Tight approximation bounds for greedy frugal coverage algorithms

Ioannis Caragiannis, Christos Kaklamanis, and Maria Kyropoulou

Research Academic Computer Technology Institute and
Department of Computer Engineering and Informatics
University of Patras, 26504 Rio, Greece

Abstract. We consider the frugal coverage problem, an interesting variation of set cover defined as follows. Instances of the problem consist of a universe of elements and a collection of sets over these elements; the objective is to compute a subcollection of sets so that the number of elements it covers plus the number of sets not chosen is maximized. The problem was introduced and studied by Huang and Svitkina [7] due to its connections to the donation center location problem. We prove that the greedy algorithm has approximation ratio at least 0.782, improving a previous bound of 0.731 in [7]. We also present a further improvement that is obtained by adding a simple corrective phase at the end of the execution of the greedy algorithm. The approximation ratio achieved in this way is at least 0.806. Our analysis is based on the use of linear programs which capture the behavior of the algorithms in worst-case examples. The obtained bounds are proved to be tight.

1 Introduction

Set cover is among the most popular combinatorial optimization problems with many applications. In the classical version of the problem, we are given a universe of elements and a collection of sets over these elements and the objective is to compute a subcollection of sets of minimum size that covers all elements. The problem is known to be hard to approximate within sublogarithmic factors [5, 12] while the classical greedy algorithm achieves an almost tight approximation ratio of H_n, the n-th harmonic number, where n is the number of elements in the universe [9]. Several variations of the greedy algorithm have been proposed that improve this approximation bound by constant (additive) factors [1, 4, 10].

A different objective was recently considered by Huang and Svitkina [7]; they call the new combinatorial optimization problem *frugal coverage*. An instance of frugal coverage consists of a universe of elements and a collection of sets over these elements, and the objective is to compute a subcollection of sets so that the number of elements covered plus the number of sets not chosen is maximized. Without loss of generality, we can assume that each element belongs to at least one set of the input collection. So, the objective can be thought of as computing a subcollection that covers all elements so that the number of sets not chosen plus n (the number of elements in the universe) is maximized. Clearly, an optimal solution for set cover is also an optimal solution for frugal coverage. However, this does not have any direct implication to the approximation guarantee for frugal coverage algorithms.

A nice motivation for studying frugal coverage is the problem of locating donation centers (DCL). Instances of DCL consist of a bipartite graph $G = (A \cup L, E)$. An agent, represented by a node $a \in A$, is connected through an edge $e \in E$ to any donation center $l \in L$ she would be willing to make a donation. Every agent a has a preference ranking on the corresponding centers, and every donation center l has a capacity, meaning that it can accept at most some specific number of donations. The problem is to decide which centers to open in order to maximize the number of donations, under the restriction that an agent will only donate to her highest ranked open center. Huang and Svitkina [7] present an approximation preserving reduction from frugal coverage to the special case of DCL in which each center has unit capacity and each agent has a degree bound of 2. They also prove that the greedy algorithm has approximation ratio at least 0.731 for both problems.

We present a tight analysis of the greedy algorithm. This algorithm, starting from an empty solution, iteratively augments the solution by a set that contains the maximum number of uncovered elements until all elements are covered. We show that its approximation ratio is exactly $18/23 \approx 0.782$, improving the previous bound in [7]. This approximation guarantee can be further improved by adding a simple corrective phase at the end of the execution. Namely, we consider each set in the solution produced that included two uncovered elements when it was selected. If its removal does not leave an element uncovered, we simply remove this set from the solution. The approximation ratio obtained in this way is $54/67 \approx 0.806$. A simple instance shows that this bound is tight. We remark that, even though such a corrective phase can improve the solution obtained by the greedy algorithm with respect to the standard set cover objective, it does not improve the worst-case approximation guarantee.

Even though the algorithms we consider are purely combinatorial, our analysis is based on the use of *linear programs*. The technique can be briefly described as follows. Given an algorithm $\mathcal{A}$, we define a linear program that takes a value $f \in (0, 1)$ as a parameter. This linear program witnesses the fact that the algorithm computes an at most f-approximate solution for some instance. The constraints of the LP capture the properties of the algorithm as well as the structure of the corresponding optimal solution. Then, a lower bound of ρ on the approximation ratio of algorithm $\mathcal{A}$ follows by showing that the corresponding LP with parameter $f = \rho$ is infeasible. In order to do this, we exploit LP duality. This particular approach was recently proved to be useful for variations of set cover such as spanning star forest and color saving [2]. However, due to the different objective of frugal coverage (and, in particular, the appearance of the number of sets not chosen in the objective function), additional variables (and different constraints) have to be included in the parameterized LPs. Except from variations of set cover [1, 11], analysis of combinatorial algorithms through linear programs has also been performed in contexts such as facility location [8], wavelength management in optical networks [3], and the maximum directed cut problem [6]. The use of the LPs in these papers is slightly different than the approach followed here, since the problem objectives allow for LPs that reveal the approximation ratio of the algorithms.

