbb{R}_{\geq 0}$ is the cost of subsets in the i'th collection. The goal is to cover at least k elements in all layers simultaneously, at minimum total cost. In Intersection k-MST we are given a complete graph $G = (V, E)$ on n nodes, and h metric edge-weight functions $w^1, \ldots, w^h$. The goal is to compute a tree T^i for each layer such that $\sum_i w^i(T^i)$ is minimized and $|\bigcap_i V(T^i)| \geq k$. In Intersection k-Nonmetric Facility Location we are given a set $\mathcal{F}$ of *facilities*, with opening costs $o^i : \mathcal{F} \to \mathbb{R}_{\geq 0}$ on layer i, and a set $\mathcal{C}$ of *clients*, with connection costs $w^i : \mathcal{C} \times \mathcal{F} \to \mathbb{R}_{\geq 0}$ on layer i. The goal is to compute a subset $\mathcal{A}^i$ of *open* facilities on each layer i and a subset $\mathcal{C}'$ of k clients such that $\sum_i (\sum_{f \in \mathcal{A}^i} o^i(f) + \sum_{c \in \mathcal{C}'} w^i(c, \mathcal{A}^i))$ is minimized. Here $w^i(c, \mathcal{A}^i) := \min_{f \in \mathcal{A}^i} \{w^i(c, f)\}$. Intersection k-Metric Facility Location is the special case of Intersection k-Nonmetric Facility Location where connection costs satisfy triangle inequality.

2.1 Hardness

In order to show the hardness of our problems, we use reductions from the k-Densest Subgraph problem: find the induced subgraph on k nodes with the largest possible number of edges. The fact that partial coverage problems can be as hard as k-Densest Subgraph is already known. Hajiaghayi and Jain [18] use k-Densest Subgraph to show that a partial coverage version of the Steiner Forest problem has no polylogarithmic approximation. In particular they introduce the Minimum ℓ-Edge Coverage problem where one is to find the minimum number of vertices in a graph, whose induced subgraph has at least ℓ edges. Moreover Hajiaghayi and Jain show a relation between approximation ratios for k-Densest Subgraph and Minimum ℓ-Edge Coverage.

[2] In the literature k-Nonmetric Facility Location often means that we are allowed to open at most k facilities, while here we mean that we need to connect at least k clients. Similarly for k-Metric Facility Location. Sometimes k-Set Cover indicates a Set Cover instance where the largest cardinality of a set is k, while our problem is sometimes called Partial Set Cover.

Figure 1 Approximation algorithm for 2-layer Intersection k-Set Cover. For $a \in \{1,2\}$, $\overline{a}$ is the other value in $\{1,2\}$

```
procedure SCI(k, U, S^1, S^2, w^1, w^2)
    K ← ∅, A^1 ← ∅, A^2 ← ∅
    repeat
        for a=1 to 2 do
            for all X ∈ S^a do
                for b := 1 to min(k − |K|, |X \ K|) do
                    Solve one-layer Intersection k-Set Cover problem on layer ā
                    with universe X \ K and target b.
        Let (a', b', X') be the loop iterators which provide a solution (K', A')
        minimizing the ratio of cost C' to number b' of covered elements.
        K ← K ∪ K', A^{a'} ← A^{a'} ∪ {X'}, A^{ā'} ← A^{ā'} ∪ A'
    until |K| = k
    return (K, A^1, A^2)
```

In order to simplify our reductions we extend the result on Minimum ℓ-Edge Coverage to bipartite graphs. In particular, we are able to show that a $f(n)$-approximation for Minimum ℓ-Edge Coverage on bipartite graphs implies a $16(f(2n))^2$-approximation for k-Densest Subgraph on arbitrary graphs. One can naturally reduce a Minimum ℓ-Edge Coverage instance (G, ℓ) on a bipartite graph $G = (V_1 \cup V_2, e)$ to Intersection k-Set Cover (with $k = \ell$), by mapping edges into elements and each node v into the corresponding subset of incident edges $\delta(v) \subseteq E$. A similar reduction works for Intersection k-Metric Facility Location, where facility opening costs are set to 1 and distances corresponding to edges and anti-edges are set to 0 and $+\infty$, respectively. For Intersection k-MST we construct one layer (symmetrically for the other one) by connecting a dummy root node r with all nodes in V_1 at cost 1, and each $v \in V_1$ with $\delta(v)$ at cost 0. As a consequence, we obtain the following two theorems, which suggest that the existence of a polylogarithmic approximation for the considered problems is rather unlikely (or at least hard to achieve).

