

Approximation algorithms for the Label-Cover_{MAX} and Red-Blue Set Cover problems

David Peleg¹

Department of Computer Science and Applied Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel

Available online 11 May 2006

Abstract

This paper presents approximation algorithms for two extensions of the set cover problem: a graph-based extension known as the Max-Rep or Label-Cover_{MAX} problem, and a color-based extension known as the Red-Blue Set Cover problem. First, a randomized algorithm guaranteeing approximation ratio \sqrt{n} with high probability is proposed for the Max-Rep (or Label-Cover_{MAX}) problem, where n is the number of vertices in the graph. This algorithm is then generalized into a $4\sqrt{n}$ -ratio algorithm for the nonuniform version of the problem. Secondly, it is shown that the Red-Blue Set Cover problem can be approximated with ratio $2\sqrt{n \log \beta}$, where n is the number of sets and β is the number of blue elements. Both algorithms can be adapted to the weighted variants of the respective problems, yielding the same approximation ratios.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Label Cover; Red-Blue Set Cover; Max-Rep problem; Approximation algorithms

1. Introduction

1.1. Background

Recent classifications of NP-hard problems by their approximability properties have led to the identification of a group of problems termed *class III* problems in [10]. These problems can be informally characterized as ones known to have no approximation algorithm with ratio $2^{\log^{1-\epsilon} n}$ (for any $0 < \epsilon < 1$) under some plausible complexity-theoretic assumption (such as $NP \neq P$ or $NP \not\subseteq DTIME(n^{O(\text{polylog } n)})$). We henceforth refer to this property as *strong inapproximability*. This class includes problems such as the minimization and maximization versions of *Label-Cover* [2], *AND/OR Scheduling* [8], *Minimum-Monotone-Satisfying-Assignment (MMSA)* [1], *Min-Rep* and *Max-Rep* [11], *Red-Blue Set Cover* [3], *sparsest k -spanner* for $k \geq 3$ [6,7], and more.

While negative (strong inapproximability) results are known for all of those problems (and indeed, in a certain sense they define the class), less is known about positive (approximability) results. The current paper is concerned with providing such results for some of the above problems.

E-mail address: david.peleg@weizmann.ac.il (D. Peleg).

¹ Supported in part by a grant from the Israel Ministry of Science and Art.

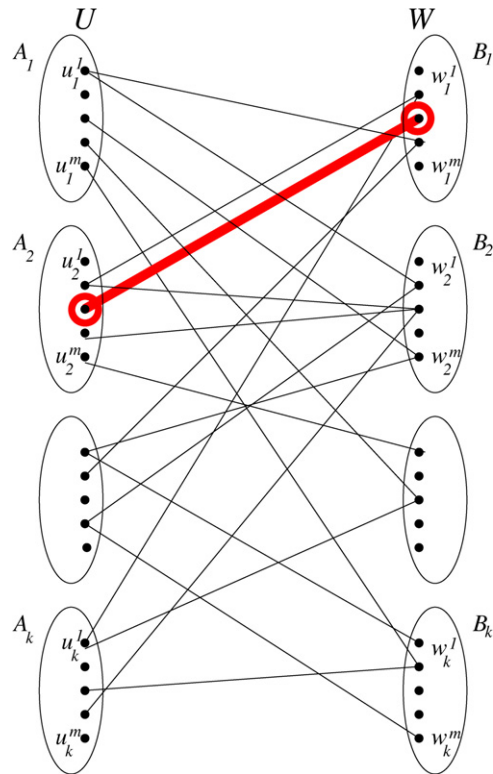


Fig. 1. An instance of the Max-Rep problem. The highlighted edge makes the super-vertices A_2 and B_1 adjacent.

1.2. The problems considered

1.2.1. Max-Rep, Label-Cover_{MAX} and related problems

The *Max-Rep* problem is defined in [11] as follows. We are given a bipartite graph $G(U, W, E)$, where U and W are each split into a disjoint union of k sets, $U = \bigcup_{i=1}^k A_i$ and $W = \bigcup_{i=1}^k B_i$. The sets A_i, B_i all have size m . Let $\mathcal{A} = \{A_1, \dots, A_k\}$ and $\mathcal{B} = \{B_1, \dots, B_k\}$. An instance of the problem consists of the 5-tuple $(U, W, E, \mathcal{A}, \mathcal{B})$. The bipartite graph G and the partitions \mathcal{A} and \mathcal{B} of U and W induce a bipartite super-graph $\mathcal{H} = (\mathcal{A}, \mathcal{B}, E_{\mathcal{H}})$. Two super-vertices A_i and B_j are adjacent in \mathcal{H} iff there exist some $u \in A_i$ and $w \in B_j$ which are adjacent in G . (See Fig. 1.)

