A 0.821-Ratio Purely Combinatorial Algorithm for Maximum k-vertex Cover in Bipartite Graphs

Édouard Bonnet[1], Bruno Escoffier[2], Vangelis Th. Paschos[3,4(✉)], and Georgios Stamoulis[3,4]

[1] Institute for Computer Science and Control,
Hungarian Academy of Sciences (MTA SZTAKI), Budapest, Hungary
bonnet.edouard@sztaki.mta.hu

[2] Sorbonne Universités,
UPMC Universite Paris 06, CNRS, LIP6 UMR 7606, Paris, France
bruno.escoffier@lip6.fr

[3] PSL* Research University, Université Paris-Dauphine, LAMSADE, Paris, France

[4] CNRS UMR 7243, Paris, France
paschos@lamsade.dauphine.fr, georgios.stamoulis@dauphine.fr

Abstract. We study the polynomial time approximation of the MAX k-VERTEX COVER problem in bipartite graphs and propose a purely *combinatorial algorithm* that beats the only such known algorithm, namely the greedy approach. We present a computer-assisted analysis of our algorithm, establishing that the worst case approximation guarantee is bounded below by 0.821.

1 Introduction

In MAX k-VERTEX COVER, a graph $G = (V, E)$ with $|V| = n$ and $|E| = m$ is given together with an integer $k \leqslant n$. The goal is to find a subset $K \subseteq V$ with k vertices such that the total number of edges covered by K is maximized. We say that an edge $e = \{u, v\}$ is covered by a subset of vertices K if $K \cap e \neq \emptyset$. MAX k-VERTEX COVER is **NP**-hard in general graphs (as a generalization of MIN VERTEX COVER) and it remains so in bipartite graphs [1,2].

The approximation of MAX k-VERTEX COVER has been originally studied in [3], where ratio $1 - (1/e)$ (≈ 0.632) is achieved by the natural greedy algorithm. This ratio is tight even in bipartite graphs [4]. Using a sophisticated linear programming method, the approximation ratio for MAX k-VERTEX COVER was improved to 3/4 [5], which, until very recently, was the best known ratio even in bipartite graphs. The best approximation ratio in bipartite graphs is now 8/9 and is still based on linear programming [2]. A direct reduction from MIN VERTEX COVER shows that MAX k-VERTEX COVER can not admit a polynomial time approximation schema (PTAS), unless $\mathbf{P} = \mathbf{NP}$ [6].

Finally, we may observe that MAX k-VERTEX COVER is easy in semiregular bipartite graphs (where all the vertices of each color class have the same degree).

Indeed, any k vertices in the color class of maximum degree yield an optimal solution. Obviously, if this color class contains less than k vertices, then one can cover *all* the edges.

Our Contribution. The principal motivation of this paper is to determine *to what extent combinatorial methods compete with linear programming* for MAX k-VERTEX COVER. In other words, *what ratio can a purely combinatorial algorithm guarantee?* To this purpose, we first devise a very simple algorithm that guarantees approximation ratio $2/3$, improving so the ratio of the greedy algorithm in bipartite graphs. The proof is given in [7]. Our main contribution consists of an approximation algorithm which computes six distinct solutions and returns the best among them.

Analyzing the performance guarantee of such an algorithm is a challenging task. Indeed, there is no obvious way to compare the different solutions and argue globally over a lower bound on the maximum value taken by the six solutions. The large number of variables (in all, 48) used to express the many solution values participates in the difficulty of the analysis.

Similar situation was faced, for example, in [8] where the authors gave a 0.921 approximation guarantee for MAX CUT in graphs of maximum degree 3 (and an improved 0.924 for 3-regular graphs) by a computer-assisted analysis of the quantities generated by theoretically analyzing a particular semi-definite relaxation of the problem at hand. Similarly, by setting up a suitable non-linear program and solving it, we give a computer-assisted analysis of a 0.821-approximation guarantee for MAX k-VERTEX COVER in bipartite graphs. We give all the details of the implementation in [7].

2   Preliminaries

Consider a bipartite graph $B = (V_1, V_2, E)$, instance of MAX k-VERTEX COVER and fix an optimal solution O (i.e., a set of k vertices covering a maximum number of edges in E) as well as its parts O_1 and O_2 lying in color classes V_1 and V_2, respectively.

