



A primal-dual algorithm for the minimum partial set multi-cover problem

Yingli Ran¹ · Yishuo Shi² · Changbing Tang¹ · Zhao Zhang¹ 

Published online: 13 December 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

In a minimum partial set multi-cover problem (MinPSMC), given an element set E , a collection of subsets $\mathcal{S} \subseteq 2^E$, a cost w_S on each set $S \in \mathcal{S}$, a covering requirement r_e for each element $e \in E$, and an integer k , the goal is to find a sub-collection $\mathcal{F} \subseteq \mathcal{S}$ to fully cover at least k elements such that the cost of \mathcal{F} is as small as possible, where element e is fully covered by \mathcal{F} if it belongs to at least r_e sets of \mathcal{F} . On the application side, the problem has its background in the seed selection problem in a social network. On the theoretical side, it is a natural combination of the minimum partial (single) set cover problem (MinPSC) and the minimum set multi-cover problem (MinSMC). Although both MinPSC and MinSMC admit good approximations whose performance ratios match those lower bounds for the classic set cover problem, previous studies show that theoretical study on MinPSMC is quite challenging. In this paper, we prove that MinPSMC cannot be approximated within factor $O(n^{\frac{1}{2(\log \log n)^c}})$ for some constant c under the ETH assumption. Furthermore, assuming r_{\max} is a constant, where $r_{\max} = \max_{e \in E} r_e$ is the maximum covering requirement and f is the maximum number of sets containing a common element, we present a primal-dual algorithm for MinPSMC and show that its performance ratio is $O(\sqrt{n})$. We also improve the ratio for a restricted version of MinPSMC which possesses a graph-type structure.

Keywords Partial set multi-cover problem · Positive influence dominating set · Densest l -subgraph problem · Primal-dual · Approximation algorithm

✉ Zhao Zhang
hxhzz@163.com

¹ College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua 321004, Zhejiang, China

² Institute of Information Science, Academia Sinica, Taipei 11529, Taiwan

1 Introduction

In this paper, we study approximation algorithms for the minimum partial set multi-cover problem whose formal definition is given as follows.

Definition 1.1 (*Minimum Partial Set Multi-Cover (MinPSMC)*) Given an element set E consisting of n elements, a collection of subsets $\mathcal{S} : 2^E \mapsto \mathbb{R}^+$, a nonnegative weight w_S for each subset $S \in \mathcal{S}$, a covering requirement r_e for each element $e \in E$, and an integer $k \leq n$, the MinPSMC problem is to find a minimum weight sub-collection $\mathcal{F} \subseteq \mathcal{S}$ such that at least k elements are fully covered by \mathcal{F} , where an element e is fully covered by \mathcal{F} means that e is contained in at least r_e sets of \mathcal{F} , and the weight of sub-collection \mathcal{F} is $w(\mathcal{F}) = \sum_{S \in \mathcal{F}} w_S$. An instance of MinPSMC is denoted as $(E, \mathcal{S}, w, r, k)$.

In particular, when $r_e \equiv 1$, then the MinPSMC problem is the *minimum partial set cover problem* (MinPSC), when $k = n$, then the MinPSMC problem is the *minimum set multi-cover problem* (MinSMC). In this paper, it is always assumed that picking all sets will fully cover all elements.

The study of the MinPSMC problem is motivated by a *seed selection problem* in a social network. Social network is an important medium for the spread of information and opinions. How information is spread depends on the structure of the network and how opinions are spread depends on the mechanism of influence. One of the most important topics people concern about is to which extent an opinion can be accepted. Following the seminal work of Kempe et al. (2003) on the influence maximization problem, there are a huge body of studies in this field. Most of the studies are on various probabilistic spreading models. One widely studied model is the *linear threshold model* in which a node is influenced only when the ratio between the number of his influenced neighbors and the number of all his neighbors exceeds a threshold, where the threshold function on the nodes is distributed uniformly and independently in $[0, 1]$. Good performance ratios were achieved by exploring the submodularity of the influence function (Kempe et al. 2003, 2005). One may be wondering about a deterministic model in which the threshold function is *predetermined*. Previous studies show that researches on such a deterministic model is very hard. In fact, Chen (2009) proved that under such an influence mechanism, the minimum seeding problem, the goal of which is to select the minimum number of initially selected nodes (called seeds) to influence all nodes, does not admit an $O(2^{\log^{1-\epsilon} n})$ -approximation unless $NP \subseteq DTIME(n^{\text{polylog}(n)})$, where n is the number of nodes in the network. However, if one only considers the *minimum one-step seeding problem*, the goal of which is to select the minimum number of seeds to influence all the people in the network *in one time slot*, then the problem is a special case of the MinSMC problem (see the explanation in Sect. 1.1), and thus admits good approximations. In the real world, because of economic considerations, it is often more cost-effective to influence only a fraction of people. Such a consideration leads to the *minimum partial seeding problem*, which is a special case of the MinPSMC problem.

The MinPSMC problem is a combination of the MinSMC problem and the MinPSC problem. There are a lot of studies on MinPSC and MinSMC, achieving tight performance ratios matching those lower bounds for the classic set cover problem. However,

the study on MinPSMC is very rare. According to recent studies (Ran et al. 2017a, b), this problem is quite challenging theoretically.

1.1 Related works

The minimum one-step seeding problem is to select the minimum number of seeds to influence all the people in one time slot. This problem is also known as the *minimum positive dominating set problem* (MinPDS) in Dinh et al. (2014) which can be defined as follows: given a graph $G = (V, E)$, a constant $0 < \rho \leq 1$, the goal is to find a node set $D \subseteq V$ with the minimum size such that every vertex v in V has at least $\lceil \rho d(v) \rceil$ neighbors in D , where $d(v)$ is the degree of node v in G . It can be viewed as a special case of MinSMC by setting $E = V(G)$, $\mathcal{S} = \{N_G(v) : v \in V\}$ where $N_G(v)$ is the set of neighbors of v in G , and $r_v = \lceil \rho d(v) \rceil$. Wang et al. (2009, 2011) proved that the MinPDS problem is APX-hard and proposed a greedy algorithm with performance ratio $H(\Delta)$, where Δ is the maximum degree of the graph and $H(\Delta) = \sum_{i=1}^{\Delta} 1/i$ is the Harmonic number. The same ratio was obtained by Dinh et al. (2014) by observing the relation between MinPDS and MinSMC.

The minimum set cover problem (MinSC) was one of the 21 problems shown to be NP-hard in Karp's seminal paper (Karp 1972). Feige (1996) proved that unless $NP \subseteq DTIME(n^{O(\log \log n)})$, MinSC does not admit performance ratio $\rho \ln n$ for any $0 < \rho < 1$, where n is the number of elements. Dinur and Steurer (2014) proved that this lower bound holds if $P \neq NP$. For the cardinality version of MinSC, Johnson (1974) and Lovász (1975) obtained a greedy $H(\Delta)$ -approximation algorithm, where Δ is the maximum cardinality of a set and $H(\Delta) = \sum_{i=1}^{\Delta} 1/i$ is the Harmonic number. The same performance ratio was obtained for the weighted version of MinSC by Chvatal (1979). Another well-known performance ratio for MinSC is f , the maximum number of sets containing a common element (Vazirani 2001), which can be achieved by either an LP rounding algorithm (Hochbaum 1982) or a local ratio method (Bar-Yehuda and Even 1985). By Khot and Regev (2008), ratio f is also best possible.

For MinSMC, Vazirani (2001) showed that a greedy algorithm achieves performance ratio $H(n)$ using *dual fitting* analysis. The same performance ratio was obtained by a primal-dual algorithm presented by Rajagopalan and Vazirani (1993).

The MinPSC problem was first studied by Kearns and Ortiz (2003), and a greedy algorithm was presented with performance ratio at most $2H(n) + 3$. Slavík (1997) improved the algorithm, obtaining performance ratio $\min\{H(\Delta), H(\lceil pn \rceil)\}$. Using primal-dual method, Gandhi et al. (2004) gave an approximation algorithm with performance ratio f . The same performance ratio f was also obtained by Bar-Yehuda (2001) using local ratio method.

It can be seen from the above related work that both MinSMC and MinPSC have approximation algorithms with the best possible performance ratios, matching those lower bounds for the classic set cover problem. On the contrary, study on MinPSMC seems very difficult. Ran et al. (2017a) were the first to obtain a guaranteed performance ratio for the MinPSMC problem. However, their ratio is meaningful only when the covering percentage $p = k/n$ is very close to 1. Afterwards, in (2017b), Ran et al. presented a simple greedy algorithm for MinPSMC achieving performance ratio Δ .

Notice that Δ can be as large as n , and in terms of Δ , the performance ratio for MinSMC and MinPSC is of order $\ln \Delta$. In Ran et al. (2017b), the authors presented a local ratio algorithm for MinPSMC, which reveals a “shock wave” phenomenon: their performance ratio is f for both MinPSC and MinSMC (which is best possible), but for MinPSMC, the ratio jumps abruptly to $O(n)$ even when the covering percentage p is smaller than 1 by a very small constant. In view of these results, the study of MinPSMC seems to be very challenging.

1.2 Our contribution

In this paper, we obtain the following results for MinPSMC.

