Greedy differential approximations for MIN SET COVER

Cristina Bazgan Jérôme Monnot Vangelis Th. Paschos Fabrice Serrière

LAMSADE, Université Paris-Dauphine
Place du Maréchal De Lattre de Tassigny, 75775 Paris Cedex 16, France
{bazgan,monnot,paschos,serriere}@lamsade.dauphine.fr

January 26, 2004

Abstract

We present in this paper di erential approximation results for and . We first show that the di erential approximation ratio of the natural greedy algorithm for is bounded below by $1.365/\Delta$ and above by $4/(\Delta+1)$, where Δ is the maximum set-cardinality in the -instance. Next, we study an approximation algorithm for and provide a tight lower bound of $1/\Delta$.

1 Introduction

Given a family $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ of subsets of a ground set $C = \{c_1, c_2, \ldots, c_n\}$ (we assume that $\cup_{S_i \in \mathcal{S}} S_i = C$), a set-cover of C is a sub-family $\mathcal{S}' \subseteq \mathcal{S}$ such that $\cup_{S_i \in \mathcal{S}'} S_i = C$; is the problem of determining a minimum-size set-cover of C. consists of considering that sets of $\mathcal{S}$ are weighted by positive weights; the objective becomes then to determine a minimum total-weight cover of C.

Definition 1. Given an instance $(\mathcal{S}, C)$ of , its characteristic graph $B(S, C)$ is a bipartite graph B with color-classes $S = \{s_1, \ldots, s_m\}$, corresponding to the members of the family $\mathcal{S}$ and $C = \{c_1, \ldots, c_n\}$, corresponding to the elements of the ground set C; the edge-set E of B is defined as $E = \{(s_i, c_j) : c_j \in S_i, S_i \in \mathcal{S}, c_j \in C\}$. ∎

A cover $\mathcal{S}'$ of C is said minimal (or minimal for the inclusion) if removal of any set $S \in \mathcal{S}'$ results in a family that is not anymore a cover for C.

Remark 1. Consider an instance $(\mathcal{S}, C)$ of and a minimal set-cover $\mathcal{S}'$ for it. Then, for any $S_i \in \mathcal{S}'$, there exists $c_j \in C$ such that S_i is the only set in $\mathcal{S}'$ covering c_j. Such a c_j will be called *non-redundant with respect to* $S_i \in \mathcal{S}'$; furthermore, S_i itself will be called *non-redundant for* $\mathcal{S}'$. With respect to the induced subgraph $B(S', C)$ of $B(S, C)$, for any $s \in S'$, there exists a $c \in C$ such that $\delta(c) = 1$, where, for a vertex v of a graph G, $\delta(v)$ denotes the degree of v. ∎

In this paper we study di erential approximability for in both unweighted and weighted versions. Di erential approximability is analyzed using the so-called di erential approximation ratio defined, for an instance I of an **NPO** problem Π and an approximation algorithm $\mathtt{A}$ computing a solution S for Π in I, as $\delta_{\mathtt{A}}(I) = |\omega(I) - m_{\mathtt{A}}(I, S)|/|\omega(I) - \mathrm{opt}(I)|$ where $\omega(I)$ is the value of the worst Π-solution for I, $m_{\mathtt{A}}(I, S)$ is the value of S and $\mathrm{opt}(I)$ is the value of an optimal Π-solution for I. For an instance $I = (\mathcal{S}, C)$ of , $\omega(I) = m$, the size of the family $\mathcal{S}$. Obviously, this is the maximum-size cover of I. Finally, standard approximability is analyzed using the standard approximation ratio defined as $m_{\mathtt{A}}(I, S)/\mathrm{opt}(I)$.

We study in this paper the performance of two approximation algorithms. The first one is the classical greedy algorithm studied, for the unweighted case and for the standard approximation ratio, in [6, 7] and, more recently, in [8]. For this algorithm, we provide a di erential approximation ratio bounded below by $1.365/\Delta$ when $\Delta = \max_{S_i \in \mathcal{S}}\{|S_i|\}$ grows to ∞. We next deal with and analyze the di erential approximation performance of a simple greedy algorithm that starts from the whole $\mathcal{S}$ considering it as solution for and then it reduces it by removing the heaviest of the remaining sets of $\mathcal{S}$ per time until the first moment $B(S', C)$ does not fulfill Remark 1. We show that this algorithm achieves di erential approximation ratio $1/\Delta$.

Di erential approximability for both and have already been studied in [1]. The di erential approximation ratios provided there are $1/\Delta$, for the former, and $1/(\Delta+1)$, for the latter. Our current work improves (quite importantly for the unweighted case), these old results. Note also that an approximation algorithm for has been analyzed also in [3] under the assumption $m \geqslant n$, the size of the ground set C. It has been shown that, under this assumption, is approximable within di erential approximation ratio 1/2. More recently, in [5], under the same assumption, has been proved approximable within di erential approximation ratio 289/360.

The following inapproximability result, proved in [3], holds for and shows that approximation ratios of the same type as in standard approximation (for example, $O(1/\ln\Delta)$, or $O(1/\log n)$) are extremely unlikely for in di erential approximation. Consequently, di erential approximation results for cannot be trivially achieved by simply transposing the existing standard approximation results to the di erential framework. This is a further motivation of our work.

Proposition 1. *If **P** $\neq$ **NP**, then inapproximability bounds for standard (and differential) approximation for hold as differential inapproximability bounds for . Consequently, unless **P** $=$ **NP**, is not differentially approximable within $O(n^{\epsilon-(1/2)})$, for any $\epsilon > 0$.*

In what follows, we deal with non-trivial instances of . An instance I is non-trivial for if the two following conditions hold simultaneously:

- no set $S_i \in \mathcal{S}$ is subset of a set $S_j \in \mathcal{S}$;
- no element in C is contained in I by only one subset of $\mathcal{S}$ (i.e., there is no non-redundant set for $\mathcal{S}$).