We present the analysis of the greedy algorithm in Section 2 and consider its extension with an additional corrective phase in Section 3. We conclude in Section 4.

2 The greedy algorithm

We first consider the greedy algorithm (henceforth called GREEDY) which is described as follows. While there is a set that covers at least one uncovered element, choose the set that covers the maximum number of uncovered elements. We show that its approximation ratio for frugal coverage is exactly $18/23 \approx 0.782$. This improves the previous lower bound of 0.732 from [7].

2.1 The parameterized LP lemma

Let $f \in (0, 1)$ and consider an instance $\mathcal{I} = (\mathcal{U}, \mathcal{C})$ of the frugal coverage problem, on which GREEDY computes an at most f-approximate solution. Without loss of generality, we assume that every element of $\mathcal{U}$ belongs to at least one set of $\mathcal{C}$. We denote by $\mathcal{O}$ the optimal solution of $(\mathcal{U}, \mathcal{C})$, i.e., the subcollection of $\mathcal{C}$ of minimum size that covers all elements in $\mathcal{U}$. We assume that any set of $\mathcal{C}$ belongs either to the optimal solution or is chosen by GREEDY; if this is not the case for the original instance, we can easily transform it to one that satisfies this assumption so that the solution computed by GREEDY is at most f-approximate.

The algorithm can be thought of as running in phases, starting from phase k, where k is the size of the largest set in the instance. In each phase i, for $i = k, ..., 1$, the algorithm chooses a maximal collection of sets, each containing i yet uncovered elements. Let $(\mathcal{U}_i, \mathcal{C}_i)$, for $i = k, ..., 1$, denote the corresponding instance which remains to be solved just before entering phase i of the algorithm. Naturally, $\mathcal{U}_i$ consists of the elements in $\mathcal{U}$ that have not been covered in previous phases and $\mathcal{C}_i$ consists of the sets of $\mathcal{C}$ which contain at least one such element. We denote by $\mathcal{O}_i$ the sets in $\mathcal{O}$ that also belong to $\mathcal{C}_i$. We consider an assignment of all elements to the sets of $\mathcal{O}$ so that each element is assigned to exactly one among the sets in $\mathcal{O}$ that contains it (if more than one). A set in $\mathcal{O}_i$ is called an (i, j)-set if exactly j elements that have not been covered until the beginning of phase i are assigned to it. For the phase i of the algorithm with $i = 4, ..., 1$, we denote by $\alpha_{i,j}$ the ratio of the number of (i, j)-sets in $\mathcal{O}_i$ over the number $|\mathcal{O}|$ of optimal sets. Furthermore, let x_5 be the ratio of the number of sets in $\mathcal{O}$ that are chosen by GREEDY at the phases $k, k-1, ..., 5$ over $|\mathcal{O}|$. Also, let x_i, for $i = 4, ..., 1$, denote the ratio of the number of sets in $\mathcal{O}$ that are also selected by GREEDY in phase i, over $|\mathcal{O}|$.

By definition, it holds that $|\mathcal{O}_i| = |\mathcal{O}|\sum_{j=1}^{i}\alpha_{i,j}$ and $|\mathcal{O}_i| \leq |\mathcal{O}| - |\mathcal{O}|\sum_{j=i+1}^{5} x_j$, for $i = 1, ..., 4$. Combining these expressions, we get that

$$\sum_{j=1}^{i}\alpha_{i,j} + \sum_{j=i+1}^{5} x_j \leq 1. \tag{1}$$

We denote by T the ratio $|\mathcal{U}|/|\mathcal{O}|$. Our definitions imply that $|\mathcal{U}_i| = |\mathcal{O}|\sum_{j=1}^{i} j\alpha_{i,j}$ for $i = 1, ..., 4$. By the definitions of T and x_5, we have

$$T \geq \sum_{j=1}^{4} j\alpha_{4,j} + 5x_5. \tag{2}$$

Clearly, for $i = 1, ..., 4$, it holds that $|\mathcal{U}_i \setminus \mathcal{U}_{i-1}| = |\mathcal{U}_i| - |\mathcal{U}_{i-1}| \geq ix_i \cdot |\mathcal{O}|$. We have

$$\sum_{j=1}^{i} j\alpha_{i,j} \geq \sum_{j=1}^{i-1} j\alpha_{i-1,j} + ix_i. \tag{3}$$