- (i) We prove a lower bound for MinPSMC by a reduction from the famous *densest l -subgraph problem* (D/S). Combining this reduction with the hardness result for D/S obtained in Manurangsi (2017), under the ETH assumption, MinPSMC can not be approximated within factor $O(n^{\frac{1}{(\log \log n)^c}})$ for some constant c .
- (ii) Under the assumption that the maximum covering requirement $r_{\max} = \max\{r_e : e \in E\}$ is upper bounded by a constant, we present a primal-dual algorithm for MinPSMC, obtaining performance ratio $B + \sqrt{B \cdot n}$, where $B = \max\{\binom{f_e}{r_e} : e \in E\}$ and f_e is the number of sets containing element e . To use the primal-dual method, how to design a linear program based on which a good approximation can be achieved is a crucial step. We propose a novel integer program the relaxation of which (using Lovász extension Lovász 1975) is a convex program. Using the fact that for a submodular function, its Lovász extension coincides with its convex closure, we modify it into a linear program. Although the linear program has exponential number of variables, we show that our primal-dual algorithm can be executed in polynomial time, making use of an efficient algorithm for minimizing a submodular function divided by a modular function (Fleisher and Iwata 2003). Our algorithm consists of two stages. The first stage is a primal-dual algorithm. After the first stage, the sub-collection of sets selected by the last iteration may fully cover much more elements than required by the remaining covering requirement. Hence the second stage refines the solution by iteratively implementing submodular minimization algorithms (Fleisher and Iwata 2003).
- (iii) We improve the performance ratio for a restricted version of MinPSMC, in which $w_S \equiv 1$, $r_e \equiv 2$, and $f_e \equiv 2$, where $f_e = |\{S \in \mathcal{S} : e \in S\}|$ is the frequency of element e . This restricted version looks more like an optimization problem on a graph. Making use of structural properties of graphs, the performance ratio can be improved to $1 + \sqrt{2}n^{1/4}$.

The paper is organized as follows. In Sect. 2, we give the definitions of some concepts and terminologies which are used in this paper. In Sect. 3, we prove a lower bound for MinPSMC. In Sect. 4, we present a primal-dual algorithm for MinPSMC and provide its performance analysis. In Sect. 5, the performance ratio is improved for the restricted MinPSMC problem. Section 6 concludes the paper with some discussions on future work.

2 Preliminaries

In this section, we introduce some concepts, terminologies, and tools we shall use.

Definition 2.1 (*Submodular*) For a finite nonempty set V , a set function $\rho : 2^V \mapsto \mathbb{R}$ is *submodular* if for any subsets $X, Y \subseteq V$,

$$\rho(X) + \rho(Y) \geq \rho(X \cap Y) + \rho(X \cup Y).$$

An equivalent definition of submodular function is that for any $X \subseteq Y \subseteq V$ and $v \notin Y$,

$$\Delta_v \rho(X) \geq \Delta_v \rho(Y),$$

where $\Delta_v \rho(X) = \rho(X \cup \{v\}) - \rho(X)$.

Definition 2.2 (*Lovász extension*) Suppose $\rho : 2^V \mapsto \mathbb{R}$ is a set function. The Lovász extension of ρ is a function $\hat{\rho} : \mathbb{R}^V \mapsto \mathbb{R}$ defined as follows. For any vector $\mathbf{z} \in \mathbb{R}^V$, order elements of V as v_1, v_2, \dots, v_n such that $1 = z_0 \geq z_1 \geq z_2 \geq \dots \geq z_n \geq z_{n+1} = 0$ where z_i is the component of \mathbf{z} indexed by v_i . Let $V_0 = \emptyset$ and $V_i = \{v_1, \dots, v_i\}$ for $i = 1, \dots, n$. The value of $\hat{\rho}$ at \mathbf{z} is

$$\hat{\rho}(\mathbf{z}) = \sum_{j=0}^n (z_j - z_{j+1}) \rho(V_j).$$

The following result reveals the relationship between submodularity and convexity.

Theorem 2.3 (Lovász 1983) *A set function ρ is submodular if and only if its Lovász extension $\hat{\rho}$ is a convex function.*

Definition 2.4 (*Convex Closure*) For a set function $\rho : 2^V \mapsto \mathbb{R}$, the convex closure of ρ is the point-wise highest convex function $\rho^- : \mathbb{R}^V \mapsto \mathbb{R}$ that always lowerbounds ρ . For any $\mathbf{x} \in \mathbb{R}^V$, $\rho^-(\mathbf{x})$ has the following expression:

$$\begin{aligned} &\rho^-(\mathbf{x}) \\ &= \min \left\{ \sum_{A \subseteq V} \rho(A) z_A : \sum_{A: v_i \in A \subseteq V} z_A = x_i, \forall v_i \in V; \sum_{A \subseteq V} z_A = 1; z_A \geq 0, \forall A \subseteq V \right\} \end{aligned} \tag{1}$$

The following result reveals the relationship between Lovász extension and convex closure.

Theorem 2.5 (Edmonds 1970) *If ρ is submodular, then its Lovász extension is the same as its convex closure.*

To prove a lower bound for MinPSMC, we shall reduce the Densest l -Subgraph problem to a special case of the MinPSMC problem.

Definition 2.6 (*Densest l -Subgraph (DIS)*) Given a graph $G = (V, E)$ with $|V| = n$ and an integer $l \leq n$, the DIS problem asks for a vertex subset C on l vertices such that the subgraph of G induced by C , denoted as $G[C]$, has the maximum number of edges among all subgraphs of G on l vertices. An instance of DIS is denoted as (G, l) .

Definition 2.7 (*Minimum Restricted Partial Set Multi-Cover (MinRPSMC)*) MinRPSMC is a special MinPSMC in which $w_S \equiv 1$, $r_e \equiv 2$, and $f_e \equiv 2$, where $f_e = |\{S \in \mathcal{S} : e \in S\}|$ is the frequency of element e . An instance of MinRPSMC is denoted as (E, \mathcal{S}, k) .

The following notations and assumptions will be used in this paper. For a sub-collection $\mathcal{F} \subseteq \mathcal{S}$, denote by $\mathcal{C}(\mathcal{F})$ the set of elements fully covered by \mathcal{F} . For an element $e \in E$, let f_e be the number of sets containing e , and let $f = \max\{f_e : e \in E\}$. We assume that $r_e \leq f_e$ holds for every $e \in E$, since an element e with $r_e > f_e$ can be removed from consideration. Denote by

$$B = \max_{e \in E} \binom{f_e}{r_e}. \quad (2)$$

Notice that $B \leq f^{r_{\max}}$. This paper studies the MinPSMC problem under the assumption that r_{\max} is a constant. To avoid ambiguity, we shall use \mathbb{E} to denote *expectation* in order to distinguish it from the symbol E which denotes the edge set of a graph or the element set of an MinPSMC instance.

3 Lower bound for MinPSMC

In this section, we prove a lower bound for MinPSMC by a reduction from DIS.

Theorem 3.1 *Suppose β is a lower bound for the performance ratio of DIS. Then $\sqrt{\frac{\beta}{2}}$ is a lower-bound for the performance ratio of MinRPSMC.*

Proof We shall show that if MinRPSMC has a polynomial-time γ -approximation, then DIS has a polynomial-time $2\gamma^2$ -approximation.

Given a DIS instance (G, l) , for an integer k between 1 and $|E|$, construct an instance of MinRPSMC as follows: the ground set is E ; for each $v \in V(G)$, let S_v be the subset of edges incident with vertex v in G ; the collection of sets $\mathcal{S} = \{S_v : v \in V(G)\}$; the covering requirement is k . Notice that every edge is incident with exactly two vertices, and thus $f_e \equiv 2$. So, the constructed instance satisfies the frequency requirement of MinRPSMC.

For any sub-collection of sets $\mathcal{F} \subseteq \mathcal{S}$, let vertex set $C_{\mathcal{F}} = \{v : S_v \in \mathcal{F}\}$. Then

$$|C_{\mathcal{F}}| = |\mathcal{F}|. \quad (3)$$

Notice that by $r_e \equiv 2$, if an element e is fully covered by \mathcal{F} , then both of its two ends belong to $C_{\mathcal{F}}$, which implies that e corresponds to an edge in $G[C_{\mathcal{F}}]$. Hence

$$|E(G[C_{\mathcal{F}}])| = |\mathcal{C}(\mathcal{F})|. \tag{4}$$

Conversely, for each vertex set $C \subseteq V(G)$, let sub-collection $\mathcal{F}_C = \{S_v : v \in C\}$, then

$$|\mathcal{F}_C| = |C| \tag{5}$$

and

$$|\mathcal{C}(\mathcal{F}_C)| = |E(G[C])|. \tag{6}$$

Assume that C^* is an optimal solution to *DIS* and $opt_D = |E(G[C^*])|$ is the optimal value, that is, opt_D is the number of edges in a densest subgraph of G on l vertices. □

Claim 1 There exists a vertex set C' on l/γ vertices such that $G[C']$ has at least $opt_D/(2\gamma^2)$ edges when $\gamma \leq (l + 1)/2$.