2 A global lower bound

Consider an instance $I = (\mathcal{S}, C)$ for , with $|\mathcal{S}| = m$ and an approximation algorithm A, computing a minimal cover $\mathcal{S}'$ for C. Fix an optimal cover $\mathcal{S}^*$ and set: $\mathcal{S}_c = \mathcal{S}^* \cap \mathcal{S}'$, $\mathcal{S}_r = \mathcal{S} \setminus (\mathcal{S}^* \cup \mathcal{S}')$, $\mathcal{S}^*_p = \mathcal{S}^* \setminus \mathcal{S}'$ and $\mathcal{S}'_p = \mathcal{S}' \setminus \mathcal{S}^*_p$. As we have already mentioned, $\omega(I) = m$. Based upon these notations (that will be used throughout the paper), we have:

$$m = \omega(I) = \left|\mathcal{S}^* \cup \mathcal{S}'\right| + |\mathcal{S}_r| = \left|\mathcal{S}^*_p\right| + |\mathcal{S}_c| + \left|\mathcal{S}'_p\right| + |\mathcal{S}_r| \tag{1}$$

Then, the di erential approximation ratio of A for can be expressed as follows:

$$\delta\left(I, \mathcal{S}'\right) = \frac{\omega(I) - |\mathcal{S}'|}{\omega(I) - |\mathcal{S}^*|} \stackrel{(1)}{=} \frac{\left|\mathcal{S}^*_p\right| + |\mathcal{S}_r|}{\left|\mathcal{S}'_p\right| + |\mathcal{S}_r|} \tag{2}$$

Consider the sub-instance of I induced by $(\hat{\mathcal{S}}^*_p \cup \hat{\mathcal{S}}'_p, \hat{C})$, where $\hat{C} = C \setminus (\cup_{S_i \in S_c} S_i)$ and, for all $S \in \mathcal{S}^*_p \cup \mathcal{S}'_p$, $\hat{\mathcal{S}}^*_p = \mathcal{S}^*_p \cap \hat{C}$ and $\hat{\mathcal{S}}'_p = \mathcal{S}'_p \cap \hat{C}$; remark also that $\forall S \in \mathcal{S}^*_p \cup \mathcal{S}'_p$, $S \neq \emptyset$. Then, it is easy to see that both $\hat{\mathcal{S}}^*_p$ and $\hat{\mathcal{S}}'_p$ are covers for $\hat{C}$. Moreover,

$$\left|\hat{\mathcal{S}}^*_p\right| \geqslant \frac{\left|\hat{C}\right|}{\Delta} \tag{3}$$

$$\left|\hat{\mathcal{S}}'_p\right| \leqslant \left|\hat{C}\right| \tag{4}$$

Revisit (2) and remark that $(|\mathcal{S}_p^*| + |\mathcal{S}_r|)/(|\mathcal{S}'_p| + |\mathcal{S}_r|) \geqslant |\mathcal{S}_p^*|/|\mathcal{S}'_p| \geqslant |\hat{\mathcal{S}}_p^*|/|\hat{\mathcal{S}}'_p|$. Then, using (3) and (4), we get: $\delta(I, \mathcal{S}') \geqslant 1/\Delta$.

Proposition 2. *The differential ratio of any minimal approximation algorithm for is bounded below by* $1/\Delta$.

3 The natural greedy algorithm for MIN SET COVER

We analyze in this section the di erential approximation performance of the following very classical greedy algorithm for , called SCGREEDY in the sequel:

1. set $\mathcal{S}'' = \emptyset$;
2. compute $S \in \operatorname{argmax}_{S_i \in \mathcal{S}}\{|S_i|\}$;
3. set: $\mathcal{S}'' = \mathcal{S}'' \cup \{S\}$;
4. update I setting: $\mathcal{S} = \mathcal{S} \setminus \{S\}$, $C = C \setminus S$ and, for any $S_j \in \mathcal{S}$, $S_j = S_j \setminus S$;
5. repeat Steps 2 to 4 until $C = \emptyset$;
6. range $\mathcal{S}''$ in the order sets have been chosen; set $\mathcal{S}'' = \{S''_1, S''_2 \ldots, S''_{|\mathcal{S}''|}\}$;
7. if $\mathcal{S}''$ is not minimal for the inclusion (cf., Remark 1), then for $j = |\mathcal{S}''|$ downto 1: if $S'_j \in \mathcal{S}''$ is redundant for $\mathcal{S}''$ (i.e., for any $c \in S'_j$ there exists another set in $\mathcal{S}''$ covering c), then remove S'_j from $\mathcal{S}''$;
8. output $\mathcal{S}' = \{S_1, S_2 \ldots, S_{|\mathcal{S}'|}\}$ the minimal cover computed in Step 8.

Theorem 1. *For* Δ *arbitrarily large, algorithm* SCGREEDY *achieves differential approximation ratio* $1.365/\Delta$.

Proof. Consider sets in $\mathcal{S}''$ (computed in Step 6) with their residual cardinalities, i.e., as they have beeen chosen during Steps 2 and 4, and remark that, so-considered, $\mathcal{S}'' = \{S''_1, S''_2 \ldots, S''_{|\mathcal{S}''|}\}$ forms a partition on C. On the other hand, consider $\mathcal{S}' = \{S_1, S_2 \ldots, S_{|\mathcal{S}'|}\}$ (as it has been computed in Step 7) and consider, in a first time, not the whole sets $S_i \in \mathcal{S}'$ but their residual parts, i.e., for $i = 1, \ldots, |\mathcal{S}'|$, $S'_i = S_i \setminus (\cup_{S_j \in \mathcal{S}'', j<i} S_j)$ (remark that family $\{S'_1, S'_2 \ldots, S'_{|\mathcal{S}'|}\}$ does not cover C).

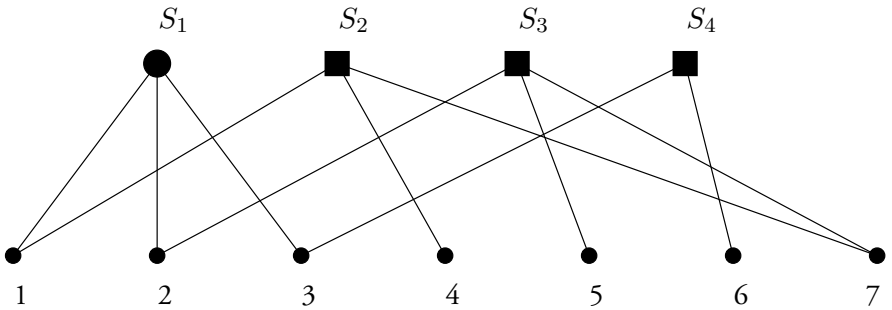

Figure 1: An example of application of Step 7 of SCGREEDY.