▶ **Theorem 1.** *If there exists an $f(n)$-approximation algorithm for unweighted* Intersection k-Set Cover *on two layers or for* Intersection k-Metric Facility Location *on two layers, then there exists a $16(f(2m))^2$-approximation algorithm for k-*Densest Subgraph.

▶ **Theorem 2.** *If there exists an $f(n)$-approximation algorithm for* Intersection k-MST *on two layers, then there exists a $16(f(2n+2m+2))^2$-approximation algorithm for k-*Densest Subgraph.

2.2 Intersection k-Set Cover

The basic idea behind our Intersection k-Set Cover algorithm is as follows. We consider any set X in any layer, and any number $j \leq k$ of elements in X. We solve recursively, on the remaining layers, the intersection problem induced by X with target j. The base of the induction is obtained by solving a one-layer Intersection k-Set Cover problem, using the greedy algorithm which provides a $(1 + \ln k)$-approximation [26]. We choose the set X and the cardinality j for which we obtain the best ratio of cost to number of covered elements. Next, we include covered elements in the solution under construction, and the problem is reduced consequently.

In order to highlight the main ideas of our approach, we focus on the special case $h = 2$, and we neglect polylogarithmic factors in the analysis.

▶ **Theorem 3.** *There is a $\tilde{O}(\sqrt{k})$-approximation algorithm for* Intersection k-Set Cover *on two layers.*

Proof. Consider the algorithm in Figure 1. Its running time is polynomial, since `SCI` procedure calls the one-layer greedy algorithm $O(Nk^2)$ times.

Let $(\mathcal{O}^1, \mathcal{O}^2) \subseteq \mathcal{S}^1 \times \mathcal{S}^2$ be the optimal solution, and let $K_{\mathcal{O}} \subseteq (\cup_{S \in \mathcal{O}^1} S) \cap (\cup_{S \in \mathcal{O}^2} S)$ be any set of k elements in the intersection. For each element $x \in K_{\mathcal{O}}$ and layer $i = 1, 2$, let us fix a set $\mathcal{O}^i(x) \in \mathcal{O}^i$ that covers x. We prove that at each iteration of the main loop $C'/b' = \tilde{O}(opt/\sqrt{k - |K|})$. This implies that the total cost of the constructed solution is bounded by $\sum_{i=0}^{k-1} \tilde{O}(opt/\sqrt{k-i}) = opt \cdot \tilde{O}(\sqrt{k})$.

Let $\kappa := \sqrt{k - |K|}$. We consider two cases, depending on whether there exists a set X in the optimal solution that covers at least κ elements of $K_{\mathcal{O}} \setminus K$.

Case 1. Assume that there exists $1 \le a \le 2$ and $X \in \mathcal{O}^a$, such that for at least κ elements x of $K_{\mathcal{O}} \setminus K$ we have $\mathcal{O}^a(x) = X$. Let us focus on the moment when our algorithm considers taking the set X. Obviously we have $\kappa \le k - |K|$, therefore our algorithm considers covering $b := \kappa$ elements of X. As the optimal solution does it, it may be done with cost opt, so the call to the one layer algorithm returns a solution with cost $\tilde{O}(opt)$. Hence we have $C'/b' = \tilde{O}(opt/\sqrt{k - |K|})$.

Case 2. For each $1 \le a \le 2$ and every $X \in \mathcal{O}^a$, at most $\kappa - 1$ elements of $K_{\mathcal{O}} \setminus K$ satisfy $\mathcal{O}^a(x) = X$. For each $x \in K_{\mathcal{O}} \setminus K$, let $w(x) := w^1(\mathcal{O}^1(x)) + w^2(\mathcal{O}^2(x))$ be the sum of the costs of sets covering x in the optimal solution. We have

$$\sum_{x \in K_{\mathcal{O}} \setminus K} w(x) = \sum_{a=1}^{2} \sum_{X \in \mathcal{O}^a} \sum_{x \in K_{\mathcal{O}} \setminus K:\ \mathcal{O}^a(x) = X} w^a(\mathcal{O}^a(x)) \le \sum_{a=1}^{2} \sum_{X \in \mathcal{O}^a} w^a(X)\kappa \le \kappa \cdot opt.$$