We first show that by uniformly randomly picking $\frac{l}{\gamma}$ vertices from C^* to form a vertex set C' , the expected number of edges in $G[C']$ is at least $opt_D/(2\gamma^2)$. In fact, for each edge $e \in E(G[C^*])$, it belongs to $E(G[C'])$ if and only if both of its two ends are picked. Since there are $\binom{|C^*|}{\frac{l}{\gamma}}$ possible choices of $\frac{l}{\gamma}$ vertices and $\binom{|C^*|-2}{\frac{l}{\gamma}-2}$ of them include the two ends of edge e , hence the probability that $e \in E(G[C'])$ is

$$Pr[e \in E(G[C'])] = \frac{\binom{|C^*|-2}{\frac{l}{\gamma}-2}}{\binom{|C^*|}{\frac{l}{\gamma}}} = \frac{\frac{l}{\gamma}(\frac{l}{\gamma}-1)}{|C^*|(|C^*|-1)}. \tag{7}$$

Combining this with $|C^*| = l$ and the assumption $\gamma \leq (l + 1)/2$, we have

$$Pr[e \in E(G[C'])] \geq \frac{1}{2\gamma^2}.$$

So, the expected number of edges in $G[C']$ is

$$\begin{aligned} \mathbb{E}[|E(G[C'])|] &= \sum_{e \in E(G[C^*])} Pr[e \in E(G[C'])] \\ &\geq \sum_{e \in E(G[C^*])} \frac{1}{2\gamma^2} = \frac{1}{2\gamma^2} |E(G[C^*])| = \frac{1}{2\gamma^2} opt_D \end{aligned}$$

Then the claim follows from the Pigeon Hole Principle.

Claim 2 Making use of a γ -approximation algorithm for MinRPSMC for $|E|$ rounds, a feasible solution C to D/S on graph G can be found such that $|E(G[C])| \geq opt_D/(2\gamma^2)$.

Notice that D/S has a trivial l -approximation algorithm: arbitrarily picking $\lfloor l/2 \rfloor$ edges, the subgraph induced by the picked edges has at most l vertices; extending it to a subgraph on exactly l vertices results in a subgraph with at least $(l - 1)/2$ edges while an optimal solution has at most $\frac{l(l-1)}{2}$ edges. So, if $l \leq 2\gamma^2$, then the claim is true. In the following, we assume that $l > 2\gamma^2$. In this case,

$$\gamma < \sqrt{l/2} < (l + 1)/2.$$

Suppose \mathcal{A} is a γ -approximation algorithm for MinRPSMC. For $k = 1, 2, \dots, |E|$, denote by \mathcal{F}_k the output of algorithm \mathcal{A} on instance $(E, \mathcal{S}, k/(2\gamma^2))$ and \mathcal{F}_k^* an optimal solution to $(E, \mathcal{S}, k/(2\gamma^2))$ (we simplify the derivation by regarding $k/(2\gamma^2)$ as an integer). It follows from Claim 1 and Eq. (6) that there exists a vertex set $C' \subseteq V(G)$ on l/γ vertices such that

$$|C(\mathcal{F}_{C'})| = |E(G[C'])| \geq \frac{opt_D}{2\gamma^2}.$$

Hence $\mathcal{F}_{C'}$ is a feasible solution to $(E, \mathcal{S}, opt_D/(2\gamma^2))$. Making use of Claim 1 and equalities (3) and (5), we have

$$|C_{\mathcal{F}_{opt_D}}| = |\mathcal{F}_{opt_D}| \leq \gamma |\mathcal{F}_{opt_D}^*| \leq \gamma |\mathcal{F}_{C'}| = \gamma |C'| = \gamma \cdot \frac{l}{\gamma} = l. \tag{8}$$

Notice that the algorithm \mathcal{A} has the following property: for any two integers k_1 and k_2 with $k_1 < k_2$, $|\mathcal{F}_{k_1}| \leq |\mathcal{F}_{k_2}|$. Run algorithm \mathcal{A} on instances $(E, \mathcal{S}, k/(2\gamma^2))$ for $k = 1$ up to $|E|$, and let k^* be the largest integer satisfying $|\mathcal{F}_{k^*}| \leq l$. By Eqs. (3) and (4),

$$|C_{\mathcal{F}_{k^*}}| \leq l \text{ and } |E(G[C_{\mathcal{F}_{k^*}}])| \geq \frac{k^*}{2\gamma^2}.$$

Since k^* is the largest index with $|\mathcal{F}_{k^*}| \leq l$ and \mathcal{F}_{opt_D} also satisfies $|\mathcal{F}_{opt_D}| \leq l$ (see inequality (8)), so $k^* \geq opt_D$. It follows that $C_{\mathcal{F}_{k^*}}$ is a feasible solution to D/S with

$$|E(G[C_{\mathcal{F}_{k^*}}])| \geq \frac{opt_D}{2\gamma^2}.$$

Then the theorem follows directly from Claim 2. □

Up to now, the best known performance ratio for D/S is $O(n^{\frac{1}{4}+\epsilon})$, where $\epsilon > 0$ is an arbitrary real number (Bhaskara et al. 2010). Very recently, Manurangsi (2017) showed that D/S has no $n^{\frac{1}{(\log \log n)^c}}$ -approximation assuming the exponential time hypothesis (ETH), where $c > 0$ is a constant independent of n . Hence we have the following corollary.

Corollary 3.2 *MinPSMC cannot be approximated within factor $O(n^{\frac{1}{2(\log \log n)^c}})$ under the ETH assumption.*

4 Primal-dual algorithm

4.1 Linear program for MinPSMC

Before we give the integer program formulation for MinPSMC, we introduce some notations. For an element $e \in E$, an r_e -cover set is a sub-collection $\mathcal{A} \subseteq \mathcal{S}$ with $|\mathcal{A}| = r_e$ such that $e \in S$ for every $S \in \mathcal{A}$. Denote by Ω_e the family of all r_e -cover sets and $\Omega = \cup_{e \in E} \Omega_e$. The following example illustrates these concepts.

Example 4.1 Let $E = \{e_1, e_2, e_3\}$, $\mathcal{S} = \{S_1, S_2, S_3\}$ with $S_1 = \{e_1, e_2\}$, $S_2 = \{e_1, e_3\}$, $S_3 = \{e_1, e_2, e_3\}$, $r(e_i) = 2$ for $i = 1, 2, 3$. For this example, $\Omega_{e_1} = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ with $\mathcal{A}_1 = \{S_1, S_2\}^{e_1}$, $\mathcal{A}_2 = \{S_1, S_3\}^{e_1}$, $\mathcal{A}_3 = \{S_2, S_3\}^{e_1}$, $\Omega_{e_2} = \{\mathcal{A}_4\}$ with $\mathcal{A}_4 = \{S_1, S_3\}^{e_2}$, $\Omega_{e_3} = \{\mathcal{A}_5\}$ with $\mathcal{A}_5 = \{S_2, S_3\}^{e_3}$, and $\Omega = \{\mathcal{A}_1, \dots, \mathcal{A}_5\}$.

Remark 4.2 Notice that different elements may have a same collection of sets as an r_e -cover set. For the above example, $\{S_1, S_3\}$ is an r_{e_1} -cover set as well as an r_{e_2} -cover set. In this case, this collection of sets should be viewed as different r_e -cover sets. We use superscript in the above example to distinguish them. The idea behind this definition is that if an r_e -cover set $\mathcal{A} \in \Omega$ is taken, then e is fully covered by those sets in \mathcal{A} .

For a sub-family $\Omega' \subseteq \Omega$, let

$$\mathcal{S}_{\Omega'} = \bigcup_{\mathcal{A} \in \Omega'} \mathcal{A}$$

be the sub-collection of \mathcal{S} consisting of those sets which belong to some cover set of Ω' . For an instance, in the above example, if $\Omega' = \{\mathcal{A}_3, \mathcal{A}_5\}$, then $\mathcal{S}_{\Omega'} = \{S_2, S_3\}$ (the superscripts are ignored while taking the union).

Let $\rho : 2^\Omega \mapsto \mathbb{R}$ be a function on sub-families of Ω defined by

$$\rho(\Omega') = \sum_{S \in \mathcal{S}_{\Omega'}} w_S \tag{9}$$

for $\Omega' \subseteq \Omega$. For the above example with $\Omega' = \{\mathcal{A}_3, \mathcal{A}_5\}$, we have $\rho(\Omega') = w_{S_2} + w_{S_3}$.