As an example, assume some -instance $(\mathcal{S}, C)$ with $C = \{1, \ldots, 6\}$ and suppose that execution of Steps 2 to 6 has produced a cover $\mathcal{S}'' = \{S_1, S_2, S_3, S_4\}$ (using notations introduced above, $\{S''_1, S''_2, S''_3, S''_4\} = \{S_1, S_2, S_3, S_4\}$). Figure 1 illustrates graph $B(S'', C)$, i.e., the subgraph of $B(S, C)$ (Definition 1) induced by the vertices of S corresponding to $\mathcal{S}''$ and by the vertices associated with the

ground set C. It is easy to see that $\mathcal{S}''$ is not minimal and application of Step 7 of SCGREEDY drops S_1 out of $\mathcal{S}''$; hence, $\mathcal{S}'_1 = \{S_2, S_3, S_4\}$. The residual parts of S_2, S_3 and S_4 are $S'_2 = \{4, 7\}$, $S'_3 = \{5\}$ and $S'_4 = \{6\}$, respectively.

Let $X = \cup_{S'_i \in \mathcal{S}'} S'_i$; obviously, $\mathcal{S}'$ form a partition on X. With respect to the example of Figure 1, $X = \{4, 5, 6, 7\}$. Since $\mathcal{S}'$ is minimal for the inclusion, any set $S_i \in \mathcal{S}'$ contains at least an element of C non-redundant with respect to S_i; a contrario, S_i should be removed from $\mathcal{S}''$ during Step 7 of SCGREEDY. Consider set $X' \subseteq X$ consisting of one non-redundant element per set of $\mathcal{S}'$; in the case where a set of $\mathcal{S}'$ has more than one such elements, randomly choose one of them. Assume that elements in X' are ranged in the order they have been covered. Recall that $\mathcal{S}'_p = \mathcal{S}' \setminus \mathcal{S}^*$ and denote by C'_1 the subset of X' the elements of which are covered by the sets in $\mathcal{S}'_p$. Finally, denote by $\mathcal{S}^{*\prime\prime}_p$ the family of sets in $\mathcal{S}^*$ covering the elements of C'_1.

Obviously, $\mathcal{S}^{*\prime\prime}_p \cap \mathcal{S}' = \emptyset$; a contrario, there would exist at least a $c \in C'_1$ covered in sub-instance $(\mathcal{S}^{*\prime\prime}_p \cup \mathcal{S}'_p, C'_1)$ twice: one time by a set in $\mathcal{S}'_p$ and one time by a set in $\mathcal{S}^{*\prime\prime}_p$ which also belongs to $\mathcal{S}'_p$, absurd by the construction of C'_1. Moreover, assume that $\mathcal{S}^{*\prime\prime}_p$ forms a partition on C'_1, considering that if an element $c \in C'_1$ is covered by more than one sets in $\mathcal{S}^{*\prime\prime}_p$, then c is randomly assigned to one of them.

Observe that, using the notations just introduced, (2) can be rewritten as:

$$\delta\left(I, \mathcal{S}'\right) = \frac{\left|\mathcal{S}^{*\prime\prime}_p\right| + \left(\left|\mathcal{S}^*_p \setminus \mathcal{S}^{*\prime\prime}_p\right| + |\mathcal{S}_r|\right)}{\left|\mathcal{S}'_p\right| + |\mathcal{S}_r|} \tag{5}$$

Consider graph $B' = B(S^{*\prime\prime}_p \cup S'_p, C'_1)$ (Definition 1). Following the remarks just above, B' is not connected and, furthermore, any of its connected components is of the form of Figure 2, where we assume that $S^* \in \mathcal{S}^{*\prime\prime}_p$ is of cardinality 3, $S'_i, S'_j, S'_k \in \mathcal{S}'_p$ and $c_l, c_m, c_n \in C'_1$.

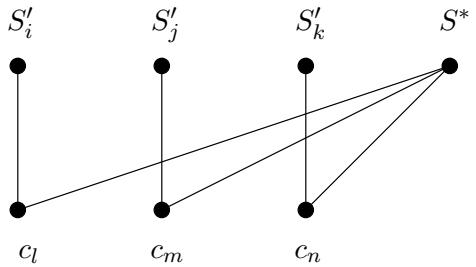

Figure 2: A connected component of $B' = B(S^{*\prime\prime}_p \cup S'_p, C'_1)$.

For $i = 1, \ldots, \Delta$, denote by x_i the number of connected components of B'corresponding to sets $S^* \in \mathcal{S}^{*\prime\prime}_p$ of cardinality i (with respect to sub-instance $(\mathcal{S}'_p \cup \mathcal{S}^{*\prime\prime}_p, C'_1)$). Then, by construction of this sub-instance, we have:

$$\left|\mathcal{S}^{*\prime\prime}_p\right| = \sum_{i=1}^{\Delta} x_i \tag{6}$$

$$\left|\mathcal{S}'_p\right| = \left|C'_1\right| = \sum_{i=1}^{\Delta} i \times x_i \tag{7}$$

Consider a $k > 1$ (it will be fixed later) and set $i_0 = \Delta/k$. We study the two following cases: $|C'_1| \leqslant i_0 |\mathcal{S}^{*\prime\prime}_p|$ and $|C'_1| \geqslant i_0 |\mathcal{S}^{*\prime\prime}_p|$.