Now, consider phase i of the algorithm, for $i = 4, 3, 2$, and a set chosen by GREEDY during this phase. The i newly covered elements of this set are assigned to at most i sets in $\mathcal{O}_i$. Furthermore, since GREEDY selects a maximal collection of sets in phase i, we know that any set of $\mathcal{O}_i$ with i assigned elements intersects some of the sets selected by GREEDY during this phase. This means that the number $|\mathcal{U}_i \setminus \mathcal{U}_{i-1}|/i$ of sets selected by GREEDY during phase i is at least $\alpha_{i,i}|\mathcal{O}|/i$. We obtain that

$$\sum_{j=1}^{i} j\alpha_{i,j} - \sum_{j=1}^{i-1} j\alpha_{i-1,j} \geq \alpha_{i,i}. \tag{4}$$

Now, let $\mathcal{S}_\mathcal{G}$ denote the ratio of the number of sets chosen by GREEDY over $|\mathcal{O}|$. Clearly, GREEDY selects at most $|\mathcal{U} \setminus \mathcal{U}_4|/5$ sets in phases $k, k-1, ..., 5$, and exactly $|\mathcal{U}_i \setminus \mathcal{U}_{i-1}|/i$ sets in phase i, for $i = 4, ..., 1$. We have that

$$\begin{aligned}
\mathcal{S}_\mathcal{G} &\leq \left(\frac{1}{5}|\mathcal{U} \setminus \mathcal{U}_4| + \sum_{i=1}^{4}\frac{1}{i}|\mathcal{U}_i \setminus \mathcal{U}_{i-1}|\right)/|\mathcal{O}| \\
&= \frac{1}{5}\left(T - \sum_{j=1}^{4} j\alpha_{4,j}\right) + \sum_{i=1}^{4}\frac{1}{i}\left(\sum_{j=1}^{i} j\alpha_{i,j} - \sum_{j=1}^{i-1} j\alpha_{i-1,j}\right) \\
&= \frac{1}{5}T + \sum_{i=1}^{4}\frac{1}{i(i+1)}\sum_{j=1}^{i} j\alpha_{i,j}.
\end{aligned}$$

Let $\mathrm{OPT}(\mathcal{I})$ denote the benefit of the optimal frugal coverage of $\mathcal{I}$. This naturally corresponds to solution $\mathcal{O}$ for $(\mathcal{U}, \mathcal{C})$. It holds that

$$\mathrm{OPT}(\mathcal{I}) = \left(T + \mathcal{S}_\mathcal{G} - \sum_{i=1}^{5} x_i\right)|\mathcal{O}|.$$

Furthermore, let GREEDY($\mathcal{I}$) denote the profit GREEDY obtains on $\mathcal{I}$. We have that

$$\mathrm{GREEDY}(\mathcal{I}) = \left(T + 1 - \sum_{i=1}^{5} x_i\right) |\mathcal{O}|.$$

The assumption that the solution obtained by GREEDY is at most f-approximate, i.e., GREEDY($\mathcal{I}$) $\leq f \cdot$ OPT($\mathcal{I}$), implies that

$$T + 1 - \sum_{i=1}^{5} x_i \leq f \left(\frac{6}{5}T + \sum_{i=1}^{4} \frac{1}{i(i+1)} \sum_{j=1}^{i} j\alpha_{i,j} - \sum_{i=1}^{5} x_i\right). \tag{5}$$

By expressing inequalities (1)-(5) in standard form, we obtain our parameterized LP lemma.

Lemma 1. *If there exists an instance $\mathcal{I}$ of the frugal coverage problem for which the greedy algorithm computes a solution of benefit* GREEDY($\mathcal{I}$) $\leq f\cdot$OPT($\mathcal{I}$) *for some $f \in [0, 1]$, then the following linear program LP(f) has a feasible solution.*