Lemma 4.3 *The function ρ defined in (9) is a nonnegative monotone nondecreasing submodular function on 2^Ω .*

Proof The nonnegativity and the monotonicity are obvious. To show the submodularity, consider two sub-families $\Omega_1 \subseteq \Omega_2 \subseteq \Omega$ and a cover set $\mathcal{B} \in \Omega \setminus \Omega_2$. Since $\mathcal{B} \setminus (\cup_{\mathcal{A} \in \Omega_2} \mathcal{A}) \subseteq \mathcal{B} \setminus (\cup_{\mathcal{A} \in \Omega_1} \mathcal{A})$, we have

$$\Delta_{\mathcal{B}} \rho(\Omega_2) \geq \Delta_{\mathcal{B}} \rho(\Omega_1).$$

By Definition 2.1, ρ is submodular. □

The MinPSMC problem can be formulated as an integer program as follows. We assign a binary variable $x_{\mathcal{A}} \in \{0, 1\}$ to each cover set $\mathcal{A} \in \Omega$ which takes value 1 if and only if \mathcal{A} is selected. A binary variable $y_e \in \{0, 1\}$ is assigned to each element $e \in E$ which takes value 1 if and only if e is *not* fully covered. For an element $e \in E$, if there exists at least one r_e -cover set $\mathcal{A} \in \Omega_e$ with $x_{\mathcal{A}} = 1$, then e is fully covered. Hence the first constraint of the following integer program indicates that if e is not fully covered (that is $x_{\mathcal{A}} = 0$ for all $\mathcal{A} \in \Omega_e$), then $y_e = 1$. The second constraint indicates that there are at most $n - k$ elements which are not fully covered.

$$\begin{aligned} \min \quad & \rho \left(\bigcup_{\mathcal{A}: x_{\mathcal{A}}=1} \mathcal{A} \right) \\ \text{s.t.} \quad & \begin{cases} \sum_{\mathcal{A} \in \Omega_e} x_{\mathcal{A}} + y_e \geq 1, & \forall e \in E, \\ \sum_{e \in E} y_e \leq n - k \\ x_{\mathcal{A}} \in \{0, 1\}, & \forall \mathcal{A} \in \Omega, \\ y_e \in \{0, 1\}, & \forall e \in E, \end{cases} \end{aligned} \tag{10}$$

The corresponding integer program relaxation is to relax the domain of $x_{\mathcal{A}}$ and y_e allowing real numbers between 0 and 1. Since the right hand side of the first constraint is 1, so the upper bound 1 on $x_{\mathcal{A}}$ and y_e can be dropped. The objective function is relaxed to its Lovász extension function $\hat{\rho}$.

$$\begin{aligned} \min \quad & \hat{\rho}(\mathbf{x}) \\ \text{s.t.} \quad & \begin{cases} \sum_{\mathcal{A} \in \Omega_e} x_{\mathcal{A}} + y_e \geq 1, & \forall e \in E, \\ \sum_{e \in E} y_e \leq n - k \\ x_{\mathcal{A}} \geq 0, & \forall \mathcal{A} \in \Omega, \\ y_e \geq 0, & \forall e \in E, \end{cases} \end{aligned} \tag{11}$$

Since ρ is a submodular function, $\hat{\rho}$ is a convex function by Theorem 2.3. So, (11) is a convex program. Next, we linearize it.

By Theorem 2.5, $\hat{\rho}(\mathbf{x})$ coincides with the convex closure $\rho^-(\mathbf{x})$, which is a solution to the following linear program (see Definition 2.4):

$$\begin{aligned} \min \quad & \sum_{\Omega': \Omega' \subseteq \Omega} \rho(\Omega') \xi_{\Omega'} \\ \text{s.t.} \quad & \begin{cases} \sum_{\mathcal{A} \in \Omega'} \xi_{\Omega'} = x_{\mathcal{A}}, & \forall \mathcal{A} \in \Omega, \\ \sum_{\Omega' \subseteq \Omega} \xi_{\Omega'} = 1, \\ \xi_{\Omega'} \geq 0, & \forall \Omega' \subseteq \Omega \end{cases} \end{aligned} \tag{12}$$

Substituting (12) into (11), we have the following relaxed linear program for MinPSMC:

$$\begin{aligned}
 \min \quad & \sum_{\Omega': \Omega' \subseteq \Omega} \rho(\Omega') \xi_{\Omega'} \\
 \text{s.t.} \quad & \begin{cases} \sum_{\Omega': \mathcal{A} \in \Omega'} \xi_{\Omega'} = x_{\mathcal{A}}, & \forall \mathcal{A} \in \Omega, \\ \sum_{\mathcal{A}: \mathcal{A} \in \Omega_e} x_{\mathcal{A}} + y_e \geq 1, & \forall e \in E, \\ \sum_{e \in E} y_e \leq n - k \\ \xi_{\Omega'} \geq 0, & \forall \Omega' \subseteq \Omega, \\ x_{\mathcal{A}} \geq 0, & \forall \mathcal{A} \in \Omega, \\ y_e \geq 0, & \forall e \in E. \end{cases} \tag{13}
 \end{aligned}$$

In obtaining this linear program, we drop off the second constraint of (12) for the simplicity of the algorithm designed below. Such a drop-off can be viewed as a further relaxation of the problem, still providing a lower bound for the optimal value of the original problem.

The dual program of (13) is:

$$\begin{aligned}
 \max \quad & \sum_{e \in E} u_e - (n - k)t \\
 \text{s.t.} \quad & \begin{cases} \sum_{\mathcal{A}: \mathcal{A} \in \Omega'} z_{\mathcal{A}} \leq \rho(\Omega'), & \forall \Omega' \subseteq \Omega, \\ \sum_{e: \mathcal{A} \in \Omega_e} u_e \leq z_{\mathcal{A}}, & \forall \mathcal{A} \in \Omega, \\ u_e \leq t, & \forall e \in E, \\ u_e \geq 0, & \forall e \in E, \\ t \geq 0 \end{cases} \tag{14}
 \end{aligned}$$

Remark 4.4 Suppose $\mathbf{z} \in \mathbb{R}^{\Omega}$ is a vector whose components are indexed by elements of Ω . For a sub-family $\Omega' \subseteq \Omega$, denote

$$z(\Omega') = \sum_{\mathcal{A} \in \Omega'} z_{\mathcal{A}}.$$

Using such a notation, the first constraint of (14) can be abbreviated as $z(\Omega') \leq \rho(\Omega')$. For any vector $\mathbf{z} \in P(\rho)$, sub-family $\Omega' \subseteq \Omega$ is said to be *z-tight* if $z(\Omega') = \rho(\Omega')$.

Remark 4.5 By Remark 4.2, for every $\mathcal{A} \in \Omega$, there exists a *unique* element e such that $\mathcal{A} \in \Omega_e$ (we say that \mathcal{A} belongs to e , denoted as $\mathcal{A} \vdash e$). For example, consider set $\mathcal{A}_3 = \{S_2, S_3\}^{e_1}$ and $\mathcal{A}_5 = \{S_2, S_3\}^{e_3}$ in Example 4.1, although they are identical as sets, they are viewed as different sets in Ω , which belongs to e_1 and e_3 , respectively. So, the second constraint of (14) is in fact $u_e \leq z_{\mathcal{A}}$ for $\mathcal{A} \vdash e$.

Remark 4.6 Notice that $|\Omega| = \sum_{e \in E} \binom{f_e}{r_e} \leq B|E|$. Hence the number of dual variables in (14) (namely $z_{\mathcal{A}}$ for $\mathcal{A} \in \Omega$, u_e for $e \in E$, and t) is at most $(B + 1)|E| + 1$, which is polynomial under the assumption that r_{\max} is a constant.

4.2 The algorithm

Recall that $\mathcal{C}(\mathcal{F})$ denotes the set of elements fully covered by sub-collection $\mathcal{F} \subseteq \mathcal{S}$. This notation can be naturally extended for a sub-family of cover sets: for $\Omega' \subseteq \Omega$, let $\mathcal{C}(\Omega') = \mathcal{C}(\mathcal{S}_{\Omega'})$ be the set of elements fully covered by Ω' .

The algorithm is formally described in Algorithm 1. It keeps a feasible solution $(\{u_e\}_{e \in E}, \{z_{\mathcal{A}}\}_{\mathcal{A} \in \Omega}, t)$ of program (14) and a sub-family $\Gamma \subseteq \Omega$. Initially, all dual variables are set to be zeros, and Γ is set to be \emptyset . All dual variables are *active* at the beginning. While \mathcal{S}_{Γ} is not a feasible solution to MinPSMC, the algorithm uniformly increases active dual variables until a sub-family $\Omega' \subseteq \Omega \setminus \Gamma$ becomes \mathbf{z} -tight. We shall prove later that this process can be accomplished in polynomial time. Then the algorithm updates $\Gamma \leftarrow \Gamma \cup \Omega'$ and *deactivates* some dual variables, where a dual variable is deactivated means that its value will not increase any more. To be more concrete, when e is *newly* fully covered by Ω' , neither u_e nor $z_{\mathcal{A}}$ (for any $\mathcal{A} \in \Omega_e$) can continue to increase their values in latter iterations. Such an operation is executed iteratively until \mathcal{S}_{Γ} is a feasible solution to MinPSMC. In the algorithm, we use Γ_j to represent the sub-family Γ , set E_j to represent the set of elements which are not fully covered, and family Ω_j to represent the sub-family of cover-sets consisting of those Ω_e with e not fully covered, at the end of the j th iteration. We also use $\Omega^{(j)}$ to represent the sub-family Ω' which becomes \mathbf{z} -tight in the j th iteration. Suppose the while loop is executed g rounds. One problem is that the last sub-family $\Omega^{(g)}$ might fully cover too many elements the number of which is much more than required by the remaining covering requirement after the $(g - 1)$ th iteration. To overcome such a problem, the algorithm continues to construct another sub-family $\tilde{\Omega}$ satisfying the remaining covering requirement, and outputs the better solution of $\Gamma_{g-1} \cup \Omega^{(g)}$ and $\Gamma_{g-1} \cup \tilde{\Omega}$ (notice that the index j in the second while loop, having the value when the algorithm jumps out of the first while loop, is g). The sub-family $\tilde{\Omega}$ is constructed by iteratively selecting a cover set minimizing ρ until the union of selected cover sets satisfies the remaining covering requirement.