3.0.1 $|C_1'| \leqslant i_0|\mathcal{S}_p^{*\prime\prime}|$

In this case, we have immediately from (5), (6) and (7):

$$\delta\left(I, \mathcal{S}'\right) \geqslant \frac{\left|\mathcal{S}_p^{*\prime\prime}\right|}{\left|\mathcal{S}_p'\right|} = \frac{\left|\mathcal{S}_p^{*\prime\prime}\right|}{|C_1'|} \geqslant \frac{1}{i_0} \geqslant \frac{k}{\Delta} \tag{8}$$

3.0.2 $|C_1'| \geqslant i_0|\mathcal{S}_p^{*\prime\prime}|$

Here, one can easily see that the average cardinality of sets in $\mathcal{S}_p^{*\prime\prime}$ is greater than, or equal to, i_0. Indeed, denote by $\overline{S}$ the average cardinality in $\mathcal{S}_p^{*\prime\prime}$; then,

$$\overline{S} = \frac{1}{\left|\mathcal{S}_p^{*\prime\prime}\right|} \sum_{S_i \in \mathcal{S}_p^{*\prime\prime}} |S_i| = \frac{|C_1'|}{\left|\mathcal{S}_p^{*\prime\prime}\right|} \geqslant i_0 \tag{9}$$

where the second equality holds because we have assumed that $\mathcal{S}_p^{*\prime\prime}$ forms a partition on C_1'.

Consider once more the component of Figure 2, suppose that set S^* has cardinality i with respect to $(\mathcal{S}_p^{*\prime\prime} \cup \mathcal{S}_p', C_1')$ and denote it by $S^* = \{c_{\ell_1}, \ldots, c_{\ell_i}\}$. Denote by S_{ℓ_j}, $j = 1, \ldots, i$, the sets of $\mathcal{S}_p'$ (singletons with respect to $(\mathcal{S}_p^{*\prime\prime} \cup \mathcal{S}_p', C_1')$) covering $c_{\ell_1}, \ldots, c_{\ell_i}$, respectively. Recall that these last elements have been ranged in the order they have been covered by algorithm `SCGREEDY`; furthermore, $S_{\ell_1}, \ldots, S_{\ell_i}$ have been chosen in this order (cf., Steps 6 and 7 of `SCGREEDY`). Then, $|S_{\ell_1}| \geqslant i - 1$ and, more generally, $|S_{\ell_p}| \geqslant i - p$, $p = 1, \ldots, i-1$. Consequently, associated with S^* there exist $(i-1) + (i-2) + \ldots + 1 = i(i-1)/2$ elements of C not included in C_1'. Iterating this observation for any connected component of B' and assuming that there exist x_i sets of cardinality i (with respect to $(\mathcal{S}_p^{*\prime\prime} \cup \mathcal{S}_p', C_1')$), $i = 1, \ldots, \Delta$, in $\mathcal{S}_p^{*\prime\prime}$, we can conclude that there exists a set $C_2 \subseteq C$, outside set C_1, of size

$$|C_2| \geqslant \sum_{i=1}^{\Delta} \frac{i(i-1)}{2} x_i \tag{10}$$

Elements of C_2 are obviously covered, with respect to $\mathcal{S}^*$, by sets either from $\mathcal{S}_p^*$, or from $\mathcal{S}_C$ (recall that $\mathcal{S}_p'$ is not part of $\mathcal{S}^*$). Suppose that sets of $\mathcal{S}_p^{*\prime\prime}$ of cardinality i (there exist x_i such sets), $i = 1, \ldots, \Delta$, cover a total of $k_i x_i$ elements of C_2. Therefore, there exists a subset $C_2' \subseteq C_2$ of size at least

$$|C_2'| \geqslant \sum_{i=1}^{\Delta} \left(\frac{i(i-1)}{2} - k_i \right) x_i$$

elements of which are covered in $\mathcal{S}^*$ by sets in $\mathcal{S}^* \setminus \mathcal{S}_p^{*\prime\prime}$. In order that C_2' is covered, a family $\hat{\mathcal{S}}^* \subseteq \mathcal{S}^* \setminus \mathcal{S}_p^{*\prime\prime}$ of size

$$\left|\hat{\mathcal{S}}^*\right| \geqslant \frac{\sum_{i=1}^{\Delta} \left(\frac{i(i-1)}{2} - k_i \right) x_i}{\Delta} \tag{11}$$

is needed. Dealing with $\hat{\mathcal{S}}^*$, suppose that for a $y \in [0, 1]$:

1. $(1-y)|\hat{\mathcal{S}}^*|$ sets of $\hat{\mathcal{S}}^*$ belong to $S_p^* \setminus \mathcal{S}_p^{*\prime\prime}$ and
2. $y|\hat{\mathcal{S}}^*|$ sets of $\hat{\mathcal{S}}^*$ belong to $\mathcal{S}_c$; denote by $\hat{\mathcal{S}}_y^*$ the family of these sets.

Sub-family $\hat{\mathcal{S}}_y^* \subseteq \mathcal{S}_c$ is, by hypothesis, common to both $\mathcal{S}'$ (the cover computed by `SCGREEDY`) and $\mathcal{S}^*$. Minimality of $\mathcal{S}'$ implies that, for any set $S_i \in \hat{\mathcal{S}}^*$, there exists at least one element of C non-redundant with respect to S_i. So, there exist at least $|C_3| = |\hat{\mathcal{S}}_y^*|$ elements of C outside C_1' and C_2.

Some elements of C_3 can belong to (and be covered by) sets in $\mathcal{S}_p^{*\prime\prime}$. In any case, for the sets of $\mathcal{S}_p^{*\prime\prime}$ of cardinality i (recall that there exist x_i such sets), there exist at most $(\Delta - (i + k_i))x_i$ elements of C_3 that can belong to them. So, there exist at least

$$\begin{aligned} \left|C_3'\right| &= |C_3| - \sum_{i=1}^{\Delta} (\Delta - (i + k_i))\, x_i = \left|\hat{\mathcal{S}}_y^*\right| - \sum_{i=1}^{\Delta} (\Delta - (i + k_i))\, x_i \\ &= y\left|\hat{\mathcal{S}}^*\right| - \sum_{i=1}^{\Delta} (\Delta - (i + k_i))\, x_i \end{aligned} \tag{12}$$

elements of C_3 (recall that these elements are non-redundant with respect to $\mathcal{S}'$) covered by sets in $\mathcal{S}_c$. Since initial instance $(\mathcal{S}, C)$ is non-trivial, elements of C_3' are also contained in sets either from $\mathcal{S}_p^* \setminus \mathcal{S}_p^{*\prime\prime}$, or from $\mathcal{S}_r$. So, a family $\mathcal{S}_2$ of

$$|\mathcal{S}_2| \geqslant \frac{|C_3'|}{\Delta} \overset{(12)}{=} \frac{y\left|\hat{\mathcal{S}}^*\right| - \sum\limits_{i=1}^{\Delta} (\Delta - (i + k_i))\, x_i}{\Delta} \tag{13}$$

contains elements of C_3'; sets of $\mathcal{S}_2$ belong either to $\mathcal{S}_p^* \setminus \mathcal{S}_p^{*\prime\prime}$, or to $\mathcal{S}_r$. Denote by $\mathcal{S}_2'$ the sub-family of $\mathcal{S}_2$ belonging to $\mathcal{S}_r$.