The algorithm (called k-VC_ALGORITHM) proposed for solving MAX k-VERTEX COVER can be sketched as follows:
1. guess the cardinality k_1 and therefore $k_2 = k - k_1$ of its subsets O_1 and O_2 lying in the color classes V_1 and V_2, respectively;
2. compute the sets S_i of k_i vertices in V_i, $i = 1, 2$ that cover the most of edges; obviously S_i is a set of the k_i largest degree vertices in V_i (breaking ties arbitrarily);
3. guess the cardinalities k_i' of the intersections $S_i \cap O_i$, $i = 1, 2$;
4. compute the sets X_i of the $k_i - k_i'$ vertices from V_i, $i = 1, 2$, that cover the most of edges in graphs $B[(V \setminus S_1) \cup V_2]$ and $B[V_1 \cup (V_2 \setminus S_2)]$, respectively;
5. choose the best among six solutions built as described in Sect. 3.

Let us note that our $2/3$-approximation algorithm in [7] guarantees ratio $4/5$, when both $k_i' = 0$, $i = 1, 2$.
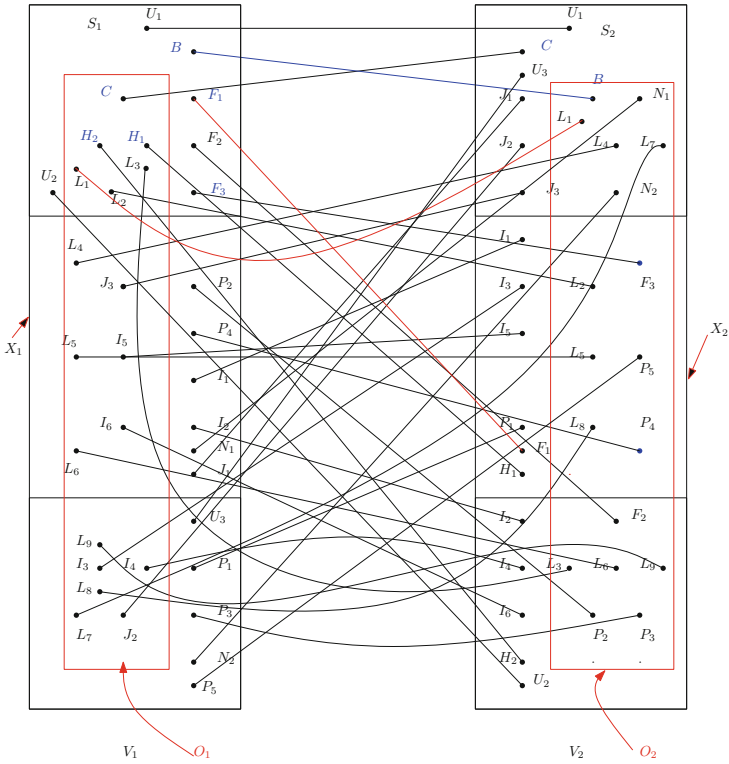
Fig. 1. Sets S_i, O_i, X_i $i = 1, 2$ and cuts between them.

Sets S_i, X_i and O_i separate each color class in 6 regions, namely, $S_i \cap O_i$, $S_i \setminus O_i$, $X_i \cap O_i$, $X_i \setminus O_i$, $O_i \setminus (S_i \cup X_i)$ (denoted by $\bar{O}_i$, in what follows) and $V_i \setminus (S_i \cup X_i \cup O_i)$. In total, there exist 36 groups of edges (cuts) among them, the group $(V_1 \setminus (S_1 \cup X_1 \cup O_1), V_2 \setminus (S_2 \cup X_2 \cup O_2))$ being irrelevant as it will become clear in the sequel. We will use the following notations to refer to the values of the 35 relevant cuts (illustrated in Fig. 1.):

B: the number of edges in the cut $(S_1 \setminus O_1, S_2 \cap O_2)$;

C: the number of edges in the cut $(S_2 \setminus O_2, S_1 \cap O_1)$;

F_1, F_2, F_3: the number of edges in the cuts $(S_1 \setminus O_1, X_2 \setminus O_2)$, $(S_1 \setminus O_1, O_2 \setminus (X_2 \cup S_2))$ and $(S_1 \setminus O_1, O_2 \cap X_2)$, respectively;

H_1, H_2: the number of edges in the cuts $(S_1 \cap O_1, X_2 \setminus O_2)$ and $(S_1 \cap O_1, V_2 \setminus (S_2 \cup X_2 \cup O_2))$, respectively;

$\{I_i\}_{i \in [6]}$: the number of edges in the cuts $(X_1 \setminus O_1, X_2 \setminus O_2)$, $(X_1 \setminus O_1, V_2 \setminus (S_2 \cup X_2 \cup O_2))$, $(O_1 \setminus (S_1 \cup X_1), X_2 \setminus O_2)$, $(O_1 \setminus (S_1 \cup X_1), V_2 \setminus (S_2 \cup X_2 \cup O_2))$, $(X_1 \cap O_1, X_2 \setminus O_2)$ and $(X_1 \cap O_1, V_2 \setminus (S_2 \cup X_2 \cup O_2))$, respectively;