A set of vertices $C \subseteq U \cup W$ is said to *cover* the super-edge (A_i, B_j) if it contains a pair of vertices u, w such that $u \in A_i, w \in B_j$ and $(u, w) \in E$. The set C is a *legal cover* for \mathcal{H} if it contains at most one vertex from each super-vertex. The goal is to select a legal cover C for \mathcal{H} covering the maximum number of super-edges possible.

A minimization version of this problem, called *Min-Rep*, is also introduced in [11]. In this version, a cover C must cover *every* super-edge, but it may contain any number of vertices from each super-vertex, and the goal is to select a minimum size cover C for \mathcal{H} .

A closely related problem is the *Label-Cover* problem, introduced in [2] and presented in [10] as one of six canonical problems for proving hardness of approximation. This problem has minimization and maximization versions called Label-Cover_{MIN} and Label-Cover_{MAX}, which can be represented as variants of the Min-Rep and Max-Rep problems respectively, except that the notion of super-edge coverage is slightly different. Namely, a super-edge (A_i, B_j) is said to be *covered* if for *every* vertex $u \in A_i \cap C$ there is a vertex $w \in B_j \cap C$ such that $(u, w) \in E$. Note that the Label-Cover_{MAX} problem is equivalent to the Max-Rep problem.

1.2.2. The Red-Blue Set Cover problem

The *Red-Blue Set Cover* problem was introduced in [3]. It is a natural generalization of the set-cover problem, defined as follows. Consider a finite universe partitioned into two disjoint sets, $U = R \cup B$, where $R = \{r_1, \dots, r_\rho\}$ is

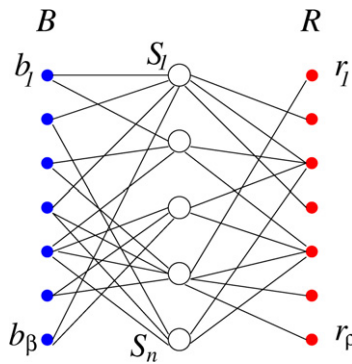


Fig. 2. An instance of the Red-Blue Set Cover problem. In this instance, the optimal cover consists of sets $\{S_2, S_3, S_5\}$. An alternative, more expensive, cover is $\{S_1, S_4\}$. (For colours see the web version of this article.)

a set of *red* elements and $B = \{b_1, \dots, b_\beta\}$ is a set of *blue* elements. We are given a collection of sets over the universe U , $S = \{S_1, S_2, \dots, S_n\}$. For any subcollection $S' \subseteq S$, let

$$U(S') = \bigcup_{S_i \in S'} S_i, \quad B(S') = U(S') \cap B, \quad R(S') = U(S') \cap R.$$

(See Fig. 2.) The goal is to choose a subcollection S' of S that covers all the elements of B (i.e., s.t. $B \subseteq B(S')$) while minimizing $|R(S')|$, the number of red elements in S' .

The Red-Blue Set Cover problem is also shown in [3] to be equivalent to $MMSA_3$, the third level of the *Minimum-Monotone-Satisfying-Assignment* problem introduced in [1].

1.3. Previous work

Both versions of the Label-Cover problem were introduced in [2], where the minimization version was also shown to be strongly inapproximable. More precisely, it was shown that it is quasi-NP-hard to approximate the problem with ratio $2^{\log^{1-\epsilon} n}$ for any constant $0 < \epsilon < 1$ (or in other words, such approximation is impossible unless $NP \subseteq DTIME(n^{O(\text{polylog} n)})$). This result was recently improved in [5], by weakening the complexity-theoretic assumption to $NP \neq P$ and allowing ϵ to be as small as $\log \log^{-c} n$ for any $c < 1/2$.

The Max-Rep and Min-Rep problems were introduced in [11], where it was also shown that both are strongly inapproximable. This implies the same for the $\text{Label-Cover}_{\text{MAX}}$ problem, due to its equivalence to the Max-Rep problem.

The strong inapproximability of the Red-Blue Set Cover problem was shown independently in [3] and [6]. Specifically, it is shown that it is quasi-NP-hard to approximate the Red-Blue Set Cover problem with ratio $2^{\log^\epsilon n}$, for any $0 < \epsilon < 1$. Applications of the Red-Blue Set Cover problem in a variety of domains, such as data mining applications, information retrieval or general machine learning and classification, are discussed at length in [3], as well as a number of special-case variants of the problem and related problems, including Set Cover, Group Steiner and Directed Steiner, and Minimum Color Path.