By the discussion just above, and using Items 1 and 2 about $\hat{\mathcal{S}}^*$, we have starting from (5):

$$\begin{aligned} \delta\left(I, \mathcal{S}'\right) &\geqslant \frac{\left|\mathcal{S}_p^{*\prime\prime}\right| + |\mathcal{S}_2 \setminus \mathcal{S}_2'| + (1-y)\left|\hat{\mathcal{S}}^*\right| + |\mathcal{S}_r|}{\left|\mathcal{S}_p'\right| + |\mathcal{S}_r|} \overset{|\mathcal{S}_r| \geqslant |\mathcal{S}_2|}{\geqslant} \frac{\left|\mathcal{S}_p^{*\prime\prime}\right| + (1-y)\left|\hat{\mathcal{S}}^*\right| + |\mathcal{S}_2|}{\left|\mathcal{S}_p'\right| + |\mathcal{S}_2'|} \\ &\overset{1-y \geqslant 0}{\geqslant} \frac{\left|\mathcal{S}_p^{*\prime\prime}\right| + (1-y)\left|\hat{\mathcal{S}}^*\right| + |\mathcal{S}_2|}{\left|\mathcal{S}_p'\right| + (1-y)\left|\hat{\mathcal{S}}^*\right| + |\mathcal{S}_2|} \end{aligned} \tag{14}$$

Using (13) and performing some easy algebra, one can prove that the sum $(1-y)|\hat{\mathcal{S}}^*| + |\mathcal{S}_2|$, appearing in both the numerator and the denominator of (14), verifies:

$$(1-y)\left|\hat{\mathcal{S}}^*\right| + |\mathcal{S}_2| \geqslant \frac{(1-y)\left|\hat{\mathcal{S}}^*\right|}{\Delta} + |\mathcal{S}_2| \geqslant \sum_{i=1}^{\Delta} \left(\frac{i(i-1)}{2\Delta^2} + \frac{i}{\Delta} - 1\right) x_i \tag{15}$$

Furthermore, note that function $(a + x)/(b + x)$ is increasing with x, for $a \leqslant b$. Therefore, using (6), (7) and (15), Expression (14) becomes:

$$\delta\left(I, \mathcal{S}'\right) \geqslant \frac{\sum\limits_{i=1}^{\Delta} \left(\frac{i(i-1)}{2\Delta^2} + \frac{i}{\Delta}\right) x_i}{\sum\limits_{i=1}^{\Delta} \left(i + \frac{i(i-1)}{2\Delta^2} + \frac{i}{\Delta} - 1\right) x_i} \tag{16}$$

Set now $f(x) = (x(x-1)/2\Delta^2) + (x/\Delta)$. Then, (16) can be expressed as:

$$\delta\left(I, \mathcal{S}'\right) \geqslant \frac{\sum\limits_{i=1}^{\Delta} f(i) x_i}{\sum\limits_{i=1}^{\Delta} \left(f(i) + i - 1\right) x_i} \tag{17}$$

Setting $\bar{i} = (\sum_{i=1}^{\Delta} i x_i)/(\sum_{i=1}^{\Delta} x_i)$, we will now show the following inequality:

$$\frac{\sum\limits_{i=1}^{\Delta} f(i)x_i}{\sum\limits_{i=1}^{\Delta} \left(f(i)+i-1\right)x_i} \geqslant \frac{f\left(\bar{i}\right)}{f\left(\bar{i}\right)+\bar{i}-1} \tag{18}$$

In order to prove (18), we can equivalently show that

$$\begin{aligned}
& f\left(\bar{i}\right)\sum_{i=1}^{\Delta}(i-1)x_i \leqslant (i-1)\sum_{i=1}^{\Delta} f(i)x_i \\
\iff \quad & f\left(\bar{i}\right)\frac{\sum\limits_{i=1}^{\Delta}(i-1)x_i}{\sum\limits_{i=1}^{\Delta} x_i} \leqslant (\bar{i}-1)\frac{\sum\limits_{i=1}^{\Delta} f(i)x_i}{\sum\limits_{i=1}^{\Delta} x_i} = \left(\frac{\sum\limits_{i=1}^{\Delta}(i-1)x_i}{\sum\limits_{i=1}^{\Delta} x_i}\right)\left(\frac{\sum\limits_{i=1}^{\Delta} f(i)x_i}{\sum\limits_{i=1}^{\Delta} x_i}\right) \\
\iff \quad & f\left(\bar{i}\right) \leqslant \frac{\sum\limits_{i=1}^{\Delta} f(i)x_i}{\sum\limits_{i=1}^{\Delta} x_i}
\end{aligned} \tag{19}$$

Since f is convex, $\sum_{i=1}^{\Delta} z_i f(i) \geqslant f(\sum_{i=1}^{\Delta} i z_i)$, where $z_i \in [0,1]$, $\sum_{i=1}^{\Delta} z_i = 1$ (Jensen's theorem). Setting $z_i = x_i / \sum_{i=1}^{\Delta} x_i$, (19) and, equivalently, (18) follow.