Lemma 4.7 *The running time of the above algorithm is polynomial.*

Proof By Remark 4.6, both $|\Omega|$, $|E|$ and the number of dual variables are polynomial. So, to prove the lemma, it suffices to show that line 4 and line 5 of Algorithm 1 can be accomplished in polynomial time (notice that the number of $\Omega' \subseteq \Omega$ is exponential). Notice that the objective to be minimized is a submodular function divided by a modular function. Such an objective can be minimized in polynomial time (Fleisher and Iwata 2003). \square

Remark 4.8 For a sub-family Ω , denote by $E(\Omega')$ the set of elements indexing those cover sets of Ω' . Notice that $|\mathcal{C}(\Omega')|$ might be larger than $|E(\Omega')|$. For example, in

Algorithm 1 $PD(E, \Omega, \rho, r, k)$

Input: An instance (E, Ω, ρ, r, k) of MinPSMC.
Output: A sub-collection \mathcal{T} fully covering at least k elements.
1: $j \leftarrow 0, \Omega_0 \leftarrow \Omega, E_0 \leftarrow E, \Gamma_0 \leftarrow \emptyset.$
 $t \leftarrow 0, z_{\mathcal{A}} \leftarrow 0$ for every $\mathcal{A} \in \Omega_0$ and $u_e \leftarrow 0$ for every $e \in E.$
2: **while** $|\mathcal{C}(\Gamma_j)| < k$ **do**
3: $j \leftarrow j + 1.$
4: $\alpha_j \leftarrow \min\{\frac{\rho(\Omega') - z(\Omega')}{|\Omega'|} : |\Omega'| \geq 1, \Omega' \subseteq \Omega_{j-1}\}.$
5: $\Omega^{(j)} \leftarrow \arg \min\{\frac{\rho(\Omega') - z(\Omega')}{|\Omega'|} : |\Omega'| \geq 1, \Omega' \subseteq \Omega_{j-1}\}.$
6: $t \leftarrow t + \alpha_j$
7: For each $\mathcal{A} \in \Omega_{j-1}, z_{\mathcal{A}} \leftarrow z_{\mathcal{A}} + \alpha_j$
8: For each $e \in E_{j-1}, u_e \leftarrow u_e + \alpha_j.$
9: $\Gamma_j \leftarrow \Gamma_{j-1} \cup \Omega^{(j)}.$
10: $E_j \leftarrow E_{j-1} \setminus \{e : e \text{ is newly fully covered}\}$
11: $\Omega_j \leftarrow \Omega_{j-1} \setminus \{\Omega_e : e \text{ is newly fully covered}\}$
12: **end while**
13: $\tilde{\Omega} \leftarrow \emptyset.$
14: **while** $|\mathcal{C}(\Gamma_{j-1} \cup \tilde{\Omega})| < k$ **do**
15: $\mathcal{A} \leftarrow \arg \min_{\mathcal{A} \in \Omega_{j-1}} \rho(\mathcal{A})$
16: $\tilde{\Omega} \leftarrow \tilde{\Omega} \cup \mathcal{A}$
17: $\Omega_{j-1} \leftarrow \Omega_{j-1} \setminus \{\Omega_e : e \text{ is newly fully covered}\}$
18: **end while**
19: $\Lambda \leftarrow \arg \min\{\rho(\Gamma_{j-1} \cup \Omega^{(j)}), \rho(\Gamma_{j-1} \cup \tilde{\Omega})\}$
20: Output $\mathcal{T} \leftarrow \mathcal{S}_{\Lambda}$

Example 4.1, if $\Omega' = \{\mathcal{A}_3\}$, then $E(\Omega') = \{e_1\}$ while $\mathcal{C}(\Omega') = \{e_1, e_3\}$. However, the sub-family $\Omega^{(j)}$ found in line 5 of Algorithm 1 always satisfies

$$|\mathcal{C}(\Omega^{(j)})| = |E(\Omega^{(j)})|. \tag{15}$$

In fact, notice that for any cover set $\mathcal{A} \in \Omega_{j-1}$, the dual variable $z_{\mathcal{A}}$ is active. Hence $z_{\mathcal{A}} = t$ for any $\mathcal{A} \in \Omega_{j-1}$. It follows that

$$\frac{\rho(\Omega') - z(\Omega')}{|\Omega'|} = \frac{\rho(\Omega')}{|\Omega'|} - t. \tag{16}$$

So, for two sub-families Ω' and Ω'' with $\rho(\Omega') = \rho(\Omega'')$, the one with larger cardinality will be preferred (since line 5 chooses the one with the minimum ratio). Hence

$$\Omega^{(j)} \text{ must include all cover sets which are subsets of } \mathcal{S}_{\Omega^{(j)}}. \tag{17}$$

In order words, if e is newly covered by $\mathcal{S}_{\Omega^{(j)}}$, then there is an r_e -cover set which is a subset of $\mathcal{S}_{\Omega^{(j)}}$, adding such an r_e -cover set into $\Omega^{(j)}$ will not change the value of $\rho(\Omega^{(j)})$ while the cardinality of $|\Omega^{(j)}|$ will be larger. Consider Example 4.1, if we choose $\Omega' = \{\mathcal{A}_3\}$, then $\mathcal{S}_{\Omega'} = \{\mathcal{S}_2, \mathcal{S}_3\}$, which can fully cover e_1 and e_3 . In this case, adding $\mathcal{A}_5 = \{\mathcal{S}_2, \mathcal{S}_3\}^{e_3}$ into Ω' will lead to a smaller ratio. Then claim (15) follows from (17).

4.3 Performance analysis

First, we show that throughout the algorithm, a dual feasible solution is maintained.

Lemma 4.9 *Algorithm 1 maintains the feasibility of (14).*

Proof Notice that when a sub-family Ω' becomes \mathbf{z} -tight, for every $\mathcal{A} \in \Omega'$, $z_{\mathcal{A}}$ is deactivated. Hence

$$\text{a } \mathbf{z}\text{-tight set will remain to be } \mathbf{z}\text{-tight to the end of the algorithm.} \quad (18)$$

For those sub-families which are not \mathbf{z} -tight, the choice of α in line 4 guarantees that the first constraint of (14) is not violated.

Since all active dual variables increase at the same rate, and for any element e which is fully covered in some iteration, u_e and every $\mathcal{A} \in \Omega_e$ are deactivated at the same time (namely, the time when e is newly fully covered), hence

$$u_e = z_{\mathcal{A}} \text{ holds for every } \mathcal{A} \in \Omega_e \text{ throughout of the algorithm.} \quad (19)$$

Combining this with Remark 4.5, the second constraint of (14) holds.

Since u_e increases at the same rate as t until it is deactivated, hence $u_e \leq t$ holds for every element e , and

$$u_e = t \text{ for every } e \text{ which is not fully covered yet.} \quad (20)$$

The third constraint of (14) is maintained. The lemma is proved. \square

Denote by $(\{u_e^{(j)}\}_{e \in E}, \{z_{\mathcal{A}}^{(j)}\}_{\mathcal{A} \in \Omega}, t^{(j)})$ the dual feasible solution after the j th iteration.

Lemma 4.10 *For any index j , the sub-family Γ_j satisfies $z^{(j)}(\Gamma_j) = \rho(\Gamma_j)$.*

Proof We prove the lemma by induction on j . This is obvious for $j = 0$ since $\Gamma_0 = \emptyset$ and $\rho(\emptyset) = z^{(0)}(\emptyset) = 0$. Assume that $j \geq 1$ and the lemma is true for $j - 1$. In the j th iteration, a sub-family $\Omega^{(j)}$ is found, and Γ_j is set to be $\Gamma_{j-1} \cup \Omega^{(j)}$. Notice that $\Gamma_{j-1} \cap \Omega^{(j)} = \emptyset$. By the choice of α_j in line 4 and the update of \mathbf{z} in line 7, we have

$$\rho(\Omega^{(j)}) = z^{(j-1)}(\Omega^{(j)}) + \alpha_j |\Omega^{(j)}| = \sum_{\mathcal{A} \in \Omega^{(j)}} (z_{\mathcal{A}}^{(j-1)} + \alpha_j) = \sum_{\mathcal{A} \in \Omega^{(j)}} z_{\mathcal{A}}^{(j)} = z^{(j)}(\Omega^{(j)}).$$

Since $z_{\mathcal{A}}$ is deactivated when \mathcal{A} is chosen into Γ_{j-1} , we have $z_{\mathcal{A}}^{(j)} = z_{\mathcal{A}}^{(j-1)}$ for every $\mathcal{A} \in \Gamma_{j-1}$. Hence

$$z^{(j)}(\Gamma_{j-1}) = z^{(j-1)}(\Gamma_{j-1}) = \rho(\Gamma_{j-1}),$$

where the second equality comes from induction hypothesis. It follows that

$$\rho(\Gamma_{j-1}) + \rho(\Omega^{(j)}) = z^{(j)}(\Gamma_{j-1}) + z^{(j)}(\Omega^{(j)}) = z^{(j)}(\Gamma_j). \quad (21)$$