$$\begin{aligned}
&\sum_{j=1}^{i} \alpha_{i,j} + \sum_{j=i+1}^{5} x_j \leq 1, \quad \textit{for } i = 1, ..., 4 \\
&-(i-1)\alpha_{i,i} - \sum_{j=1}^{i-1} j\alpha_{i,j} + \sum_{j=1}^{i-1} j\alpha_{i-1,j} \leq 0, \quad \textit{for } i = 2, 3, 4 \\
&-\sum_{j=1}^{i} j\alpha_{i,j} + \sum_{j=1}^{i-1} j\alpha_{i-1,j} + ix_i \leq 0, \quad \textit{for } i = 1, ..., 4 \\
&-T + \sum_{j=1}^{4} j\alpha_{4,j} + 5x_5 \leq 0 \\
&(1 - \frac{6f}{5})T - \sum_{i=1}^{4}\sum_{j=1}^{i} \frac{fj}{i(i+1)}\alpha_{i,j} - \sum_{i=1}^{5}(1-f)x_i \leq -1 \\
&\alpha_{i,j} \geq 0, \quad \textit{for } i = 1, ..., 4 \textit{ and } j = 1, ..., i \\
&x_j \geq 0, \quad \textit{for } j = 1, ..., 5 \\
&T \geq 0
\end{aligned}$$

The proof of the approximation bound is based on the following lemma.

Lemma 2. *For every $f < 18/23$, LP(f) has no feasible solution.*

Proof. We can assume that LP(f) is a maximization linear program with objective 0. By duality, if it were feasible, then the optimal objective value of the dual minimization linear program should be 0 as well. We show that this is not the case and that the dual has a solution with strictly negative objective value. This implies the lemma.

In the dual LP, we use the thirteen variables β_1, β_2, β_3, β_4, γ_2, γ_3, γ_4, δ_1, δ_2, δ_3, δ_4, η and ζ corresponding to the constraints of LP(f).Variables β_i correspond to the first set of constraints (inequality (1)) of LP(f), variables γ_i correspond to the second set of constraints (inequality (4)), δ_i correspond to the third set of constraints (inequality (3)), η corresponds to the next constraint (inequality (2)), and ζ corresponds to the last constraint (inequality (5)). So, the dual of LP(f) is depicted in Table 1.

$$
\begin{array}{lll}
min & \beta_1+\beta_2+\beta_3+\beta_4-\zeta & \\
 & & \\
s.t. & \beta_1+\gamma_2-\delta_1+\delta_2-\zeta f/2\geq 0 & \beta_4-3\gamma_4-3\delta_4+3\eta-\zeta 3f/20\geq 0\\
 & \beta_2-\gamma_2+\gamma_3-\delta_2+\delta_3-\zeta f/6\geq 0 & \beta_4-3\gamma_4-4\delta_4+4\eta-\zeta f/5\geq 0\\
 & \beta_2-\gamma_2+2\gamma_3-2\delta_2+2\delta_3-\zeta f/3\geq 0 & \delta_1-(1-f)\zeta\geq 0\\
 & \beta_3-\gamma_3+\gamma_4-\delta_3+\delta_4-\zeta f/12\geq 0 & \beta_1+2\delta_2-(1-f)\zeta\geq 0\\
 & \beta_3-2\gamma_3+2\gamma_4-2\delta_3+2\delta_4-\zeta f/6\geq 0 & \beta_1+\beta_2+3\delta_3-(1-f)\zeta\geq 0\\
 & \beta_3-2\gamma_3+3\gamma_4-3\delta_3+3\delta_4-\zeta f/4\geq 0 & \beta_1+\beta_2+\beta_3+4\delta_4-(1-f)\zeta\geq 0\\
 & \beta_4-\gamma_4-\delta_4+\eta-\zeta f/20\geq 0 & \beta_1+\beta_2+\beta_3+\beta_4+5\eta-(1-f)\zeta\geq 0\\
 & \beta_4-2\gamma_4-2\delta_4+2\eta-\zeta f/10\geq 0 & -\eta+(1-6f/5)\zeta\geq 0\\
 & \beta_i,\delta_i\geq 0,\quad \text{for } i=1,...,4 & \\
 & \gamma_i\geq 0,\quad \text{for } i=2,3,4 & \\
 & \zeta,\eta\geq 0 &
\end{array}
$$

Table 1. The dual of LP(f) in the proof of Lemma 2.

The solution $\beta_1 = f/2+1/46, \beta_2 = f/4, \beta_3 = f/3, \beta_4 = 15/46 - f/4, \gamma_2 = f/4, \gamma_3 = f/6, \gamma_4 = 5/46 - f/12, \delta_1 = 1-f, \delta_2 = \delta_3 = 45/46 - 5f/4, \delta_4 = 1-5f/4, \eta = 1-6f/5$ and $\zeta = 1$ satisfies all the constraints. Observe that $\beta_1+\beta_2+\beta_3+\beta_4-\zeta = \frac{5f}{6} - \frac{15}{23}$, which implies that the objective value of the dual program is strictly negative since $f < \frac{18}{23}$. The lemma follows. □

Theorem 1. *The approximation ratio of the greedy algorithm is at least* 18/23.