Few positive results exist for the above problems. The Red-Blue Set Cover problem admits naive approximation algorithms with ratios β , ρ , or $n \log \beta$. A number of better approximation algorithms are given in [3] for this problem. Specifically, letting k_B (respectively, k_R) denote the maximum number of blue (resp., red) elements in any of the sets S_i , the paper presents approximation algorithms with ratio $Z_1 = 2\sqrt{k_B \cdot n}$ or $Z_2 = O(n^{1-1/k_R} \log n)$. Hence the approximation ratios of these algorithms are low when k_B or k_R are small, but may be as high as $Z_1 = \Omega(\sqrt{n\beta})$ or $Z_2 = \Omega(n \log n)$, respectively, in the general case, when k_B or k_R are large.

In [6] it is shown that the Min-Rep and $\text{Label-Cover}_{\text{MIN}}$ problems admit a \sqrt{n} -approximation ratio. It is also shown that the Min-Rep and $\text{Label-Cover}_{\text{MIN}}$ problems restricted to the cases where the girth of the induced super-graph is greater than t , admit an $n^{2/t}$ approximation ratio. In particular, it follows that the Min-Rep and the $\text{Label-Cover}_{\text{MIN}}$ problems with girth greater than $\log^\epsilon n$ (for some constant $\epsilon > 0$) are not strongly inapproximable, i.e., they admit an $O(2^{\log^{\epsilon'} n})$ -approximation ratio, for some $0 < \epsilon' < 1$.

1.4. Contributions

The current paper presents approximation algorithms for two of the above problems.

The first algorithm, presented in Section 2, is a randomized algorithm guaranteeing approximation ratio \sqrt{n} with high probability for the Max-Rep problem (or for Label-Cover_{MAX}). (A simple deterministic variant was pointed out by Y. Hassin [9].) This algorithm is then generalized into a $4\sqrt{n}$ -ratio algorithm for the nonuniform version of the problem, in which there may be a different number of sets in \mathcal{A} and \mathcal{B} , and these sets may have different sizes. The algorithm can be generalized also to the weighted version of the problem, where super-edges have real nonnegative weights, and the goal is to maximize the total weight of covered super-edges.

In Section 3 we present an algorithm with approximation ratio $2\sqrt{n \log \beta}$ for the Red-Blue Set Cover problem. The algorithm can be generalized also to the weighted version of the problem, in which every red element $r_i \in R$ has a positive real weight associated with it, and the goal is to minimize the weight of the selected cover.

2. An approximation algorithm for the Max-Rep problem

2.1. The uniform case

Let us start with some terminology. For every $1 \leq i \leq k$, let $A_i = \{u_i^1, \dots, u_i^m\}$ and $B_i = \{w_i^1, \dots, w_i^m\}$. We think of the graph as drawn with the vertices of U on the left and the vertices of W on the right.

Consider a cover C . Without loss of generality we may assume that $C = U^C \cup W^C$, where U^C (respectively, W^C) contains exactly one vertex u_i^C (resp., w_i^C) in each super-vertex A_i (resp., B_i) on the left (resp., right). In particular, we denote by $C^* = U^* \cup W^*$ the optimal solution to the problem, and let u_i^* (resp., w_i^*) denote the unique vertex of $U^* \cap A_i$ (resp., $W^* \cap B_i$).

For vertex subsets $U' \subseteq U$ and $W' \subseteq W$, let $G(U', W')$ denote the subgraph of G induced by U' and W' . For a vertex $u \in U'$ (resp., $w \in W'$), let $\deg(u, W')$ (resp., $\deg(w, U')$) denote its degree in $G(U', W')$. For $u \in A_i$, let $\Gamma(u, \mathcal{H})$ denote the set of super-vertices B_j neighboring u , namely, such that there is an edge $(u, w_j^j) \in E$ for some $w_j^j \in B_j$. Let $\text{sdeg}(u, \mathcal{H})$ denote the super-degree of u , namely, the cardinality of $\Gamma(u, \mathcal{H})$. The super-degree of a vertex represents the number of super-edges it can potentially cover.

For any cover C , let $\widehat{E}(C)$ denote the set of super-edges of \mathcal{H} covered by C . Note that these are precisely the super-edges corresponding to the edges of the graph $G(U^C, W^C)$. Denote the cardinality of this set by $f(C) = |\widehat{E}(C)|$.

We first present two approximation procedures for the problem. The first of the two has approximation ratio k , so it applies well in case there are few sets.

Procedure FEW_SETS

1. Calculate the super-degree $\text{sdeg}(u, \mathcal{H})$ of every vertex $u \in U$.
2. Find the vertex $\hat{u} \in U$ with maximum super-degree.
3. Construct a set \widehat{W} consisting of one neighbor $w_i^{j_i}$ of \hat{u} in every super-vertex $B_i \in \Gamma(\hat{u}, \mathcal{H})$.
4. Complete the set $\widehat{W} \cup \{\hat{u}\}$ into a cover \widehat{C} arbitrarily.
5. Output the cover \widehat{C} .