Remark now that $\bar{i}$ is nothing else than $\overline{S}$ defined in (9). Consequently, (16) and (17) can be rewritten as:

$$\delta\left(I, \mathcal{S}'\right) \geqslant \frac{\frac{|\overline{S}|(|\overline{S}|-1)}{2\Delta^2} + \frac{|\overline{S}|}{\Delta}}{|\overline{S}| + \frac{|\overline{S}|(|\overline{S}|-1)}{\Delta} - 1} \tag{20}$$

Expression (20) is increasing with $|\overline{S}|$, which, for the case we deal with (revisit (9)), is greater than i_0. Hence, (20) becomes:

$$\delta\left(I, \mathcal{S}'\right) \geqslant \frac{\frac{i_0(i_0-1)}{2\Delta^2} + \frac{i_0}{\Delta}}{i_0 + \frac{i_0(i_0-1)}{\Delta} - 1} \tag{21}$$

Recall that we have fixed $i_0 = \Delta/k$; assume, furthermore, that Δ is arbitrarily large. Then, $i_0 \approx i_0 - 1$ and (21) becomes:

$$\delta\left(I, \mathcal{S}'\right) \geqslant \frac{\frac{1}{k} + \frac{1}{2k^2}}{\frac{\Delta}{k} + \frac{1}{2k^2} + \frac{1}{k} - 1} \approx \frac{\frac{1}{k} + \frac{1}{2k^2}}{\frac{\Delta}{k}} = \frac{2k+1}{2k} \times \frac{1}{\Delta} \tag{22}$$

Ratio given by (8) is increasing with k, while the one of (22) is decreasing with k. Equality of both ratios is reached when $2k^2 - 2k - 1 = 0$, i.e., for

$$k = \frac{2+\sqrt{12}}{4} \approx 1.365 \tag{23}$$

Using (23) with (8) or (22), the di erential approximation ratio achieved by `SCGREEDY` becomes greater than, or equal to, $1.365/\Delta$ and the proof of the theorem is now complete. ∎

Proposition 3. *There exist -instances such that the approximation ratio of `SCGREEDY` is bounded above by* $4/(\Delta+2)$.

Proof. Assume a fixed $t > 1$, a ground set $C = \{c_{ij} : i = 1, \ldots, t-1, j = 2, \ldots, t, j > i\}$ and a system $\mathcal{S} = \{S_1, \ldots, S_t\}$, where $S_i = \{c_{ji} : j < i\} \cup \{c_{ij} : j > i\}$, for $i = 1, \ldots, t$. Denote by $I_t = (\mathcal{S}, C)$ the instance of defined on C and $\mathcal{S}$.

Remark that the smallest cover for C includes at least $t-1$ sets of $\mathcal{S}$. Indeed, consider a family $\mathcal{S}' \subseteq \mathcal{S}$ of size less than $t-1$. Then, there exists $i_0 < j_0$ such that neither S_{i_0}, nor S_{j_0} belong to $\mathcal{S}'$. In this case element $c_{i_0 j_0} \in C$ is not covered by $\mathcal{S}'$. Note finally that, for I_t, the maximum size of the subsets of $\mathcal{S}$ is $\Delta = t-1$. Indeed, for any $i = 1, \ldots, t$, $|\{c_{ji} : j < i\}| = i-1$ and $|\{c_{ij} : j < i\}| = t-i$; so, $|S_i| = t-1$, since

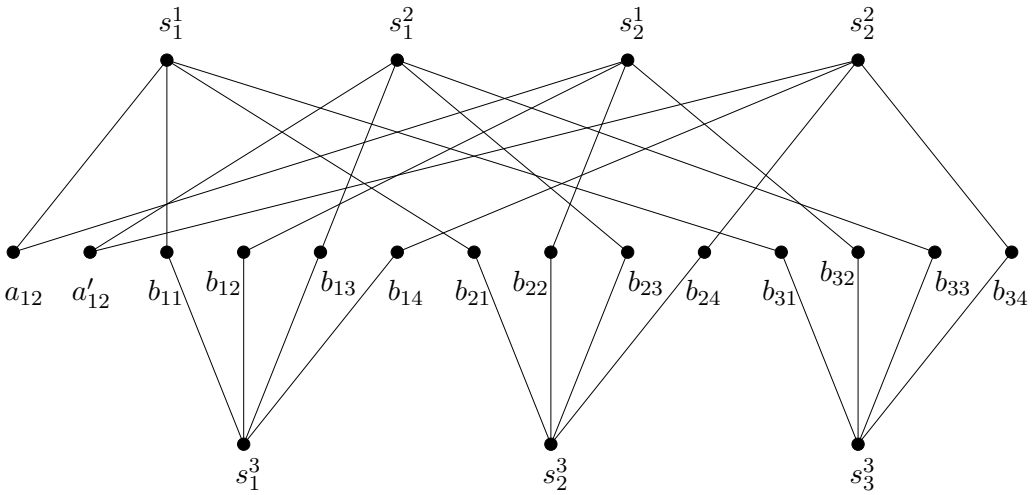

Figure 3: The characteristic graph of instance $\hat{I}$.

Fix now an even Δ and construct the following instance $(\mathcal{S}, C)$ for :

$$
\begin{aligned}
C &= \left\{a_{ij}, a'_{ij} : i = 1, \ldots, \frac{\Delta}{2}, j = 1, \ldots, \frac{\Delta}{2}, j > i\right\} \cup \left\{b_{ij} : i = 1, \ldots, \frac{\Delta+2}{2}, j = 1, \ldots, \Delta\right\} \\
S_i^1 &= \{a_{ji} : j < i\} \cup \{a_{ij} : j > i\} \cup \left\{b_{ji} : j = 1, \ldots, \frac{\Delta+2}{2}\right\} \quad i = 1, \ldots, \frac{\Delta}{2} \\
S_i^2 &= \left\{a'_{ij} : j < i\right\} \cup \left\{a'_{ij} : j > i\right\} \cup \left\{b_{jk} : j = 1, \ldots, \frac{\Delta+2}{2}, k = i + \frac{\Delta}{2}\right\} \quad i = 1, \ldots, \frac{\Delta}{2} \\
S_i^3 &= \{b_{ij} : j = 1, \ldots, \Delta\} \quad i = 1, \ldots, \frac{\Delta+2}{2} \\
\mathcal{S}^1 &= \left\{S_i^1 : i = 1, \ldots, \frac{\Delta}{2}\right\} \\
\mathcal{S}^2 &= \left\{S_i^2 : i = 1, \ldots, \frac{\Delta}{2}\right\} \\
\mathcal{S}^3 &= \left\{S_i^3 : i = 1, \ldots, \frac{\Delta+2}{2}\right\} \\
\mathcal{S} &= \mathcal{S}^1 \cup \mathcal{S}^2 \cup \mathcal{S}^3
\end{aligned}
$$