J_1, J_2, J_3: the number of edges in the cuts $(S_2 \setminus O_2, X_1 \setminus O_1)$, $(S_2 \setminus O_2, O_1 \setminus (S_1 \cup X_1))$ and $(S_2 \setminus O_2, O_1 \cap X_1)$, respectively;

$\{L_i\}_{i \in [9]}$: the number of edges in the cuts $(S_1 \cap O_1, S_2 \cap O_2)$, $(S_1 \cap O_1, X_2 \cap O_2)$, $(S_1 \cap O_1, O_2 \setminus (S_2 \cup X_2))$, $(X_1 \cap O_1, S_2 \cap O_2)$, $(X_1 \cap O_1, X_2 \cap O_2)$, $(X_1 \cap O_1, O_2 \setminus (S_2 \cup X_2))$, $(O_1 \setminus (S_1 \cup X_1), S_2 \cap O_2)$, $(O_1 \setminus (S_1 \cup X_1), X_2 \cap O_2)$, and $(O_1 \setminus (S_1 \cup X_1), O_2 \setminus (S_2 \cup X_2))$, respectively;

N_1, N_2: the number of edges in the cuts $(S_2 \cap O_2, X_1 \setminus O_1)$ and $(S_2 \cap O_2, V_1 \setminus (S_1 \cup X_1 \cup O_1))$, respectively;

$\{P_i\}_{i \in [5]}$: the number of edges in the cuts $(X_2 \setminus O_2, V_1 \setminus (S_1 \cup X_1 \cup O_1))$, $(O_2 \setminus (S_2 \cup X_2), X_1 \setminus O_1)$, $(O_2 \setminus (S_2 \cup X_2), V_1 \setminus (S_1 \cup X_1 \cup O_1))$, $(X_2 \cap O_2, X_1 \setminus O_1)$, and $(X_2 \cap O_2, V_1 \setminus (S_1 \cup X_1 \cup O_1))$, respectively;

U_1, U_2, U_3: the number of edges is the cuts, $(S_1 \setminus O_1, S_2 \setminus O_2)$, $(S_1 \setminus O_1, V_2 \setminus (S_2 \cup X_2 \cup O_2))$ and $(S_2 \setminus O_2, V_1 \setminus (S_1 \cup X_1 \cup O_1))$, respectively.

Denoting by $\delta(V')$, $V' \subseteq V$, the number of edges covered by V' and by $\mathrm{opt}(B)$ the value of an optimal solution (i.e., the number edges covered) for MAX k-VERTEX COVER in the input graph B, the following holds (see also Fig. 1):

$$\delta(S_1) = B + C + F_1 + F_2 + F_3 + H_1 + H_2 + L_1 + L_2 + L_3 + U_1 + U_2 \quad (1)$$

$$\delta(S_2) = B + C + J_1 + J_2 + J_3 + L_1 + L_4 + L_7 + N_1 + N_2 + U_1 + U_3 \quad (2)$$

$$\delta(X_1) = I_1 + I_2 + I_5 + I_6 + J_1 + J_3 + \sum_{i=4}^{6} L_i + N_1 + P_2 + P_4 \quad (3)$$

$$\delta(X_2) = F_1 + F_3 + H_1 + I_1 + I_3 + I_5 + L_2 + L_5 + L_8 + P_1 + P_4 + P_5 \quad (4)$$

$$\delta(O_1) = C + H_1 + H_2 + I_3 + I_4 + I_5 + I_6 + J_2 + J_3 + \sum_{i=1}^{9} L_i \quad (5)$$

$$\delta(O_2) = B + F_2 + F_3 + \sum_{i=1}^{9} L_i + N_1 + N_2 + \sum_{i=2}^{5} P_i \quad (6)$$

$$\mathrm{opt}(B) = B + C + \sum_{i=2}^{3} F_i + \sum_{i=1}^{2} H_i + \sum_{i=3}^{6} I_i + \sum_{i=2}^{3} J_i + \sum_{i=1}^{9} L_i$$
$$+ \sum_{i=1}^{2} N_i + \sum_{i=2}^{5} P_i \quad (7)$$