Proof. By Lemmas 1 and 2, we have that for any $f < 18/23$ and for any instance $\mathcal{I}$ of the frugal coverage problem, the greedy algorithm computes a solution of benefit GREEDY($\mathcal{I}$) $> f \cdot$ OPT($\mathcal{I}$). Hence, its approximation ratio is at least 18/23. □

2.2 The upper bound

The instance depicted in Figure 1 proves that our analysis is tight. It consists of 18 elements and 11 sets. GREEDY starts by choosing the 2 horizontal sets of size 3, then it chooses the 3 horizontal sets of size 2, and, finally, it chooses all the vertical sets in order to cover the last 6 elements. Thus, GREEDY uses all 11 sets in order to cover the 18 elements for a profit of 18. The optimal solution consists of the 6 vertical sets of size 3 and has a profit of 23 since it covers the 18 elements without using 5 out of the 11 sets. The upper bound follows.

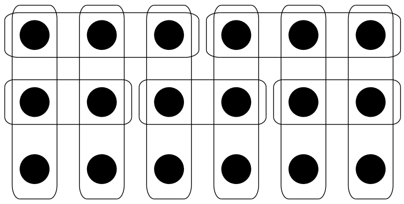

Fig. 1. The tight upper bound for GREEDY.

3 Adding a corrective phase

In this section we show that an improvement in the performance of GREEDY is obtained by adding a simple corrective phase at the end of its execution. We call the new algorithm GREEDY$_c$ and prove that its approximation ratio is exactly $54/67 \approx 0.8059$ for frugal coverage. The definition of the corrective phase is very simple. After the execution of GREEDY, we examine every set s that was chosen during phase 2; if by removing s we still have a cover of the elements, we simply remove s from the solution.

3.1 The parameterized LP Lemma

We slightly modify the parameterized LP for GREEDY in order to capture the behavior of the corrective phase. We will need some additional definitions. We denote by $\mathcal{G}_2$ and $\mathcal{G}_1$ the sets included by the algorithm in phases 2 and 1, respectively. A $(2,1)$-set of $\mathcal{O}_2$ is of type Γ if its uncovered element is included in a set of $\mathcal{G}_2$. A $(2,2)$-set of $\mathcal{O}_2$ is of type A if both of its uncovered elements are included in sets of $\mathcal{G}_2$ and is of type B otherwise. A set of $\mathcal{G}_2$ is of type AA if it does not belong to $\mathcal{O}_2$ and its newly covered elements (during phase 2) belong to $(2,2)$-sets of $\mathcal{O}_2$ of type A and is of type $kl \in \{AB, A\Gamma, BB, B\Gamma\}$ if its newly covered elements belong to two sets of $\mathcal{O}_2$ of types k and l, respectively. A set of $\mathcal{G}_1$ is of type Δ if its newly covered element belongs to a $(2,1)$-set of $\mathcal{O}_2$. We introduce the variables $t_{AA}, t_{AB}, t_{A\Gamma}, t_{BB}, t_{B\Gamma}, t_{\Gamma\Gamma}$, and t_Δ to denote the ratio of the number of sets of types AA, AB, $A\Gamma$, BB, $B\Gamma$, $\Gamma\Gamma$, and Δ over $|\mathcal{O}|$, respectively. We also use variable d to denote the ratio of the number of sets removed during the corrective phase over $|\mathcal{O}|$. The next lemma provides a lower bound on d, which we will use as a constraint in the parameterized LP for GREEDY$_c$.

Lemma 3. $x_1 - t_{AB} - t_{BB} - t_{B\Gamma} - t_\Delta - d \leq 0$.

Proof. We partition $(2,2)$-sets of $\mathcal{O}_2$ of type B into two subtypes. Such a $(2,2)$-set s is of type $\hat{B}$ if one of its elements is included in a set of $\mathcal{G}_1$ that also belongs to $\mathcal{O}_2$ and is of type $\bar{B}$ otherwise. Now, we extend the notation t_k to denote the ratio of the number of sets of $\mathcal{G}_2$ of type $k \in \{A\hat{B}, A\bar{B}, \hat{B}\hat{B}, \hat{B}\bar{B}, \bar{B}\bar{B}, \hat{B}\Gamma, \bar{B}\Gamma\}$ over $|\mathcal{O}|$.