As an example, the following instance $\hat{I} = (\mathcal{S}, C)$ can be constructed, according to above schema, with $\Delta = 4$ (its characteristic graph is illustrated in Figure 3):

$$
\begin{aligned}
C &= \left\{a_{12}, a'_{12}, b_{11}, b_{12}, b_{13}, b_{14}, b_{21}, b_{22}, b_{23}, b_{24}, b_{31}, b_{32}, b_{33}, b_{34}\right\} \\
S_1^1 &= \{a_{12}, b_{11}, b_{21}, b_{31}\}
\end{aligned}
$$

$$
\begin{aligned}
S_1^2 &= \left\{a'_{12}, b_{13}, b_{23}, b_{33}\right\} \\
S_1^3 &= \{b_{11}, b_{12}, b_{13}, b_{14}\} \\
S_2^1 &= \{a_{12}, b_{12}, b_{22}, b_{32}\} \\
S_2^2 &= \left\{a'_{12}, b_{14}, b_{24}, b_{34}\right\} \\
S_2^3 &= \{b_{21}, b_{22}, b_{23}, b_{24}\} \\
S_3^3 &= \{b_{31}, b_{32}, b_{33}, b_{34}\} \\
\mathcal{S}^1 &= \left\{S_1^1, S_2^1\right\} \\
\mathcal{S}^2 &= \left\{S_1^2, S_2^2\right\} \\
\mathcal{S}^3 &= \left\{S_1^3, S_2^3, S_3^3\right\} \\
\mathcal{S} &= \left\{S_1^1, S_2^1, S_1^2, S_2^2, S_1^3, S_2^3, S_3^3\right\}
\end{aligned}
$$

It is easy to see that, $\forall S_i \in \mathcal{S}$, $|S_i| = \Delta$. Hence, during its first iteration, `SCGREEDY` can choose a set in $\mathcal{S}^3$. Such a choice does not reduce cardinalities of the remaining sets in this sub-family; so, during its first $(\Delta+2)/2$ iterations, `SCGREEDY` can exclusively choose all sets in $\mathcal{S}^3$. Remark that such choices entail that the surviving instance is the union of two disjoint instances $I_{\Delta/2}$ (i.e., instances of type I_t, as the ones defined in the beginning of this section, with $t = \Delta/2$), induced by the sub-systems $(\mathcal{S}^1, \{a_{ij}\})$ and $(\mathcal{S}^2, \{a'_{ij}\})$. According to what has been discussed at the beginning of the section, any cover for such instances uses at least $(\Delta/2) - 1$ sets. We so have, for a set-cover $\mathcal{S}'$ computed by `SCGREEDY` (remark that $\mathcal{S}'$ is minimal):

$$
m\left(\hat{I}, \mathcal{S}'\right) \geqslant \frac{\Delta+2}{2} + 2\left(\frac{\Delta}{2} - 1\right) = \frac{3\Delta}{2} - 1 \tag{24}
$$

Furthermore, given that sub-family $\mathcal{S}^1 \cup \mathcal{S}^2$ is a cover for C, we have:

$$
\begin{aligned}
\text{opt}\left(\hat{I}\right) &\leqslant \Delta & (25)\\
\omega\left(\hat{I}\right) &= \frac{3\Delta}{2} + 1 & (26)
\end{aligned}
$$

Combination of (24), (25) and (26) concludes $\delta(\hat{I}, \mathcal{S}') \leqslant 4/(\Delta+2)$.

Assume now that Δ is odd and consider the following instance $(\mathcal{S}, C)$ for :

$$
\begin{aligned}
C &= \left\{a_{ij} : i = 1, \ldots, \frac{\Delta-1}{2}, j = 2, \ldots, \frac{\Delta-1}{2}, j > i\right\} \\
&\quad \cup \left\{a'_{ij} : i = 1, \ldots, \frac{\Delta+1}{2}, j = 2, \ldots, \frac{\Delta+1}{2}, j > i\right\} \\
&\quad \cup \left\{b_{ij} : i = 1, \ldots, \frac{\Delta+1}{2}, j \in 1, \ldots, \Delta\right\} \\
S_i^1 &= \{a_{ji} : j < i\} \cup \{a_{ij} : j > i\} \cup \left\{b_{ji} : j = 1, \ldots, \frac{\Delta+1}{2}\right\} \quad i = 1, \ldots, \frac{\Delta-1}{2} \\
S_i^2 &= \left\{a'_{ji} : j < i\right\} \cup \{a_{ij} : j > i\} \\
&\quad \cup \left\{b_{jk} : j = 1, \ldots, \frac{\Delta+1}{2}, k = i + \frac{\Delta-1}{2}\right\} \quad i = 1, \ldots, \frac{\Delta+1}{2} \\
S_i^3 &= \{b_{ij} : j = 1, \ldots, \Delta\} \quad i = 1, \ldots, \frac{\Delta+1}{2}
\end{aligned}
$$

The same arguments conclude an upper bound of $4/(\Delta+1)$, for odd values of Δ. ∎.

4 Differential approximation for MIN WEIGHTED SET COVER

Consider an instance $I = (\mathcal{S}, C, \vec{w})$ of , where $\vec{w}$ denotes the vector of the weights on the subsets of $\mathcal{S}$ and the following algorithm, denoted by WSC in what follows:

- set $\mathcal{S}' = \mathcal{S}$;
- order sets in $\mathcal{S}$ in decreasing weight-order;
- let $\mathcal{S} = \{S_1, \ldots, S_m\}$ the ordered family obtained;
- for $i = 1$ to m: if $\mathcal{S}' \setminus \{S_i\}$ covers C, then set $\mathcal{S}' = \mathcal{S}' \setminus \{S_i\}$;
- output $\mathcal{S}'$.