Lemma 2.1. *Procedure FEW_SETS yields a k -approximation for the Max-Rep problem.*

Proof. By the choice of \hat{u} ,

$$\text{sdeg}(u_i^*, \mathcal{H}) \leq f(\{\hat{u}\} \cup \widehat{W}) \leq f(\widehat{C})$$

for every $1 \leq i \leq k$. Subsequently,

$$f(C^*) \leq \sum_{1 \leq i \leq k} \text{sdeg}(u_i^*, \mathcal{H}) \leq k \cdot f(\widehat{C}). \quad \square$$

Our second approximation procedure has approximation ratio $2m$, so it applies well in case the sets are small.

Procedure SMALL_SETS

1. For every $1 \leq i \leq k$, draw a vertex $\tilde{u}_i \in A_i$ uniformly at random.
2. Let $\tilde{U} = \{\tilde{u}_1, \dots, \tilde{u}_k\}$.
3. For every $1 \leq i \leq k$ do:
 - (a) Compute the degree $\deg(w_i^j, \tilde{U})$ for every vertex $w_i^j \in W_i$.
 - (b) Let \tilde{w}_i be the vertex with maximum degree.
4. Let $\tilde{W} = \{\tilde{w}_1, \dots, \tilde{w}_k\}$.
5. Output the cover $\tilde{C} = (\tilde{U}, \tilde{W})$.

For the analysis, we need to argue that the cover $\tilde{C} = (\tilde{U}, \tilde{W})$ constructed by Procedure SMALL_SETS is not much worse than the optimal cover $C^* = U^* \cup W^*$. To do that, let us first consider the intermediary “mixed” cover $\bar{C} = (\tilde{U}, W^*)$.

By the choice of \tilde{W} , it is clear that once \tilde{U} is fixed, W^* is no better than \tilde{W} . Hence comparing \tilde{C} to \bar{C} , the following claim is immediate.

Lemma 2.2. $f(\tilde{C}) \geq f(\bar{C})$.

On the other hand, comparing \bar{C} to C^* we have:

Lemma 2.3. $\mathbb{E}(f(\bar{C})) \geq \frac{1}{m} \cdot f(C^*)$.

Proof. Let $\tilde{d}_i = \deg(\tilde{u}_i, W^*)$ and $d_i^* = \deg(u_i^*, W^*)$. Observe that if $\tilde{u}_i = v_i^*$ then $\tilde{d}_i = d_i^*$. As this happens with probability $1/m$, we have that $\mathbb{E}(\tilde{d}_i) \geq \frac{1}{m} \cdot d_i^*$.

Noting that $f(C^*) = \sum_{i=1}^k d_i^*$ and $f(\bar{C}) = \sum_{i=1}^k \tilde{d}_i$, we conclude that

$$\mathbb{E}(f(\bar{C})) = \sum_{i=1}^k \mathbb{E}(\tilde{d}_i) \geq \frac{1}{m} \sum_{i=1}^k d_i^* = \frac{1}{m} \cdot f(C^*). \quad \square$$

Corollary 2.4. $\mathbb{E}(f(\tilde{C})) \geq \frac{1}{m} \cdot f(C^*)$.

To get this result with high probability, we apply the following procedure.

Procedure SMALL_SETS_2

1. Set $\ell = 2m \log n$.
2. Invoke Procedure SMALL_SETS for ℓ times.
3. Select the best result.

We rely on the following elementary fact.

Fact 2.5. *If X is a random variable in the range $[0, m\alpha]$ with expectation $\mathbb{E}(X) = \alpha$, then the probability that $X \leq \alpha/2$ is at most $1 - \frac{1}{2m}$.*

By applying the above fact to the random variable $f(\tilde{C})$ with $\alpha = f(C^*)/m$, we get that in each invocation of Procedure SMALL_SETS, the probability that the gain of the resulting cover \tilde{C} is $f(\tilde{C}) \leq f(C^*)/2m$ is at most $1 - \frac{1}{2m}$. Subsequently, the probability that the gain of none of the ℓ covers exceeds $f(C^*)/2m$ is at most $(1 - \frac{1}{2m})^\ell \approx 1/n$.

Corollary 2.6. *With probability at least $1 - 1/n$, Procedure SMALL_SETS_2 yields a $2m$ -approximation for the Max-Rep problem.*

Finally, by combining Lemma 2.1 and Corollary 2.6 we conclude that applying both procedures FEW_SETS and SMALL_SETS_2 and selecting the better result yields an approximation ratio of $\min\{k, 2m\}$. As $n = 2km$, either $k \leq \sqrt{n}$ or $2m \leq \sqrt{n}$ must hold, hence we have the following.

Theorem 2.7. *There is a randomized algorithm yielding an approximation with ratio \sqrt{n} with probability at least $1 - 1/n$ for the Max-Rep problem.*

Let us remark that a simple deterministic variant of Procedure SMALL_SETS, and hence of the entire algorithm, was pointed out by Y. Hassin [9].