Thus there exists $x_0 \in K_{\mathcal{O}} \setminus K$ such that $w(x_0) \le \kappa \cdot opt/|K_{\mathcal{O}} \setminus K|$. If we take any a and consider the iteration with $X = \mathcal{O}^a(x_0)$ and $b = 1$, the algorithm computes a set of minimum cost $C_0 \le w(x_0)$ covering x_0. We can conclude that

$$\frac{C'}{b'} \le C_0 \le \frac{\kappa \cdot opt}{|K_{\mathcal{O}} \setminus K|} = \tilde{O}(opt/\sqrt{k - |K|}).$$

◀

It is easy, just more technical, to extend the same approach to $h > 2$ and to refine (slightly) the approximation factor. With some more work, our approach generalizes to Intersection k-Nonmetric Facility Location. The details of this generalization will be given in the full version of this paper.

▶ **Theorem 4.** *There exists a $(4k^{1-1/h} \log^{1/h}(k))$-approximation algorithm for* Intersection k-Nonmetric Facility Location *(hence for* Intersection k-Set Cover*) running in $N^{O(h)}$ time.*

2.3 Intersection k-MST

Our Intersection k-MST algorithm works as follows. We consider a new metric w defined as $w(e) := \sum_i w^i(e)$ for each $e \in E$, and compute a 2-approximate solution of the resulting (one-layer) k-MST problem using the algorithm in [12].

▶ **Lemma 5.** *Let $K \subseteq V$, and $w^i(K)$ denote the cost of a minimum spanning tree of K on layer i. Then there exist two nodes $u, v \in K$ such that $w^i(u, v) \le 4w^i(K)/|K|^{1/h}$ for $i = 1, \ldots, h$.*

Proof. Let us prove the following claim by induction on i: for any $i \in \{0, \dots, h-1\}$, there exist a nodeset $K_i \subseteq K$ and paths $P_i^1, P_i^2 \dots, P_i^i$ on K_i such that: (a) $|K_i| \geq |K|^{1-i/h}$ and (b) $w^j(P_i^j) \leq 2w^j(K)/|K|^{1/h}$ for $j = 1, \dots, i$. Trivially $K_0 = K$ satisfies the claim, hence assume $i > 0$. Let T^i be the minimum spanning tree of K on layer i. Duplicate its edges, compute an Euler tour, and shortcut duplicated nodes. Let C^i be the resulting cycle on K of length at most $2w^i(K)$. Remove up to $|K|^{1/h}$ edges from C^i so as to obtain $|K|^{1/h}$ segments of length at most $2w^i(K)/|K|^{1/h}$ each. Let P be the segment maximizing the cardinality of $K_i := V(P) \cap K_{i-1}$. Set K_i satisfies (a) since $|K_i| \geq |K_{i-1}|/|K|^{1/h} \geq |K|^{1-(i-1)/h-1/h}$. The paths P_i^i and P_i^j, $j < i$, satisfying (b) are obtained from P and P_{i-1}^j, respectively, by shortcutting the nodes not in K_i.

Similarly as above, we can split C^h into $|K|^{1/h}/2$ segments which span K and have length at most $4w^h(K)/|K|^{1/h}$ each. At least one of these segments contains $2|K_{h-1}|/|K|^{1/h} \geq 2$ nodes of K_{h-1}. Thus there are two nodes u and v such that $w^i(u,v) \leq 4w^i(K)/|K|^{1/h}$ for $i = 1, \dots, h$. ◀

▶ **Theorem 6.** *The* Intersection k-MST *algorithm above is* $16k^{1-1/h}$*-approximate.*

Proof. Consider the following process: starting with the optimal set $K_{\mathcal{O}}$ of k covered nodes, we iteratively take the edge $\{x, y\}$ guaranteed by Lemma 5 and contract it in all layers, until $K_{\mathcal{O}}$ collapses into a single node. The contracted edges form a tree T' (same for all layers) spanning k nodes, of cost

$$w(T') \leq 4 \sum_{i=1}^{h} w^i(K_{\mathcal{O}}) \sum_{i=1}^{k-1} (k-i+1)^{-\frac{1}{h}} \leq \; 8k^{1-\frac{1}{h}} \sum_{i=1}^{h} w^i(K_{\mathcal{O}}) = 8k^{1-\frac{1}{h}} opt.$$

The algorithm returns a solution of cost at most $2w(T')$. The claim follows. ◀

Via a non-trivial construction we can show that the approximation factor of our algorithm is $\Omega(\frac{1}{h}k^{1-1/h})$. The details will be given in the full version of this paper.