Proposition 4. *Algorithm WSC achieves differential approximation ratio bounded below by* $1/\Delta$. *This bound is asymptotically tight.*

Proof. We use in what follows notations introduced in Section 3; furthermore we set $\bar{\mathcal{S}}' = \mathcal{S} \setminus \mathcal{S}' = \mathcal{S}_r \cup (\mathcal{S}^* \setminus \mathcal{S}')$ and $\bar{\mathcal{S}}^* = \mathcal{S} \setminus \mathcal{S}^* = \mathcal{S}_r \cup (\mathcal{S}' \setminus \mathcal{S}^*)$. Denoting, for any $S_i \in \mathcal{S}$, by w_i its weight, and, for any sub-family $\mathcal{X} \subseteq \mathcal{S}$, by $w_{\mathcal{X}}$, its total weight, i.e., the quantity $\sum_{S_i \in \mathcal{X}} w_i$, the di erential approximation ratio of WSC becomes (with the same arguments as for (2) in Section 2):

$$\delta\left(I, \mathcal{S}'\right) = \frac{w_{\bar{\mathcal{S}}'}}{w_{\bar{\mathcal{S}}^*}} \tag{27}$$

Set $C_c = \cup_{S_i \mathcal{S}_c} S_i$ and $\bar{C} = C \setminus C_c$. It is easy to see that both $\mathcal{S}'_p$ and $\mathcal{S}^*_p$ cover $\bar{C}$. Denote by C' the set of non-redundant elements with respect to sets of $\mathcal{S}'_p$; obviously, $C' \subseteq \bar{C}$, and any element of C' is covered, with respect to $\mathcal{S}^*$, by sets in $\mathcal{S}^*_p$. Consider the sub-instance of I induced by $(\mathcal{S}^*_p \cup \mathcal{S}'_p, C')$. Obviously, the restriction to C' of sets in $\mathcal{S}^*_p$ form a partition for C'. Fix a set $S_i \in \mathcal{S}^*_p$ and denote by S'_i its restriction to C', i.e., $S'_i = S_i \cap C'$. Assume that $S'_i \neq \emptyset$; as it will be understood just below if this is not the case, then the approximation ratio of WSC will be even better. Denote by $\mathcal{S}'_i = \{S_{i_1}, \ldots, S_{i_k}\}$, the sub-family of $\mathcal{S}'_p$ covering elements of S'_i in $(\mathcal{S}^*_p \cup \mathcal{S}'_p, C')$. Obviously, since sets $S_{i_1}, \ldots, S_{i_k}$ bave been chosen by WSC, $w_{i_j} \leqslant w_i$ and, since $|S_i| \leqslant \Delta$, we get: $\sum_{S_i \in \mathcal{S}'_p} w_i \leqslant \Delta \sum_{S_j \in \mathcal{S}^*_p} w_j$. In all,

$$\underbrace{\sum_{S_i \in \mathcal{S}'_p} w_i + \sum_{S_i \in \mathcal{S}_r} w_i}_{= \sum_{S_i \in \bar{\mathcal{S}}^*} w_i = w_{\bar{\mathcal{S}}^*}} \leqslant \Delta \left(\underbrace{\sum_{S_i \in \mathcal{S}^*_p} w_i + \sum_{S_i \in \mathcal{S}_r} w_i}_{= \sum_{S_i \in \bar{\mathcal{S}}'} w_i = w_{\bar{\mathcal{S}}'}} \right) \tag{28}$$

Combination of (27) and (28) concludes the ratio claimed.

For tightness, fix $\Delta \in \mathbb{N}$, $w \in \mathbb{R}^+$ and consider the following instance $(\mathcal{S}, C, \vec{w})$ for :

$$\begin{aligned} C &= \{1, \ldots, \Delta\} \\ S_0 &= C \\ S_i &= \{i\} \quad i = 1, \ldots, \Delta \\ \mathcal{S} &= \{S_0, S_1, \ldots, S_\Delta\} \\ w_0 &= w + 1 \\ w_i &= w \quad i = 1, \ldots, \Delta \end{aligned}$$

Here, $w_{\bar{\mathcal{S}}'} = w + 1$, while $w_{\bar{\mathcal{S}}^*} = \Delta w$ and, following (27), the di erential approximation ratio of WSC is $(w+1)/w\Delta \to 1/\Delta$. ∎

5 Conclusions

As we have already mention, Proposition 1, implies that, unless $\mathbf{P} = \mathbf{NP}$, is not polynomially approximable within di erential ratios of $O(\log^{-1} n)$. Consequently, di erential approximation bounds, even trivial, cannot be directly derived from the standard ones. The results of Theorem 1 and of Proposition 3 exhibit an important gap between lower and upper bounds for SCGREEDY. It seems us very interesting to reduce this gap by improving both of them.

Other interesting questions about di erential approximability of are, for instance:

- can be as "well di erentially approximable" as , i.e, is di erentially approximable within $O(\log^2 n/n)$ ([4]) or (asymtotically) within $k/\Delta, \forall k \in \mathbb{N}$ ([2])? (recall that standard and di erential approximation ratios coincide for);
- does there exist for an upper bound tighter than the one for ?

The result of Proposition 4, even if it slightly improves the lower bound of [1], is, to our opinion, far from being optimal for . It would be interesting to improve it. Also, questions mentioned above hold for the unweighted case too. Finally another question that seem us very interesting is the investigation of structural links between and with respect to the di erential approximation. Are these problems approximate equivalent or not?

References

[1] M. Demange, P. Grisoni, and V. Th. Paschos. Di erential approximation algorithms for some combinatorial optimization problems. *Theoret. Comput. Sci.*, 209:107–122, 1998.

[2] M. Demange and V. Th. Paschos. Improved approximations for maximum independent set via approximation chains. *Appl. Math. Lett.*, 10(3):105–110, 1997.

[3] M. Demange and V. Th. Paschos. On an approximation measure founded on the links between optimization and polynomial approximation theory. *Theoret. Comput. Sci.*, 158:117–141, 1996.

[4] M. M. Halldórsson. Approximations of weighted independent set and hereditary subset problems. *J. Graph Algorithms Appli.*, 4(1):1–16, 2000.

[5] R. Hassin and S. Khuller. z-approximations. *J. Algorithms*, 41:429–442, 2001.

[6] D. S. Johnson. Approximation algorithms for combinatorial problems. *J. Comput. System Sci.*, 9:256–278, 1974.

[7] L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Math.*, 13:383–390, 1975.

[8] P. Slavík. A tight analysis of the greedy algorithm for set cover. In *Proc. STOC'96*, pages 435–441, 1996.