By the feasibility of $z^{(j)}$, we have

$$z^{(j)}(\Gamma_j) \leq \rho(\Gamma_j). \tag{22}$$

By the submodularity of ρ and the observation that $\Gamma_{j-1} \cap \Omega^{(j)} = \emptyset$ (and thus $\rho(\Gamma_{j-1} \cap \Omega^{(j)}) = 0$), we have

$$\rho(\Gamma_j) = \rho(\Gamma_{j-1} \cup \Omega^{(j)}) + \rho(\Gamma_{j-1} \cap \Omega^{(j)}) \leq \rho(\Gamma_{j-1}) + \rho(\Omega^{(j)}). \tag{23}$$

Combining (21), (22), and (23), we have $z^{(j)}(\Gamma_j) = \rho(\Gamma_j)$. The induction step is finished and the lemma is proved. \square

Remark 4.11 For any sub-collection of sets $\mathcal{F} \subseteq \mathcal{S}$, we can rewrite \mathcal{F} as a family $\Psi_{\mathcal{F}}$ of cover sets. The following example illustrates how this can be done. Let $E = \{e_1, e_2, e_3, e_4\}$, $\mathcal{S} = \{S_1, S_2, S_3, S_4\}$ with $S_1 = \{e_1, e_2, e_4\}$, $S_2 = \{e_2, e_3\}$, $S_3 = \{e_1, e_2, e_3\}$, $S_4 = \{e_3, e_4\}$, and $r(e_i) = 2$ for $i = 1, 2, 3, 4$. For this example, $\mathcal{A}_1 = \{S_1, S_3\}^{e_1}$, $\mathcal{A}_2 = \{S_1, S_2\}^{e_2}$, $\mathcal{A}_3 = \{S_1, S_3\}^{e_2}$, $\mathcal{A}_4 = \{S_2, S_3\}^{e_2}$, $\mathcal{A}_5 = \{S_2, S_3\}^{e_3}$, $\mathcal{A}_6 = \{S_2, S_4\}^{e_3}$, $\mathcal{A}_7 = \{S_3, S_4\}^{e_3}$, $\mathcal{A}_8 = \{S_1, S_4\}^{e_4}$. If $\mathcal{F} = \{S_2, S_3, S_4\}$, then $\Psi_{\mathcal{F}} = \{\mathcal{A}_4, \mathcal{A}_5, \mathcal{A}_6, \mathcal{A}_7\}$. In general, $\Psi_{\mathcal{F}}$ consists of all those cover sets which are subsets of \mathcal{F} .

Theorem 4.12 Algorithm 1 has performance ratio at most $B + \sqrt{n \cdot B}$.

Proof Let $\Omega^{(1)}, \dots, \Omega^{(g)}$ be the sequence of sub-families of cover sets found by Algorithm 1. Then $\Gamma_j = \cup_{l=1}^j \Omega^{(l)}$ for $j = 1, \dots, g$. Denote by $\mathcal{U}(\Gamma_j)$ the set of elements which are not fully covered by \mathcal{S}_{Γ_j} . Denote by opt the optimal value of the MinPSMC instance. \square

Claim 1 $\rho(\Gamma_{g-1}) \leq B \cdot opt$.

For any index j , it can be calculated that

$$\begin{aligned} \rho(\Gamma_j) &= z^{(j)}(\Gamma_j) = \sum_{\mathcal{A} \in \Gamma_j} z_{\mathcal{A}}^{(j)} \\ &= \sum_{\mathcal{A} \in \Gamma_j} \sum_{e: \mathcal{A} \vdash e} u_e^{(j)} \\ &= \sum_{e: e \in \mathcal{C}(\Gamma_j)} u_e^{(j)} \cdot |\{\mathcal{A} \in \Gamma_j : \mathcal{A} \vdash e\}| \\ &\leq B \sum_{e: e \in \mathcal{C}(\Gamma_j)} u_e^{(j)} \\ &= B \left(\sum_{e: e \in E} u_e^{(j)} - \sum_{e: e \in \mathcal{U}(\Gamma_j)} u_e^{(j)} \right) \\ &= B \left(\sum_{e: e \in E} u_e^{(j)} - |\mathcal{U}(\Gamma_j)|t^{(j)} \right), \end{aligned} \tag{24}$$

where the first equality comes from Lemma 4.10; the third equality comes from remark 4.5 and property (19); the inequality holds because the number of sets belonging to e is at most B ; the last equality comes from (20).

In particular, taking $j = g - 1$ in inequality (24),

$$\rho(\Gamma_{g-1}) \leq B \left(\sum_{e: e \in E} u_e^{(g-1)} - |\mathcal{U}(\Gamma_{g-1})|t^{(g-1)} \right). \tag{25}$$

Since the algorithm does not jump out of the while loop at the $(g - 1)$ th iteration, $|\mathcal{C}(\Gamma_{g-1})| < k$ and thus $|\mathcal{U}(\Gamma_{g-1})| > n - k$. Combining this with inequality (25) and the weak duality theorem for linear programs, we have

$$\rho(\Gamma_{g-1}) \leq B \left(\sum_{e: e \in E} u_e^{(g-1)} - (n - k)t^{(g-1)} \right) = B \cdot \text{obj}_D^{(g-1)} \leq B \cdot \text{opt},$$

where $\text{obj}_D^{(g-1)}$ is the objective value of dual program (14) for those dual variables after the $(g - 1)$ th iteration. Claim 1 is proved.

Claim 2 $\rho(\Omega^{(g)}) \leq \frac{(n - |\mathcal{C}(\Gamma_{g-1})|)}{k - |\mathcal{C}(\Gamma_{g-1})|} \cdot B \cdot \text{opt}$.

Let Ψ be the family of cover sets constructed from an optimal solution by the method described in Remark 4.11. Let

$$\Psi_{g-1} = \Psi \setminus \Gamma_{g-1}. \tag{26}$$

Notice that $\Psi_{g-1} \neq \emptyset$, since otherwise $\Psi \subseteq \Gamma_{g-1}$ which contradicts $|\mathcal{C}(\Gamma_{g-1})| < k$.

In the g th iteration, $\Omega^{(g)}$ is chosen, which means that

$$\frac{\rho(\Omega^{(g)}) - z^{(g-1)}(\Omega^{(g)})}{|\Omega^{(g)}|} \leq \frac{\rho(\Psi_{g-1}) - z^{(g-1)}(\Psi_{g-1})}{|\Psi_{g-1}|}. \tag{27}$$

By (16) in Remark 4.8, we have

$$\rho(\Omega^{(g)}) \leq \frac{|\Omega^{(g)}|}{|\Psi_{g-1}|} \rho(\Psi_{g-1}). \tag{28}$$

Then Claim 2 follows from the observation that

$$\begin{aligned} |\Omega^{(g)}| &\leq B \cdot (n - |\mathcal{C}(\Gamma_{g-1})|), \\ |\Psi_{g-1}| &\geq k - |\mathcal{C}(\Gamma_{g-1})|, \\ \rho(\Psi_{g-1}) &\leq \rho(\Psi) = \text{opt}, \end{aligned} \tag{29}$$

where the first inequality holds because there are $n - |\mathcal{C}(\Gamma_{g-1})|$ elements which are not fully covered by Γ_{g-1} , and each element is contained in at most B cover sets; the

second inequality holds because Ψ_{g-1} fully covers at least $k - |\mathcal{C}(\Omega_{g-1})|$ elements and each element has at least one cover set in Ψ_{g-1} .

Claim 3 $\rho(\tilde{\Omega}) \leq (k - |\mathcal{C}(\Gamma_{g-1})|) \cdot opt$.

This claim is obvious by observing that the second while loop picks at most $k - |\mathcal{C}(\Gamma_{g-1})|$ cover sets and every cover set picked has cost upper bounded by opt .

Combing Claims 2 and 3, the last sub-family has cost at most

$$\min \left\{ \frac{(n - |\mathcal{C}(\Gamma_{g-1})|)}{k - |\mathcal{C}(\Gamma_{g-1})|} \cdot B \cdot opt, (k - |\mathcal{C}(\Gamma_{g-1})|) \cdot opt \right\} \leq \sqrt{n \cdot B} \cdot opt. \tag{30}$$

Then the theorem follows from the combination of inequality (30) and Claim 1. \square

5 Improvement on restricted MinPSMC

In the restricted problem MinRPSMC, $f_e \equiv 2$ and $r_e \equiv 2$, so $B = 1$ by the definition of B in (2). Then it follows from Theorem 4.12 that Algorithm 1 has performance ratio $1 + \sqrt{n}$ for MinRPSMC. In this section, we present a modified algorithm achieving performance ratio $1 + \sqrt{2}n^{1/4}$ for MinRPSMC.