3 Union Problems

In this section we present our results for Union k-MST and Union k-Metric Facility Location. In (unrooted) Union k-MST we have the same input and output as in Intersection k-MST, but here the trees T^i must satisfy $|\bigcup_i V(T^i)| \geq k$. In the rooted version of the problem, we are also given a root r^i for each layer i with the constraint $r^i \in T^i$. The Union MST problem is the special case of Union k-MST with $k = n$ (rooted and unrooted versions are equivalent). In Union k-Metric Facility Location we have the same input and output as in Intersection k-Metric Facility Location, but the objective function to be minimized is $\sum_i \sum_{f \in \mathcal{A}^i} o^i(f) + \sum_{c \in \mathcal{C}'} \min_i \{w^i(c, \mathcal{A}^i)\}$.

3.1 Rooted Union k-MST

▶ **Theorem 7.** *Rooted* Union k-MST *and* Union k-Metric Facility Location *are APX-hard for any* $h \geq 1$. Union MST *is APX-hard for any* $h \geq 2$.

Proof. The first claim trivially follows from the APX-hardness [12, 17] of the considered problems for $h = 1$, by adding dummy layers with infinite edge weights.

For the second claim, we consider a reduction from the APX-hard [5] Prize-Collecting Steiner Tree problem: given an undirected graph $G = (V, E)$, edge weights $w : E \to \mathbb{R}_{\geq 0}$, a root node $r \in V$, and node prizes $p : V \to \mathbb{R}_{\geq 0}$, find a tree $T \ni r$ which minimizes

$\sum_{e \in T} w(e) + \sum_{v \notin T} p(v)$. We create a first layer, with edge weights $w^1 = w$. Then we construct a second layer, where we set $w^2(\{r, v\}) = p(v)$ for any $v \in V$. All the other layers, if any, are dummy layers defined as above. This reduction is approximation preserving. ◀

▶ **Theorem 8.** *For an arbitrary number of layers, rooted* Union k-MST *and* Union k-Metric Facility Location *are not approximable better than* $\Omega(\log k)$ *unless* $P = NP$*, even when* $k = n$*.*

Proof. We prove the claim for rooted Union k-MST, by giving a reduction from cardinality Set Cover: given a universe $\mathcal{U}$ of n' elements, and a collection $\mathcal{S} = \{S_1, \ldots, S_{m'}\}$ of m' subsets of $\mathcal{U}$, find a minimum cardinality subset $\mathcal{A} \subseteq \mathcal{S}$ which spans $\mathcal{U}$. This problem is $\Omega(\log n')$-hard to approximate [24]. We create one node per element of $\mathcal{U}$, plus two extra nodes r and s. We create one layer i for each set S_i (i.e., $h = m'$). In layer i we let $w^i(\{r, s\}) = 1$ and $w^i(\{s, v\}) = 0$ for each $v \in S_i$. We also let $r^i := r$ for each i, and assume $k = n = n' + 2$. Note that any solution to the rooted Union k-MST instance of cost α can be turned into a solution to the Set Cover instance of the same cost, and vice versa.

To prove the claim for Union k-Metric Facility Location, we use the same reduction as above, where the edge $\{r, s\}$ is replaced by a single node r, which is a facility of opening cost 1. ◀

A simple greedy algorithm guarantees a $O(\log k)$-approximation which matches the above lower bound. The basic idea is as follows. Suppose that the considered covering problem satisfies a natural *composition* property, namely two solutions satisfying k' and k'' distinct requests, can be merged (without increasing the total cost) to obtain a solution satisfying $k' + k''$ requests. (Merging might involve some polynomial-time operations). Suppose also that there exists a ρ-approximation for one layer. The idea is then to compute, for each layer separately and for each $k' \leq k$, a ρ-approximate solution to the partial covering instance induced by that layer with target k'. The solution providing the best ratio of cost to number k' of satisfied requests is *merged* with the solution under construction. Then satisfied requests are *removed* from the set of requests, k decremented by k', and the process is iterated until $k \leq 0$. Via standard techniques this algorithms provides a $O(\rho \cdot \log k)$-approximation.