Observe that for each set s of $\mathcal{O}_2$ of type $\hat{B}$, there exists a set in $\mathcal{O}_1 \cap \mathcal{G}_1$ (the one containing the uncovered element of s at the beginning of phase 1) and any other set of $\mathcal{O}_1 \cap \mathcal{G}_1$ is of type Δ. Hence, $x_1 \leq t_{A\hat{B}} + 2t_{\hat{B}\hat{B}} + t_{\hat{B}\bar{B}} + t_{\hat{B}\Gamma} + t_{\Delta}$. Now, observe that the number of sets that will be removed during the corrective phase include those of type $\hat{B}\hat{B}$. Hence,

$$d \geq t_{\hat{B}\hat{B}} \geq x_1 - t_{A\hat{B}} - t_{\hat{B}\hat{B}} - t_{\hat{B}\bar{B}} - t_{\hat{B}\Gamma} - t_{\Delta} \geq x_1 - t_{AB} - t_{BB} - t_{B\Gamma} - t_{\Delta},$$

as desired. □

Variables $a_{1,1}$, $a_{2,1}$, and $a_{2,2}$ will not be explicitly used; observe that we can replace them as follows:

$$\begin{aligned}
\alpha_{1,1} &= t_{AB} + 2t_{BB} + t_{B\Gamma} + t_{\Delta} \\
\alpha_{2,1} &= t_{A\Gamma} + t_{B\Gamma} + 2t_{\Gamma\Gamma} + t_{\Delta}, \text{ and} \\
\alpha_{2,2} &= t_{AA} + \frac{3}{2}t_{AB} + \frac{1}{2}t_{A\Gamma} + 2t_{BB} + t_{B\Gamma} + x_2.
\end{aligned}$$

The profit GREEDY$_c(\mathcal{I})$ is now $d|\mathcal{O}|$ more than GREEDY$(\mathcal{I})$, i.e.,

$$\text{GREEDY}_c(\mathcal{I}) = \left(T + 1 - \sum_{i=1}^{5} x_i + d\right)|\mathcal{O}|,$$

while OPT$(\mathcal{I})$ is the same as in the previous section. In this way, we have obtained the following parameterized LP lemma for GREEDY$_c$.

Lemma 4. *If there exists an instance $\mathcal{I}$ of the frugal coverage problem for which GREEDY$_c$ computes a solution of benefit GREEDY$_c(\mathcal{I}) \leq f \cdot OPT(\mathcal{I})$ for some $f \in [0,1]$, then the linear program LP$_c(f)$ has a feasible solution.*

$$t_{AB} + 2t_{BB} + t_{B\Gamma} + t_{\Delta} + \sum_{j=2}^{5} x_j \leq 1$$

$$t_{AA} + \frac{3}{2}t_{AB} + \frac{3}{2}t_{A\Gamma} + 2t_{BB} + 2t_{B\Gamma} + 2t_{\Gamma\Gamma} + t_{\Delta} + x_2 + \sum_{j=3}^{5} x_j \leq 1$$

$$\sum_{j=1}^{i} \alpha_{i,j} + \sum_{j=i+1}^{5} x_j \leq 1, \qquad \textit{for } i = 3, 4$$

$$-2\alpha_{3,3} - \sum_{j=1}^{2} j\alpha_{3,j} + 2t_{AA} + 3t_{AB} + 2t_{A\Gamma} + 4t_{BB} + 3t_{B\Gamma} + 2t_{\Gamma\Gamma} + t_{\Delta} + 2x_2 \leq 0$$

$$-3\alpha_{4,4} - \sum_{j=1}^{3} j\alpha_{4,j} + \sum_{j=1}^{3} j\alpha_{3,j} \leq 0$$

$$-\sum_{j=1}^{3} j\alpha_{3,j} + 2t_{AA} + 3t_{AB} + 2t_{A\Gamma} + 4t_{BB} + 3t_{B\Gamma} + 2t_{\Gamma\Gamma} + t_{\Delta} + 2x_2 + 3x_3 \leq 0$$