This restricted problem can be considered from a graph theoretical point of view. Since $f_e \equiv 2$, we may regard every set in \mathcal{S} as a vertex and every element in E as an edge which is incident with the two vertices corresponding to the two sets of \mathcal{S} containing e . Denote by G the graph on vertex set \mathcal{S} and edge set E under such a point of view. For a sub-family Ω' , denote by $V_{\Omega'}$ the set of vertices corresponding to the sets in $\mathcal{S}_{\Omega'}$, and denote by $E_{\Omega'}$ the edge set of the subgraph of G induced by vertex set $V_{\Omega'}$. Then it can be seen that $E_{\Omega'}$ is exactly the set of elements fully covered by $\mathcal{S}_{\Omega'}$. In other words,

$$|\mathcal{C}(\Omega')| = |E_{\Omega'}|. \tag{31}$$

By the definition of $\rho(\Omega')$ in (9), and because $w_S \equiv 1$ in this restricted problem,

$$\rho(\Omega') = |V_{\Omega'}|. \tag{32}$$

The modified algorithm for the restricted problem is presented in Algorithm 2. Compared with Algorithm 1, instead of outputting the better solution of $\Gamma_{g-1} \cup \Omega^{(g)}$ and $\Gamma_{g-1} \cup \tilde{\Omega}$, the modified algorithm outputs the better solution of $\Gamma_{g-1} \cup \hat{\Omega}$ and $\Gamma_{g-1} \cup \tilde{\Omega}$, where g is the value of index j when the algorithm jumps out of the first while loop, and $\hat{\Omega}$ is obtained from line 13 to line 17. The algorithm uses a parameter

$$\gamma = \sqrt{\frac{|\mathcal{C}(\Omega^{(g)})|}{2(k - |\mathcal{C}(\Gamma_{g-1})|)}}. \tag{33}$$

One question is whether the sub-family $\hat{\Omega}$ in line 16 of the algorithm exists and can be found in polynomial time? The following lemma will answer this question.

Algorithm 2 $PD(E, \Omega, \rho, r, k)$

Input: An instance (E, Ω, ρ, r, k) of MinRPSMC.

Output: A sub-collection \mathcal{T} fully covering at least k elements.

```

1:  $j \leftarrow 0, \Omega_0 \leftarrow \Omega, E_0 \leftarrow E, \Gamma_0 \leftarrow \emptyset.$ 
    $t \leftarrow 0, z_{\mathcal{A}} \leftarrow 0$  for every  $\mathcal{A} \in \Omega_0$  and  $u_e \leftarrow 0$  for every  $e \in E.$ 
2: while  $|\mathcal{C}(\Gamma_j)| < k$  do
3:    $j \leftarrow j + 1.$ 
4:    $\alpha_j \leftarrow \min\{\frac{\rho(\Omega') - z(\Omega')}{|\Omega'|} : |\Omega'| \geq 1, \Omega' \subseteq \Omega_{j-1}\}.$ 
5:    $\Omega^{(j)} \leftarrow \arg \min\{\frac{\rho(\Omega') - z(\Omega')}{|\Omega'|} : |\Omega'| \geq 1, \Omega' \subseteq \Omega_{j-1}\}.$ 
6:    $t \leftarrow t + \alpha_j$ 
7:   For each  $\mathcal{A} \in \Omega_{j-1}, z_{\mathcal{A}} \leftarrow z_{\mathcal{A}} + \alpha_j$ 
8:   For each  $e \in E_{j-1}, u_e \leftarrow u_e + \alpha_j.$ 
9:    $\Gamma_j \leftarrow \Gamma_{j-1} \cup \Omega^{(j)}.$ 
10:   $E_j \leftarrow E_{j-1} \setminus \{e : e \text{ is newly fully covered}\}$ 
11:   $\Omega_j \leftarrow \Omega_{j-1} \setminus \{\Omega_e : e \text{ is newly fully covered}\}$ 
12: end while
13: if  $|\mathcal{C}(\Omega^{(j)})| \leq 2(k - |\mathcal{C}(\Gamma_{j-1})|)$  then
14:   $\widehat{\Omega} \leftarrow \Omega^{(j)}$ 
15: else
16:   $\widehat{\Omega} \leftarrow$  a sub-family of  $\Omega^{(j)}$  with  $\rho(\widehat{\Omega}) \leq \rho(\Omega^{(g)})/\gamma$  and  $|\mathcal{C}(\widehat{\Omega})| \geq k - |\mathcal{C}(\Gamma_{j-1})|$ 
17: end if
18:  $\widetilde{\Omega} \leftarrow \emptyset.$ 
19: while  $|\mathcal{C}(\Gamma_{j-1} \cup \widetilde{\Omega})| < k$  do
20:   $\mathcal{A} \leftarrow \arg \min_{\mathcal{A} \in \Omega_{j-1}} \rho(\mathcal{A})$ 
21:   $\widetilde{\Omega} \leftarrow \widetilde{\Omega} \cup \mathcal{A}$ 
22:   $\Omega_{j-1} \leftarrow \Omega_{j-1} \setminus \{\Omega_e : e \text{ is newly fully covered}\}$ 
23: end while
24:  $\Lambda \leftarrow \arg \min\{\rho(\Gamma_{j-1} \cup \widehat{\Omega}), \rho(\Gamma_{j-1} \cup \widetilde{\Omega})\}$ 
25: Output  $\mathcal{T} \leftarrow \mathcal{S}_{\Lambda}$ 

```

Using the notation in the proof of Theorem 4.12, suppose the first while loop is executed g times, and the sub-collections found in line 5 of Algorithm 2 are $\Omega^{(1)}, \dots, \Omega^{(g)}$.

Lemma 5.1 *In the case $|\mathcal{C}(\Omega^{(g)})| > 2(k - |\mathcal{C}(\Gamma_{g-1})|)$, one can find a sub-family of cover sets $\widehat{\Omega}$ in polynomial time with $\rho(\widehat{\Omega}) \leq \rho(\Omega^{(g)})/\gamma$ and $|\mathcal{C}(\widehat{\Omega})| \geq k - |\mathcal{C}(\Gamma_{g-1})|$.*

Proof Under the assumption that $|\mathcal{C}(\Omega^{(g)})| > 2(k - |\mathcal{C}(\Gamma_{g-1})|)$, we have $\gamma > 1$. Let \widehat{V} be the vertex set obtained by uniformly randomly picking $|V_{\Omega^{(g)}}|/\gamma$ vertices from $V_{\Omega^{(g)}}$ (we simplify the derivation by regarding $|V_{\Omega^{(g)}}|/\gamma$ as an integer). Similar to the proof of Theorem 3.1,

$$\mathbb{E}[|E_{\widehat{V}}|] \geq \frac{|E_{V_{\Omega^{(g)}}}|}{2\gamma^2}.$$

where $E_{\widehat{V}}$ represents the set of edges in the subgraph induced by vertex set \widehat{V} .

The construction of vertex set \widehat{V} can be derandomized using the classic method of conditional expectation satisfying

$$|\widehat{V}| \leq \frac{|V_{\Omega^{(g)}}|}{\gamma} \text{ and } |E_{\widehat{V}}| \geq \frac{|E_{V_{\Omega^{(g)}}}|}{2\gamma^2}. \tag{34}$$

Let $\widehat{\mathcal{S}}$ be the sub-collection of sets corresponding to \widehat{V} , and consider the sub-family of cover sets $\widehat{\Omega}$ which is constructed from $\widehat{\mathcal{S}}$ by the method described in Remark 4.11. By observations (31) and (32), inequality (34) is in fact

$$\rho(\widehat{\Omega}) \leq \frac{\rho(\Omega^{(g)})}{\gamma} \text{ and } |\mathcal{C}(\widehat{\Omega})| \geq \frac{|\mathcal{C}(\Omega^{(g)})|}{2\gamma^2} \geq k - |\mathcal{C}(\Gamma_{g-1})|,$$

where the last inequality follows from the definition of γ in (33). The lemma is proved. □

Next, we analyze the performance ratio.

Theorem 5.2 *Algorithm 2 has performance ratio at most $1 + \sqrt{2}n^{1/4}$.*

Proof Denote by *opt* the optimal value of the MinRPSMC instance. We estimate $\rho(\widehat{\Omega})$ and $\rho(\widetilde{\Omega})$ as follows. □

Claim 1 $\Gamma_{g-1} \cup \widehat{\Omega}$ is a feasible solution to MinRPSMC and

$$\rho(\widehat{\Omega}) \leq \max \left\{ 2, \sqrt{\frac{2|\mathcal{C}(\Omega^{(g)})|}{k - |\mathcal{C}(\Gamma_{g-1})|}} \right\} \text{opt}.$$

Since $f_e \equiv 2$ and $r_e \equiv 2$, each element e has exactly one r_e cover set. Hence for a sub-family Ω' , the number of elements indexing the cover sets of Ω' is equal to $|\Omega'|$. Then by property (15), we have

$$|\mathcal{C}(\Omega^{(g)})| = |\Omega^{(g)}|.$$

Combining this with inequalities (28) and (29), we have

$$\rho(\Omega^{(g)}) \leq \frac{|\mathcal{C}(\Omega^{(g)})|}{k - |\mathcal{C}(\Gamma_{g-1})|} \rho(\Psi_{g-1}). \tag{35}$$

In the case $|\mathcal{C}(\Omega^{(g)})| \leq 2(k - |\mathcal{C}(\Gamma_{g-1})|)$, we have $\widehat{\Omega} = \Omega^{(g)}$. Since $\Gamma_{g-1} \cup \Omega^{(g)}$ fully covers at least k elements, $\Gamma_{g-1} \cup \widehat{\Omega}$ is a feasible solution to MinRPSMC. By (35),

$$\rho(\widehat{\Omega}) \leq 2\rho(\Psi_{g-1}). \tag{36}$$

In the case $|\mathcal{C}(\Omega^{(g)})| > 2(k - |\mathcal{C}(\Gamma_{g-1})|)$, by the choice of $\widehat{\Omega}$ in line 16 of the algorithm, $|\mathcal{C}(\widehat{\Omega})| \geq k - |\mathcal{C}(\Gamma_{g-1})|$, and thus $\Gamma_{g-1} \cup \widehat{\Omega}$ is a feasible solution to MinRPSMC. Then it follows from inequality (35) that

$$\rho(\widehat{\Omega}) \leq \frac{\rho(\Omega^{(g)})}{\gamma} \leq \sqrt{\frac{2|\mathcal{C}(\Omega^{(g)})|}{k - |\mathcal{C}(\Gamma_{g-1})|}} \rho(\Psi_{g-1}). \tag{37}$$

Claim 1 follows from (36), (37), and the observation that $\rho(\Psi_{g-1}) \leq \text{opt}$.