Removing requests transforms a given k-MST instance in each layer into a k-Steiner Tree instance: for the latter problem there is a 4-approximation algorithm [12]. Note also that all the partial solutions in each layer contain the corresponding root r^i: hence the merging step is trivial. For k-Metric Facility Location, there is a 2-approximation algorithm in [19]. In this case removing a request simply means removing one client, and the merging step is trivial. Altogether:

▶ **Theorem 9.** *There exist* $O(\log k)$*-approximation algorithms for* Union k-MST *and* Union k-Metric Facility Location*.*

We next describe an LP-based $O(h)$-approximation algorithm for rooted Union k-MST. This is an improvement over the $\Theta(\log k)$-approximation given by the greedy algorithm for the relevant case of bounded h.

For notational convenience, we assume that the roots $R := \cup_i \{r^i\}$ are not counted into the target number k of connected nodes. In other terms, we replace k by $k - |R|$. We make the same assumption also in the case of one layer. Consider the following LP relaxation for

k-Steiner Tree ($W \ni r$ is the set of terminals) denoted by $LP_{kST}(w, W, V, r, k)$:

$$\begin{array}{lll} \min & \sum_{e\in E} w(e)\, x_e & \\ s.t. & \sum_{e\in\delta(S)} x_e \geq z_v, & \forall (v,S) : S \subseteq V - \{r\}, v \in S \cap W; \\ & \sum_{v\in W} z_v \geq k; & \\ & x_e \geq 0, 1 \geq z_v \geq 0, & \forall v \in W, \forall e \in E. \end{array}$$

Here, variable x_e indicates whether edge e is included in the solution, whereas variable z_v indicates whether terminal v is connected. Moreover $\delta(S)$ denotes the set of edges with exactly one endpoint in S. Observe that $LP_{kMST}(w, V, r, k) := LP_{kST}(w, V, V, r, k)$ is an LP relaxation for k-MST. We need the following lemmas.

▶ **Lemma 10.** *[11] Let (w, V, r, k) be an instance of k-MST, $w_{max} := \max_{v\in V}\{w(r, v)\}$, and opt' be the optimal solution to $LP_{kMST}(w, V, r, k)$. There is a polynomial-time algorithm* `apx-kmst` *which computes a solution to the instance of cost at most $2opt' + w_{max}$.*

▶ **Lemma 11.** *[10] Let $G = (V \cup \{v\}, E)$ be a directed graph, with edge capacities $\alpha : E \to \mathbb{R}_{\geq 0}$ such that $\sum_{e\in\delta^+(u)} \alpha(e) = \sum_{e\in\delta^-(u)} \alpha(e)$ for all $u \in V \cup \{v\}$. Then there is a pair of edges (u, v) and (v, z), such that the following capacity reservation β supports the same flow as α between any pair of nodes in V: for $\Delta\alpha := \min\{\alpha(u, v), \alpha(v, z)\}$, set $\beta(u, v) = \alpha(u, v) - \Delta\alpha$, $\beta(v, z) = \alpha(v, z) - \Delta\alpha$, $\beta(u, z) = \alpha(u, z) + \Delta\alpha$, and $\beta(e) = \alpha(e)$ for the remaining edges e.*

▶ **Corollary 12.** *Given a feasible solution (x, z) to $LP_{kST}(w, W, V, r, k)$, there is a feasible solution (x', z') to $LP_{kMST}(w, W, r, k)$ such that $\sum_e w(e)x'_e \leq 2 \cdot \sum_e w(e)x_e$.*