2.2. The nonuniform case

Let us now generalize the approximation algorithm to the case where the partitioning of the graph is nonuniform, i.e., there are k_U sets A_i and k_W sets B_i , and each of those sets is possibly of different size.

It is easy to verify that the procedures described earlier still work correctly, albeit with weaker approximation ratios. In particular, Procedure FEW_SETS will guarantee approximation ratio at most k_U . Analogously, a dual procedure FEW_SETS_2 which reverses the roles of the sets U and W (i.e., selects the best vertex $\hat{w} \in W$ and bases the cover on \hat{w} and its neighbors in U) will yield ratio k_W . Procedure SMALL_SETS_2 will guarantee approximation ratio at most $\hat{m}(\mathcal{H}) = \max\{|A_i| \mid 1 \leq i \leq k_U\}$. Unfortunately, in a nonuniform instance, all of these bounds might be simultaneously as large as $\Omega(n)$.

To get a better bound, we partition the problem into four subproblems as follows. First, split the sets A_i and B_i into large and small ones, letting

$$\begin{aligned} \mathcal{A}_L &= \{A_i \mid |A_i| \geq \sqrt{n}\}, & \mathcal{A}_S &= \{A_i \mid |A_i| < \sqrt{n}\}, \\ \mathcal{B}_L &= \{B_i \mid |B_i| \geq \sqrt{n}\}, & \mathcal{B}_S &= \{B_i \mid |B_i| < \sqrt{n}\}, \end{aligned}$$

and taking

$$\begin{aligned} U_L &= \bigcup_{A_i \in \mathcal{A}_L} A_i, & U_S &= U \setminus U_L, \\ W_L &= \bigcup_{B_i \in \mathcal{B}_L} B_i, & W_S &= W \setminus W_L. \end{aligned}$$

The edge set E is partitioned accordingly into four subsets

$$E_{XY} = E \cap U_X \times W_Y, \quad \text{for } X, Y \in \{L, S\}.$$

The problem now splits into four subproblems, denoted Π_{XY} for $X, Y \in \{L, S\}$, where Π_{XY} is defined over the graph $G_{XY} = (U_X, W_Y, E_{XY})$ and the super-graph \mathcal{H}_{XY} induced by the super-vertices of \mathcal{A}_X and \mathcal{B}_Y .

Clearly, each cover C for the original problem induces four covers C_{XY} for the subproblems, with

$$f(C) = f(C_{LL}) + f(C_{LS}) + f(C_{SL}) + f(C_{SS}).$$

Subsequently, if it is possible to approximate *each* of the four subproblems Π_{XY} for $X, Y \in \{L, S\}$ separately, giving it a cover C_{XY} with ratio at most γ , then we can guarantee an approximation for the original problem with ratio at most 4γ , simply by taking the largest of the resulting four covers and completing it arbitrarily.

The crucial observation is that on subproblems Π_{LL} and Π_{LS} we have $k_{U_L} \leq \sqrt{n}$, so Procedure FEW_SETS yields approximation ratio \sqrt{n} . Likewise, on subproblem Π_{SL} we have $k_{W_L} \leq \sqrt{n}$, so the dual Procedure FEW_SETS_2 again yields approximation ratio \sqrt{n} . Finally, on subproblem Π_{SS} we have $\hat{m}(\mathcal{H}_{SS}) \leq \sqrt{n}$, so Procedure SMALL_SETS_2 will yield approximation ratio \sqrt{n} .

Theorem 2.8. *There is a randomized algorithm yielding an approximation with ratio $4\sqrt{n}$ with probability at least $1 - 1/n$ for the Nonuniform Max-Rep problem.*

2.3. The weighted problem

Finally, let us consider the weighted variant of the problem, in which every super-edge (A_i, B_j) has a nonnegative real weight $\omega(A_i, B_j)$ associated with it, and the goal is to maximize the weight of the selected cover. We claim that Procedures FEW_SETS, FEW_SETS_2, SMALL_SETS and SMALL_SETS_2 can be extended to the weighted setting with no change in the approximation ratio. This can be shown by defining appropriate generalizations of the deg and sdeg functions which take the super-edge weights into account.

In particular, for a set E' of super-edges, let $\omega(E') = \sum_{e \in E'} \omega(e)$. Let $\text{sdeg}(u, \mathcal{H})$ denote the *weighted* super-degree of u in \mathcal{H} , namely, the total weight $\omega(\{(A_i, B_i) \mid B_i \in \Gamma(u, \mathcal{H})\})$. The super-degree of a vertex now represents the total weight of the super-edges it can potentially cover. Denote the weight of the set $\widehat{E}(C)$ by $f(C) = \omega(\widehat{E}(C))$. Finally, for every $w \in W'$, let $B(w)$ denote the set to which w belongs. For $u \in A_I$, denote the set of neighbors of u in G by $\Gamma(u, G)$. Then define $\text{deg}(u, W') = \sum_{w \in W' \cap \Gamma(u, G)} \omega(A_i, B(w))$. With these definitions, the entire analysis goes through with little change. We thus get the following.