Claim 2 $\rho(\tilde{\Omega}) \leq \sqrt{2(k - |\mathcal{C}(\Gamma_{g-1})|)}opt.$

Suppose $\tilde{\Omega}$ contains t cover sets. Since the addition of a cover set fully covers at least one more element, we have $t \leq k - |\mathcal{C}(\Gamma_{g-1})|$. Since every cover set has cardinality 2 in this restricted problem, we have $|\mathcal{S}_{\tilde{\Omega}}| \leq 2t$. It follows from (32) that

$$\rho(\tilde{\Omega}) = |V_{\tilde{\Omega}}| = |\mathcal{S}_{\tilde{\Omega}}| \leq 2(k - |\mathcal{C}(\Gamma_{g-1})|). \tag{38}$$

Let Ψ be a family of cover sets constructed from an optimal solution to the MinRPSMC instance by the method described in Remark 4.11. By (31), $|E_{\Psi}| = |\mathcal{C}(\Psi)| \geq k$. Since any graph with at least k edges has at least $\sqrt{2k}$ vertices, we have

$$opt = \rho(\Psi) = |V_{\Psi}| \geq \sqrt{2k} \geq \sqrt{2(k - |\mathcal{C}(\Gamma_{g-1})|)}.$$

Combining this inequality with (38), Claim 2 follows.

By Claims 1 and 2, when $2 \geq \sqrt{\frac{2|\mathcal{C}(\Omega^{(g)})|}{k - |\mathcal{C}(\Gamma_{g-1})|}}$, we have

$$\min\{\rho(\hat{\Omega}), \rho(\tilde{\Omega})\} \leq 2opt,$$

and when $2 < \sqrt{\frac{2|\mathcal{C}(\Omega^{(g)})|}{k - |\mathcal{C}(\Gamma_{g-1})|}}$, we have

$$\min\{\rho(\hat{\Omega}), \rho(\tilde{\Omega})\} = \sqrt{2}|\mathcal{C}(\Omega^{(g)})|^{1/4}opt \leq \sqrt{2}n^{1/4}opt.$$

where the minimum is achieved when $k - |\mathcal{C}(\Gamma_{g-1})| = \sqrt{|\mathcal{C}(\Omega^{(g)})|}$. Notice that the second upper bound is larger than the first one in general (in fact, when $n \geq 4$). Then by Claim 1 of Theorem 4.12, recalling that $B = 1$ in this restricted problem, we have

$$\rho(\Gamma) \leq (1 + \sqrt{2}n^{1/4})opt.$$

The theorem is proved.

6 Conclusion and discussion

This paper proves a lower bound for the minimum partial set multi-cover problem MinPSMC by a reduction from the densest l -subgraph problem. Then, under the assumption that the maximum covering requirement r_{\max} is upper bounded by a constant, this paper gives a $(B + \sqrt{n \cdot B})$ -approximation algorithm for MinPSMC, where $B = \max\{\binom{f}{r_e} : e \in E\} \leq f^{r_{\max}}$ and f is the maximum number of sets containing a common element. For a restricted version of MinRPSMC, the performance ratio can be improved to $1 + \sqrt{2}n^{1/4}$. From Theorem 3.1 and the fact that the current best known performance ratio for DLS is $O(n^{\frac{1}{4} + \varepsilon})$, where $\varepsilon > 0$ is an arbitrary real number (Bhaskara et al. 2010), it is natural to ask whether the performance ratio for the restricted version can be improved to $O(n^{\frac{1}{8} + \varepsilon})$?

Our algorithm depends on the assumption that r_{max} is upper bounded by a constant. Notice that for the minimum partial positive influence dominating set problem MinPPIDS which is a special case of MinPSMC, $r_{max} = \lceil \delta/2 \rceil$, where δ is the maximum degree of the graph. In a real world network which satisfies power law, δ may be of order $O(\sqrt{n})$ (Dinh et al. 2014). How to obtain an approximation algorithm for the problem without assuming a constant upper bound on r_{max} is a topic deserving further exploration.

Acknowledgements A preliminary version of this paper appears in COCOA2018. This research is supported in part by NSFC (11771013, 61751303, 11531011) and the Zhejiang Provincial Natural Science Foundation of China (LD19A010001, LY19A010018).

References

- Bar-Yehuda R, Even S (1985) A local-ratio theorem for approximating the weighted vertex cover problem. *N-Holl Math Stud* 109:27–45
- Bar-Yehuda R (2001) Using homogeneous weights for approximating the partial cover problem. *J Algorithms* 39:137–144
- Bhaskara A, Charikar M, Chlamtac E, Feige U, Vijayaraghavan A (2010) Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k -subgraph. In: STOC, pp 201–210
- Chen N (2009) On the approximability of influence in social networks. *SIAM J Discrete Math* 23(3):1400–1415 (A preliminary version appears in SODA'08, 1029–1037)
- Chvatal V (1979) A greedy heuristic for the set-covering problem. *Math Oper Res* 4:233–235
- Dinh TN, Shen Y, Nguyen DT, Thai MT (2014) On the approximability of positive influence dominating set in social networks. *J Combin Optim* 27:487–503
- Dinur I, Steurer D (2014) Analytical approach to parallel repetition. In: STOC, pp 624–633
- Edmonds J (1970) Submodular functions, matroids, and certain polyhedra. In: Guy R, Hanani H, Sauer N, Schönheim J (eds) *Combinatorial structures and their applications*. Gordon and Breach, New York, pp 69–87
- Feige U (1996) A threshold of $\ln n$ for approximating set cover. In: STOC, pp 312–318
- Fleisher L, Iwata S (2003) A push-relabel framework for submodular function minimization and applications to parametric optimization. *Discrete Appl Math* 131:311–322
- Gandhi R, Khuller S, Srinivasan A (2004) Approximation algorithms for partial covering problems. *J Algorithms* 53(1):55–84
- Hochbaum DS (1982) Approximation algorithms for the set covering and vertex cover problems. *SIAM J Comput* 11:555–556
- Johnson DS (1974) Approximation algorithms for combinatorial problems. *J Comput Syst Sci* 9:256–278
- Karp RM (1972) Reducibility among combinatorial problems. In: Miller RE, Thatcher JW (eds) *Complexity of computer computations*. Plenum Press, New York, pp 85–103
- Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: SIGKDD, pp 137–146
- Kempe D, Kleinberg J, Tardos E (2005) Influential nodes in a diffusion model for social networks. In: ICALP, pp 1127–1138
- Kearns M, Ortiz L (2003) Algorithms for interdependent security games. In: NIPS, pp 288–297
- Khot S, Regev O (2008) Vertex cover might be hard to approximate to within $2 - \epsilon$. *J Comput Syst Sci* 74(3):335–349
- Lovász L (1975) On the ratio of optimal integral and fractional covers. *Discret Math* 13:383–390
- Lovász L (1983) Submodular functions and convexity. In: Bachem A, Korte B, Grötschel M (eds) *Mathematical programming the state of the art*. Springer, Berlin, pp 235–257
- Manurangsi P (2017) Almost-polynomial ratio ETH-hardness of approximating densest k -subgraph. In: STOC, pp 19–23
- Rajagopalan S, Vazirani VV (1993) Primal-dual RNC approximation algorithms for (multi)-set (multi)-cover and covering integer programs. In: FOCS, pp 322–331

- Ran Y, Zhang Z, Du H, Zhu Y (2017a) Approximation algorithm for partial positive influence problem in social network. *J Combin Optim* 33(2):791–802
- Ran Y, Shi Y, Zhang Z (2017b) Local ratio method on partial set multi-cover. *J Combin Optim* 34:302–313
- Slavík P (1997) Improved performance of the greedy algorithm for partial cover. *Inf Process Lett* 64(5):251–254
- Vazirani VV (2001) *Approximation algorithms*. Springer, Berlin
- Wang F, Camacho E, Xu K (2009) Positive influence dominating set in online social networks. *COCOA LNCS* 5573:313–321
- Wang F, Du H, Camacho E, Xu K, Lee W, Shi Y, Shan S (2011) On positive influence dominating sets in social networks. *Theor Comput Sci* 412:265–269

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.