Proof. Variables x_e can be interpreted as a capacity reservation which supports a fractional flow of value z_v from each $v \in W$ to the root. Let us replace each edge with two oppositely directed edges, and assign to each such edge the same weight and capacity as the original edge. This way, we obtain a capacity reservation α which costs twice the original capacity reservation, and satisfies the condition of Lemma 11. We consider any non-terminal node $v \neq r$ with some incident edge of positive capacity, and apply Lemma 11 to it. Due to triangle inequality, the cost of the capacity reservation does not increase. We iterate the process on the resulting capacity reservation. Within a polynomial number of steps, we obtain a capacity reservation β which: (1) supports the same flow from each terminal to the root r as α, (2) has value 0 on edges incident to non-terminal nodes (besides r), and (3) does not cost more than α. At this point, we remove the nodes $V - (W \cup \{r\})$, and merge the capacity of oppositely directed edges to get an undirected capacity reservation x'. By construction, the pair (x', z) is a feasible solution to $LP_{kMST}(w, W, r, k)$ of cost at most $2 \cdot \sum_e w(e)x_e$. ◀

We are now ready to describe our algorithm for rooted Union k-MST. In a preliminary step we guess the largest distance L in the optimal solution between any connected node and the corresponding root, and discard nodes at distance larger than L from their root. This introduces a factor $O(nh)$ in the running time. Note that $L \leq opt$. We let V^i be the remaining nodes in layer i.

Then we compute the optimal fractional solution $OPT^* = (x^i, z^i, z)_i$, of cost opt^*, to the following LP relaxation LP_{ukMST} for the problem, where variables x^i_e and z^i_v indicate whether edge e is included in the solution of layer i and node v is connected in layer i, respectively. Variable z_v indicates whether node v is connected in at least one layer.

$$
\begin{array}{lll}
\min & \sum_{i=1,\dots,h} \sum_{e\in E} w^i(e)\, x^i_e & \\
s.t. & \sum_{e\in\delta(S)} x^i_e \geq z^i_v, & \forall i \in \{1,\dots,h\}, \forall (v,S) : S \subseteq V^i - \{r^i\}, v \in S; \\
& \sum_{i=1,\dots,h} z^i_v \geq z_v, & \forall v \in V - R; \\
& \sum_{v\in V-R} z_v \geq k; & \\
& z^i_v, x^i_e \geq 0, 1 \geq z_v \geq 0, & \forall i \in \{1,\dots,h\}, \forall v \in V - R, \forall e \in E.
\end{array}
$$

Given OPT^*, we compute for each layer i a subset of nodes W^i, where v belongs to W^i iff $z^i_v = \max_{j=1,\dots,h}\{z^j_v\}$ (breaking ties arbitrarily). We also define $k^i := \lfloor \sum_{v\in W^i} z_v \rfloor$. For each layer i, we consider the k-MST instance on nodes $W^i \cup \{r^i\}$ with target k^i. This instance is solved using the 2-approximation algorithm `apx-kmst` of Lemma 10: the resulting tree T^i is added to the solution for layer i. Let k' be the number of connected nodes. If $k' < k$, the algorithm connects $k - k'$ extra nodes, chosen greedily, to the corresponding root in order to reach the global target k.

▶ **Theorem 13.** *There is a $O(h)$-approximation algorithm for rooted* Union k-MST. *The running time of the algorithm is $O((nh)^{O(1)})$.*

Proof. Consider the above algorithm. The claim on the running time is trivial. By construction, the solution computed is feasible (i.e., it connects k nodes). It remains to consider the approximation factor.

For each $v \in W^i$, we let $\tilde{z}^i_v = z_v$, and set $\tilde{z}^i_v = 0$ for the remaining nodes. Furthermore, we let $\tilde{x}^i_e = h \cdot x^i_e$. Observe that $(\tilde{x}^i, \tilde{z}^i, z)_i$ is a feasible fractional solution to LP_{ukMST} of cost $h \cdot opt^*$. Observe also that $(\tilde{x}^i, \tilde{z}^i)$ is a feasible solution to $LP_{kST}(w^i, W^i, V^i, r^i, k^i)$: let $a\tilde{p}x^i$ be the associated cost. By Lemma 11, there is a fractional solution to $LP_{kMST}(w^i, W^i, r^i, k^i)$ of cost at most $2a\tilde{p}x^i$. It follows from Lemma 10 that the solution computed by `apx-kmst` on layer i costs at most $4a\tilde{p}x^i + L$.