$$
\begin{aligned}
&-\sum_{j=1}^{4} j\alpha_{4,j} + \sum_{j=1}^{3} j\alpha_{3,j} + 4x_4 \le 0\\
&-T + \sum_{j=1}^{4} j\alpha_{4,j} + 5x_5 \le 0\\
&x_1 - t_{AB} - t_{BB} - t_{B\Gamma} - t_{\Delta} - d \le 0\\
&(1-\frac{6f}{5})T - \frac{f}{3}t_{AA} - ft_{AB} - \frac{f}{3}t_{A\Gamma} - \frac{5f}{3}t_{BB} - ft_{B\Gamma} - \frac{f}{3}t_{\Gamma\Gamma} - \frac{2f}{3}t_{\Delta}\\
&\quad -\sum_{i=3}^{4}\sum_{j=1}^{i}\frac{fj}{i(i+1)}\alpha_{i,j} - \sum_{i=3}^{5}(1-f)x_i - (1-f)x_1 - (1-\frac{2f}{3})x_2 + d \le -1\\
&\alpha_{i,j} \ge 0, \quad \textit{for } i = 3,4 \textit{ and } j = 1,...,i\\
&x_j \ge 0, \quad \textit{for } j = 1,...,5\\
&t_{AA}, t_{AB}, t_{A\Gamma}, t_{BB}, t_{B\Gamma}, t_{\Gamma\Gamma}, t_{\Delta} \ge 0\\
&T \ge 0
\end{aligned}
$$

The proof of the approximation bound is based on the following lemma.

Lemma 5. *For every $f < 54/67$, $LP_c(f)$ has no feasible solution.*

Proof. Similarly to the proof of Lemma 2, we assume that $\mathrm{LP}_c(f)$ is a maximization linear program with objective 0. We show that its dual has a solution with strictly negative objective value when $f < 54/67$. This implies the lemma.

To construct the dual of $\mathrm{LP}_c(f)$ we use the eleven variables β_1, β_2, β_3, β_4, γ_3, γ_4, δ_3, δ_4, η, μ and ζ corresponding to the constraints of $\mathrm{LP}_c(f)$. Variables β_i correspond to the constraints in the first three lines of $\mathrm{LP}_c(f)$, variables γ_i correspond to the constraints in the fourth and fifth line, variables δ_i correspond to the constraints in the sixth and seventh line, and η, μ, and ζ correspond to the three last constraints of $\mathrm{LP}_c(f)$, respectively. So, the dual of $\mathrm{LP}_c(f)$ is depicted in Table 2.

The solution $\beta_1 = 41/134, \beta_2 = f/3 + 13/67, \beta_3 = 0, \beta_4 = 31f/108, \gamma_3 = 0, \gamma_4 = f/12, \delta_3 = \delta_4 = 1 - 67f/54, \eta = 1 - 6f/5, \zeta = 1$ and $\mu = 1 - f$ satisfies all the constraints. Observe that $\beta_1 + \beta_2 + \beta_3 + \beta_4 - \zeta = \frac{67f}{108} - \frac{1}{2}$, which implies that the objective value of the dual program is strictly negative for $f < \frac{54}{67}$. The lemma follows. □

The next statement follows by Lemmas 4 and 5.

Theorem 2. *The approximation ratio of GREEDY$_c$ is at least* 54/67.

3.2 The upper bound

The instance which yields the tight upper bound for GREEDY_c consists of 48 elements and 31 sets; see Figure 2. There, the sets selected by GREEDY_c are shown, while the optimal solution consists of the 12 vertical disjoint sets of size 4

$$
\begin{array}{lll}
\mathit{min} & \beta_1+\beta_2+\beta_3+\beta_4-\zeta & \\
& & \\
s.t. & \beta_2+2\gamma_3+2\delta_3-\frac{f}{3}\zeta\geq 0 & \beta_4-\gamma_4-\delta_4+\eta-\frac{f}{20}\zeta\geq 0 \\
& \beta_1+\frac{3}{2}\beta_2+3\gamma_3+3\delta_3-\mu-f\zeta\geq 0 & \beta_4-2\gamma_4-2\delta_4+2\eta-\frac{f}{10}\zeta\geq 0 \\
& \frac{3}{2}\beta_2+2\gamma_3+2\delta_3-\frac{f}{3}\zeta\geq 0 & \beta_4-3\gamma_4-3\delta_4+3\eta-\frac{3f}{20}\zeta\geq 0 \\
& 2\beta_1+2\beta_2+4\gamma_3+4\delta_3-\mu-\frac{5f}{3}\zeta\geq 0 & \beta_4-3\gamma_4-4\delta_4+4\eta-\frac{f}{5}\zeta\geq 0 \\
& \beta_1+2\beta_2+3\gamma_3+3\delta_3-\mu-f\zeta\geq 0 & \mu-(1-f)\zeta\geq 0 \\
& 2\beta_2+2\gamma_3+2\delta_3-\frac{f}{3}\zeta\geq 0 & \beta_1+\beta_2+2\gamma_3+2\delta_3-(1-\frac{2f}{3})\zeta\geq 0 \\
& \beta_1+\beta_2+\gamma_3+\delta_3-\mu-\frac{2f}{3}\zeta\geq 0 & \beta_1+\beta_2+3\delta_3-(1-f)\zeta\geq 0 \\
& \beta_3-\gamma_3+\gamma_4-\delta_3+\delta_4-\frac{f}{12}\zeta\geq 0 & \beta_1+\beta_2+\beta_3+4\delta_4-(1-f)\zeta\geq 0 \\
& \beta_3-2\gamma_3+2\gamma_4-2\delta_3+2\delta_4-\frac{f}{6}\zeta\geq 0 & \beta_1+\beta_2+\beta_3+\beta_4+5\eta-(1-f)\zeta\geq 0 \\
& \beta_3-2\gamma_3+3\gamma_4-3\delta_3+3\delta_4-\frac{f}{4}\zeta\geq 0 & -\eta+(1-6f/5)\zeta\geq 0 \\
& \beta_i\geq 0,\quad \text{for } i=1,\dots,4 & -\mu+\zeta\geq 0 \\
& \gamma_i,\delta_i\geq 0,\quad \text{for } i=3,4 & \\
& \zeta,\mu,\eta\geq 0 &
\end{array}
$$