Theorem 2.9. *There is a randomized algorithm yielding an approximation with ratio $4\sqrt{n}$ with probability at least $1 - 1/n$ for the Nonuniform Weighted Max-Rep problem.*

3. An approximation algorithm for the Red-Blue Set Cover problem

Let us first give some definitions. For every red element $r_i \in R$ and set collection \mathcal{S} , let $\text{deg}(r_i, \mathcal{S})$ denote the number of sets in \mathcal{S} that contain r_i . Let $\Delta(\mathcal{S}) = \max\{\text{deg}(r_i, \mathcal{S}) \mid r_i \in R\}$.

For a set S_i , a set collection \mathcal{S} and a subset $R' \subseteq R$ of red elements, denote the set obtained by discarding the elements of R' from S_i by $\Phi(S_i, R') = S_i \setminus R'$, and let

$$\Phi(\mathcal{S}, R') = \{\Phi(S_i, R') \mid S_i \in \mathcal{S}'\}.$$

For every set $S_i \in \mathcal{S}$, let $r(S_i) = |R(\{S_i\})|$, and for every subcollection $\mathcal{S}' \subseteq \mathcal{S}$, let $r(\mathcal{S}') = |R(\mathcal{S}')|$.

Let \mathcal{S}^* denote the optimal solution for the Red-Blue Set Cover problem on the instance \mathcal{S} .

3.1. The greedy procedure

We make use of the following approximation procedure for the Red-Blue Set Cover problem.

Procedure GREEDY_RB

1. Modify \mathcal{S} into an instance \mathcal{T} of the weighted set cover problem as follows.
 - (a) Take $\mathcal{T} = \Phi(\mathcal{S}, R)$,
 - (b) Assign each set $T_i = \Phi(S_i, R)$ in \mathcal{T} a weight $\omega(T_i) = r(S_i)$.
2. Apply the greedy algorithm for weighted set cover to \mathcal{T} , and generate a cover $\widetilde{\mathcal{T}}$.
3. Take the corresponding collection $\widetilde{\mathcal{S}} = \{S_i \mid T_i \in \widetilde{\mathcal{T}}\}$ as the resulting approximation.

For every subcollection \mathcal{T}' of an instance \mathcal{T} of the weighted set cover problem, let $\omega(\mathcal{T}') = \sum_{T_i \in \mathcal{T}'} \omega(T_i)$. It is easy to verify the following.

Lemma 3.1. *For any set collection \mathcal{S}' and corresponding instance $\mathcal{T}' = \Phi(\mathcal{S}', R)$ of the weighted set cover problem,*

$$r(\mathcal{S}') \leq \omega(\mathcal{T}') \leq \Delta(\mathcal{S}) \cdot r(\mathcal{S}').$$

Proof. Note that

$$\omega(\mathcal{T}') = \sum_{T_i \in \mathcal{T}'} \omega(T_i) = \sum_{S_i \in \mathcal{S}'} r(S_i) = \sum_{r_j \in R(\mathcal{S}')} \text{deg}(r_j, \mathcal{S}').$$

As $1 \leq \deg(r_j, \mathcal{S}') \leq \Delta(\mathcal{S}') \leq \Delta(\mathcal{S})$ for every $r_j \in R(\mathcal{S}')$, we get

$$r(\mathcal{S}') \leq \sum_{r_j \in R(\mathcal{S}')} \deg(r_j, \mathcal{S}') \leq \Delta(\mathcal{S}) \cdot r(\mathcal{S}'),$$

implying the claim. \square

Lemma 3.2. Procedure GREEDY_RB has an approximation ratio of $\Delta(\mathcal{S}) \cdot \log \beta$.

Proof. Denote the minimum-weight set cover for \mathcal{T} by $\mathcal{T}^\#$, and let $\mathcal{T}^* = \Phi(\mathcal{S}^*, R)$ be the instance of the weighted set cover problem corresponding to \mathcal{S}^* . (Note that $\mathcal{T}^\#$ and \mathcal{T}^* need not necessarily be the same.) It is known that the greedy algorithm yields a $\log \beta$ approximation for the weighted set cover problem, namely, $\omega(\tilde{\mathcal{T}}) \leq \log \beta \cdot \omega(\mathcal{T}^\#)$ [4]. Therefore, by Lemma 3.1,

$$r(\tilde{\mathcal{S}}) \leq \omega(\tilde{\mathcal{T}}) \leq \log \beta \cdot \omega(\mathcal{T}^\#).$$

The optimality of $\mathcal{T}^\#$ implies that $\omega(\mathcal{T}^\#) \leq \omega(\mathcal{T}^*)$. Combined, we get that

$$r(\tilde{\mathcal{S}}) \leq \log \beta \cdot \omega(\mathcal{T}^*).$$

Applying Lemma 3.1 again we get that

$$r(\tilde{\mathcal{S}}) \leq \log \beta \cdot \Delta(\mathcal{S}) \cdot r(\mathcal{S}^*). \quad \square$$

3.2. The main procedure

For an integer parameter X , we consider the following procedure.