Since the W^i's are disjoint, the algorithm initially connects at least $\sum_i k^i \geq k - h$ nodes. Hence the cost of the final augmentation phase is at most $h \cdot L \leq h \cdot opt$. Putting everything together, the cost of the solution returned by the algorithm is at most:

$$\sum_i (4 \cdot a\tilde{p}x^i + L) + h \cdot L \leq 4h \cdot opt^* + 2h \cdot L \leq 6h \cdot opt\,.$$

◀

The constant multiplying h in the approximation factor can be reduced with a more technical analysis, at the cost of a higher running time. We also observe that the integrality gap of LP_{ukMST} is $\Omega(h)$. In fact, consider the (cardinality) Set Cover instance as in [27]: given a hyper-graph on m' nodes, with hyper-edges given by all subsets of $m'/2$ nodes, create an element for each hyper-edge, and a set for each node containing all the hyper-edges incident to that node. It is easy to see that the best fractional solution for the standard set cover LP (assigning value $2/m'$ to all elements) has cost 2, while the best integral solution contains $m'/2 + 1$ sets. The same reduction as in Theorem 8 implies the claim.

Essentially the same approach works also for Union k-Metric Facility Location. Also in this case, we can show that the corresponding LP has integrality gap $\Omega(h)$.

▶ **Theorem 14.** *There is a $O(h)$-approximation algorithm for* Union k-Metric Facility Location. *The running time of the algorithm is $O((nh)^{O(1)})$.*

3.2 Unrooted Union k-MST

▶ **Theorem 15.** *Unrooted* Union k-MST *is not approximable in polynomial time for an arbitrary number h of layers unless $P = NP$.*

Proof. We give a reduction from SAT: given a CNF boolean formula on m' clauses and n' variables, determine whether it is satisfiable or not. For each variable i, we create two nodes t_i and f_i. Intuitively, these nodes represent the fact that i is true or false, respectively. Furthermore, we have a node for each clause. Hence the overall number of nodes is $n = 2n' + m'$. We create a separate layer for each variable i (i.e., $h = n'$). In layer i, we connect with an edge of cost zero t_i (resp., f_i) to all the clauses which are satisfied by setting i to true (resp., to false)[3]. The target value is $k = n' + m'$. Note that, there is a satisfying assignment to the SAT instance iff there is a solution of cost zero to the Union k-MST instance. ◀

For $h = O(1)$, the rooted and the unrooted versions of the problem are equivalent approximation-wise. In fact, one obtains an approximation-preserving reduction from the unrooted to the rooted case by guessing one node r^i in the optimal solution per layer: this introduces a polynomial factor $O(n^h)$ in the running time. We remark that an exponential dependence on h of the running time is unavoidable in the unrooted case, due to Theorem 15. An opposite reduction is obtained by appending n dummy nodes to each root (distinct nodes for distinct layers), with edges of cost zero, and setting the target to $k + hn$. The following result follows.

▶ **Corollary 16.** *Unrooted* Union k-MST *is APX-hard for any $h \geq 1$. There is a $O(h)$-approximation algorithm for the problem of running time $O((hn)^{O(1)}n^h)$.*

4 Conclusions and Open Problems

In this paper, we introduced multi-layer covering problems, a new framework that can be used to describe a wide spectrum of yet unstudied problems. We addressed two natural ways of combining the layers: intersection and union. We gave multi-layer approximation algorithms, as well as hardness results, for a few classic covering problems (and their partial covering versions). There are several research questions that merit further study.

- There are other natural ways one can combine the layers. Consider, for example, the car/bike problem in the case where you can put your bike in the car trunk. Now you can make more than one tour by bike, the only requirement being that the bike tours all touch the (unique) car tour.
- What about min-max multi-layer problems, where the goal is to minimize the maximum cost over the layers?
- We considered covering problems: what about packing problems?
- Our algorithms for union problems give tight bounds only with respect to the corresponding natural LPs. This leaves room for improvement.
- There is a considerable gap between upper and lower bounds for intersection problems. In particular, our hardness results do not depend on h, while the approximation ratios deteriorate rather rapidly for increasing h.

[3] Without loss of generality, we can assume that each clause does not contain both a literal and its negation.