Table 2. The dual of $\mathrm{LP}_c(f)$ in the proof of Lemma 5.

(note that only half of them are shown in the figure). Clearly, the profit obtained by GREEDY$_c$ is $48+6=54$, while the profit of the optimal solution is at least $48+19=67$, and the upper bound follows. We remark that the corrective phase does not remove any set.

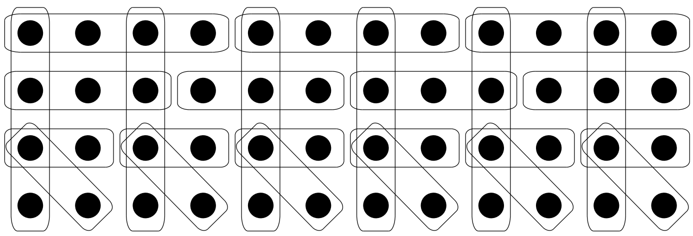

Fig. 2. The tight upper bound for GREEDY$_c$.

4 Extensions

Our focus in the current paper has been on simple algorithms for frugal coverage. A possible improvement could be obtained by extending the corrective phase so that it considers removing sets included in phases of the greedy algorithm before phase 2. Another sophisticated technique that can probably lead to further improvements has been proposed by Duh and Fürer [4] for the set cover problem with sets of bounded size and is known as *semi-local* optimization. This technique has been proved to be useful in other problems such as color saving and spanning star forest [2, 4]. In future work, we plan to extend our analysis to greedy-like algorithms that combine semi-local optimization and corrective phases.

References

1. S. Athanassopoulos, I. Caragiannis, and C. Kaklamanis. Analysis of approximation algorithms for k-set cover using factor-revealing linear programs. *Theory of Computing Systems*, 45(3), pp. 555-576, 2009.
2. S. Athanassopoulos, I. Caragiannis, C. Kaklamanis, and M. Kyropoulou. An improved approximation bound for spanning star forest and color saving. In *Proceedings of the 34th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, LNCS 5734, Springer, pp. 90-101, 2009.
3. I. Caragiannis. Wavelength management in WDM rings to maximize the number of connections. *SIAM Journal on Discrete Mathematics*, 23(2), pp. 959-978, 2009.
4. R. Duh and M. Fürer. Approximation of k-set cover by semi local optimization. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 256-264, 1997.
5. U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45(4), pp. 634-652, 1998.
6. U. Feige and S. Jozeph. Oblivious algorithms for the maximum directed cut problem. *arXiv*: 1010.0406, 2010.
7. C.-C. Huang and Z. Svitkina. Donation center location problem. In *Proceedings of the 29th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pp. 227-238, 2009.
8. 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), pp. 795-824, 2003.
9. D. S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9, pp. 256-278, 1974.
10. A. Levin. Approximating the unweighted k-set cover problem: greedy meets local search. *SIAM Journal on Discrete Mathematics*, 23(1), pp. 251-264, 2008.
11. A. Levin and U. Yovel. Uniform unweighted set cover: the power of non-oblivious local search. *Theoretical Computer Science*, 412(12-14), pp. 1033-1053, 2011.
12. R. Raz and S. Safra. A sub-constant error-probability low-degree test, and sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 475-484, 1997.