Procedure LOW_DEG(X)

1. Discard from \mathcal{S} the sets with more than X red elements, setting $\mathcal{S}_X \leftarrow \{S_i \in \mathcal{S} \mid r(S_i) \leq X\}$.
2. If $B(\mathcal{S}_X) \neq B$ then return \mathcal{S} . /* \mathcal{S}_X is not feasible */
3. Set $Y = \sqrt{n/\log \beta}$
4. Separate the red elements into *high* and *low* degree elements, setting $R_H \leftarrow \{r_i \in R \mid \deg(r_i, \mathcal{S}_X) > Y\}$ and $R_L \leftarrow R \setminus R_H$.
5. Discard the elements of R_H from \mathcal{S}_X , setting $\mathcal{S}_{X,Y} \leftarrow \Phi(\mathcal{S}_X, R_H)$.
6. Apply Procedure GREEDY_RB to $\mathcal{S}_{X,Y}$, and obtain a solution $\tilde{\mathcal{S}}_{X,Y}$.
7. Complete the sets of $\tilde{\mathcal{S}}_{X,Y}$ into the corresponding sets of \mathcal{S}_X (by adding to each set $T_i \in \tilde{\mathcal{S}}_{X,Y}$ originally obtained from $S_i \in \mathcal{S}_X$ the discarded elements $S_i \cap R_H$), and return the resulting solution $\tilde{\mathcal{S}}_X$.

Lemma 3.3. $|R_H| \leq \sqrt{n \log \beta} \cdot X$.

Proof. Each set $S_i \in \mathcal{S}_X$ has at most X red elements. Hence

$$|R_H| \cdot Y < \sum_{r_j \in R_H} \deg(r_j, \mathcal{S}_X) \leq \sum_{r_j \in R} \deg(r_j, \mathcal{S}_X) = \sum_{S_i \in \mathcal{S}_X} r(S_i) \leq |\mathcal{S}_X| \cdot X \leq nX,$$

so $|R_H| \leq nX/Y$, implying the lemma. \square

3.3. The approximation algorithm

Now let us set $\hat{X} = \max\{r(\mathcal{S}_i^*) \mid \mathcal{S}_i^* \in \mathcal{S}^*\}$, and consider the performance of Procedure LOW_DEG when invoked with the parameter $X = \hat{X}$.

Lemma 3.4. Procedure LOW_DEG(\hat{X}) yields an approximation ratio of at most $2\sqrt{n \log \beta}$.

Proof. First observe, that $\mathcal{S}_{\widehat{X}}$ is necessarily feasible. Hence the procedure will always return a solution in its Step 7 (and not Step 2).

Let \mathcal{S}^* be some optimal solution for the problem, and let $r_H^* = |R(\mathcal{S}^*) \cap R_H|$ and $r_L^* = |R(\mathcal{S}^*) \cap R_L|$. Since $\Delta(\mathcal{S}_{\widehat{X}}, Y) \leq Y$, Lemma 3.2 guarantees that the solution produced by Procedure LOW_DEG(\widehat{X}) uses at most $Y \cdot \log \beta \cdot r_L^* = \sqrt{n \log \beta} \cdot r_L^*$ red elements of R_L^* . By Lemma 3.3, the number of red elements of R_H^* contained in the solution generated by the procedure is at most $\sqrt{n \log \beta} \cdot \widehat{X}$. Combined, the total number of red elements used by the procedure satisfies $r(\widetilde{\mathcal{S}}_{\widehat{X}}) \leq \sqrt{n \log \beta} \cdot r_L^* + \sqrt{n \log \beta} \cdot \widehat{X}$. But by the definition of \widehat{X} , necessarily $r(\mathcal{S}^*) \geq \widehat{X}$, and hence $r(\widetilde{\mathcal{S}}_{\widehat{X}}) \leq 2\sqrt{n \log \beta} \cdot r(\mathcal{S}^*)$, yielding the lemma. \square

As \widehat{X} is not known to us in advance, it will be necessary to search for it. This yields our final algorithm.

Algorithm LOW_DEG2

1. **For** $X = 1$ to ρ **do**:

Invoke Procedure LOW_DEG(X).