References

1. A. Anagnostopoulos, L. Becchetti, C. Castillo, A. Gionis, and S. Leonardi. Power in unity: forming teams in large-scale community systems. In *CIKM*, pages 599–608, 2010.
2. B. Applebaum, B. Barak, and A. Wigderson. Public-key cryptography from different assumptions. In *STOC*, pages 171–180, 2010.
3. S. Arora, B. Barak, M. Brunnermeier, and R. Ge. Computational complexity and information asymmetry in financial products (extended abstract). In *ICS*, pages 49–65, 2010.
4. A. Berger, V. Bonifaci, F. Grandoni, and G. Schäfer. Budgeted matching and budgeted matroid intersection via the gasoline puzzle. *Mathematical Programming*, 128:355–372, 2011.
5. M. Bern and P. Plassmann. The Steiner problem with edge lengths 1 and 2. *Information Processing Letters*, 32:171–176, 1989.
6. A. Bhaskara, M. Charikar, E. Chlamtac, U. Feige, and A. Vijayaraghavan. Detecting high log-densities – an $O(n^{1/4})$ approximation for densest k–subgraph. In *STOC*, pages 201–210, 2010.
7. J. Edmonds. *Matroid intersection*. North-Holland, 1979.
8. F. Eisenbrand, F. Grandoni, T. Rothvoß, and G. Schäfer. Connected facility location via random facility sampling and core detouring. *Journal of Computer and System Sciences*, 76:709–726, 2010.
9. L. Fleischer, J. Könemann, S. Leonardi, and G. Schäfer. Simple cost sharing schemes for multicommodity rent-or-buy and stochastic Steiner tree. In *STOC*, pages 663–670, 2006.
10. A. Frank. On connectivity properties of Eulerian digraphs. *Annals of Discrete Mathematics*, 41:179–194, 1989.
11. N. Garg. A 3-approximation for the minimum tree spanning k vertices. In *FOCS*, pages 302–309, 1996.
12. N. Garg. Saving an epsilon: a 2-approximation for the k-MST problem in graphs. In *STOC*, pages 396–402, 2005.
13. F. Grandoni and G. F. Italiano. Improved approximation for single-sink buy-at-bulk. In *ISAAC*, pages 111–120, 2006.
14. F. Grandoni, R. Ravi, and M. Singh. Iterative rounding for multi-objective optimization problems. In *ESA*, pages 95–106, 2009.
15. F. Grandoni and R. Zenklusen. Approximation schemes for multi-budgeted independence systems. In *ESA (1)*, pages 536–548, 2010.
16. Fabrizio Grandoni, Thomas Rothvoß, and Laura Sanita. From uncertainty to nonlinearity: Solving virtual private network via single-sink buy-at-bulk. *Mathematics of Operations Research*, 36(2):185–204, 2011.
17. S. Guha and S. Khuller. Greedy strikes back: Improved facility location algorithms. In *SODA*, pages 649–657, 1998.
18. Mohammad Taghi Hajiaghayi and Kamal Jain. The prize-collecting generalized steiner tree problem via a new approach of primal-dual schema. In *SODA*, pages 631–640, 2006.
19. K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *Journal of the ACM*, 50(6):795–824, 2003.
20. Ravishankar Krishnaswamy, Amit Kumar, Viswanath Nagarajan, Yogish Sabharwal, and Barna Saha. The matroid median problem. In *SODA*, 2011. To appear.
21. T. Lappas, K. Liu, and E. Terzi. Finding a team of experts in social networks. In *KDD*, pages 467–476, 2009.
22. C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of Web sources. In *FOCS*, pages 86–92, 2000.

23 R. Ravi and M. X. Goemans. The constrained minimum spanning tree problem (extended abstract). In *SWAT*, pages 66–75, 1996.

24 R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *STOC*, pages 475–484, 1997.

25 M. Riaz, R. Nielsen, J. Pedersen, N. Prasad, and O. Madsen. A framework for planning a unified wired and wireless ict infrastructure. *Wireless Personal Communications*, 54:169–185, 2010.

26 P. Slavik. Improved performance of the greedy algorithm for partial cover. *Information Processing Letters*, 64(5):251–254, 1997.

27 V. V. Vazirani. *Approximation Algorithms*. Springer, 2003.