2. Take the best of the obtained solutions.

Theorem 3.5. Algorithm LOW_DEG2 yields an approximation ratio of $2\sqrt{n \log \beta}$ for the Red-Blue Set Cover problem.

A minor variant of this algorithm yields an approximation ratio of $(n\rho)^{1/3}$. However, the problem clearly admits also a trivial approximation algorithm of ratio ρ , and $(n\rho)^{1/3}$ is always dominated by the smaller of ρ and \sqrt{n} , so this variant is not as interesting (assuming the factor of $\log \beta$ is negligible compared to the other terms).

3.4. The weighted case

Finally, let us consider the weighted variant of the problem, in which every red element $r_i \in R$ has a positive real weight $\omega(r_i)$ associated with it, and the goal is to minimize the weight of the selected cover. We claim that Procedures GREEDY_RB and LOW_DEG can be extended to the weighted setting with no change in the approximation ratio. In particular, in addition to the previous definitions, define the weight of a set S_i to be $\omega(S_i) = \sum_{r_j \in S_i} \omega(r_j)$, and for a subcollection \mathcal{S}' let $\omega(\mathcal{S}') = \sum_{r_j \in R(\mathcal{S}')} \omega(r_j)$. In Procedure GREEDY_RB, Step 1(b) should assign each set T_i in \mathcal{T} the weight $\omega(T_i) = \omega(S_i)$. The inequalities of Lemma 3.1 become

$$\omega(\mathcal{S}') \leq \omega(\mathcal{T}') \leq \Delta(\mathcal{S}) \cdot \omega(\mathcal{S}'),$$

with minimal changes in the proof, as well as in the proof of Lemma 3.2. In Procedure LOW_DEG, the definition of \mathcal{S}_X changes to $\mathcal{S}_X \leftarrow \{S_i \in \mathcal{S} \mid \omega(S_i) \leq X\}$. As a result, Lemma 3.3 now asserts that $\omega(R_H) \leq \sqrt{n \log \beta} \cdot X$. The definition of \widehat{X} becomes $\widehat{X} = \max\{\omega(S_i^*) \mid S_i^* \in \mathcal{S}^*\}$. The proof of Lemma 3.4 uses $\omega_H^* = \omega(R(\mathcal{S}^*) \cap R_H)$ and $\omega_L^* = \omega(R(\mathcal{S}^*) \cap R_L)$ instead of r_H^* and r_L^* , respectively.

We thus get the following.

Theorem 3.6. There is an algorithm with approximation ratio $2\sqrt{n \log \beta}$ for the Weighted Red-Blue Set Cover problem.

Acknowledgements

I am grateful to Michael Elkin, Yehuda Hassin and Hadas Taubman for helpful discussions.

References

- [1] M. Alekhovich, S. Buss, S. Moran, T. Pitassi, Minimum propositional proof length is NP-hard to linearly approximate, J. Symbolic Logic 66 (1) (2001) 171–191.

- [2] S. Arora, L. Babai, J. Stern, Z. Sweedyk, The hardness of approximate optima in lattices, codes and linear equations, in: Proc. 34th IEEE Symp. on Foundations of Computer Science, 1993, pp. 724–733.
- [3] R. Carr, S. Doddi, G. Konjevod, M. Marathe, On the red-blue set cover problem, in: Proc. 11th ACM-SIAM Symp. on Discrete Algorithms, 2000, pp. 345–353.
- [4] V. Chvatal, A greedy heuristic for the set-covering problem, *Math. Oper. Res.* 4 (1979) 233–235.
- [5] I. Dinur, S. Safra, On the hardness of approximating label cover, *Inform. Process. Lett.* 89 (5) (2004) 247–254.
- [6] M. Elkin, D. Peleg, The hardness of approximating spanner problems, in: Proc. 17th Symp. on Theoretical Aspects of Computer Science, February 2000, pp. 370–381.
- [7] M. Elkin, D. Peleg, Strong inapproximability of the basic k -spanner problem, in: Proc. 27th Int. Colloq. on Automata, Languages & Prog., in: Lecture Notes in Computer Science, vol. 1853, Springer-Verlag, Berlin, July 2000, pp. 636–647.
- [8] M.H. Goldwasser, R. Motwani, Intractability of assembly sequencing: Unit disks in the plane, in: Proc. Workshop on Algorithms and Data Structures, in: Lecture Notes in Computer Science, vol. 1272, Springer-Verlag, Berlin, 1997, pp. 307–320.
- [9] Y. Hassin, Private communication, 2000.
- [10] D.S. Hochbaum, *Approximation Algorithms for NP-Hard Problems*, PWS Publishing Co., Boston, MA, 1995 (Chapter 10.4).
- [11] G. Kortsarz, On the hardness of approximating spanners, *Algorithmica* 30 (3) (2001) 432–450.