Without loss of generality, we assume $k_1 \leqslant k_2$ and we set: $k_1 = \mu k_2$ ($\mu \leqslant 1$), $k_1' = |S_1 \cap O_1| = \nu k_1$ ($0 \leqslant \nu \leqslant 1$) and $k_2' = |S_2 \cap O_2| = \xi k_2$ ($0 \leqslant \xi \leqslant 1$). Let us note that, since k_i' vertices lie in the intersections $S_i \cap O_i$, the following hold for $\bar{O}_i = O_i \setminus (S_i \cup X_i)$, $i = 1, 2$: $|\bar{O}_1| = |O_1 \setminus (S_1 \cup X_1)| \leqslant (1 - \nu)k_1 = \mu(1 - \nu)k_2$ and $|\bar{O}_2| = |O_2 \setminus (S_2 \cup X_2)| \leqslant (1 - \xi)k_2$. From the definitions of the cuts and using (1) to (6) and the expressions for $|\bar{O}_1|$ and $|\bar{O}_2|$, simple average arguments and the assumptions for k_1, k_2, k_1' and k_2' just above, the following holds:

$$
\begin{aligned}
\delta(S_1) &\geq \delta(O_1) \\
\delta(S_2) &\geq \delta(O_2) \\
\delta(X_1) + C + H_1 + H_2 + L_1 + L_2 + L_3 &\geq \delta(O_1) \\
\delta(X_2) + B + N_1 + N_2 + L_1 + L_4 + L_7 &\geq \delta(O_2) \\
\delta(S_1) &\geq 1/(1-\nu) \cdot \delta(X_1) \\
\delta(S_2) &\geq 1/(1-\xi) \cdot \delta(X_2) \\
\delta(S_1) + \delta(X_1) &\geq (2-\nu)/(1-\nu) \cdot (I_3 + I_4 + J_2 + L_7 + L_8 + L_9) \\
\delta(S_2) + \delta(X_2) &\geq (2-\xi)/(1-\xi) \cdot (F_2 + L_3 + L_6 + L_9 + P_2 + P_3) \\
B + F_1 + F_2 + F_3 + U_1 + U_2 &\geq \delta(X_1) \\
C + J_1 + J_2 + J_3 + U_1 + U_3 &\geq \delta(X_2)
\end{aligned}
\tag{8}
$$

For $i = 1, 2$, the two first inequalities in (8) hold because S_i is the set of k_i highest-degree vertices in V_i; the third and fourth ones because the lefthand side quantities are the number of edges covered by $X_i \cup (S_i \cap O_i)$; each of these sets has cardinality k_i and obviously covers more edges than O_i; the fifth and sixth inequalities because the average degree of S_i is at least the average degree of X_i and $|X_1| = (1 - \nu)k_1$ and $|X_2| = (1 - \xi)k_2$; seventh and eighth ones because the average degree of vertices in $S_i \cup X_i$ is at least the average degree of vertices in $O_i \setminus (S_i \cup X_i)$; finally, for the last two inequalities the sum of degrees of the $k_i - k_i'$ vertices in $S_i \setminus O_i$ is at least the sum of degrees of the $k_i - k_i'$ vertices of X_i.

In Sect. 3, we specify the approximation algorithm sketched above. In [7] a computer assisted analysis of its approximation-performance is presented. The non-linear program that we set up, not only computes the approximation ratio of our algorithm but it also provides an experimental study over families of graphs. Indeed, a particular configuration on the variables (i.e., a feasible value assignments on the variables that represent the set of edges $B, C, \dots$) corresponds to a particular family of bipartite graphs with similar structural properties (characterized by the number of edges belonging to the several cut considered). Given such a configuration, it is immediate to find the ratio of the algorithm, because we can simply substitute the values of the variables in the corresponding ratios and output the largest one. We can view our program as an *experimental analysis* over all families of bipartite graphs, trying to find the particular family that implements the worst case for the approximation ratio of the algorithm. Our program not only finds such a configuration, but also provides data about the range of approximation factor on other families of bipartite graphs. Experimental results show that the approximation factor for the *absolute majority* of the instances is very close to 1, i.e., ≥ 0.95. Moreover, our program is *independent* on the size of the instance. We just need a particular configuration on the relative value of the variables $B, C, \dots$, thus providing a compact way of representing families of bipartite graphs sharing common structural properties.

For the rest of the paper, we call "best" vertices a set of vertices that cover the most of *uncovered* edges[1] in B. Given a solution $\mathrm{SOL}_k(B)$, we denote by $\mathrm{sol}_k(B)$ its value. For the quantities implied in the ratios corresponding to these solutions, one can be referred to Fig. 1 and to expressions (1) to (7).

---

[1] For instance, "we take S_1 plus the k_2 best vertices in V_2" means that we take S_1 and then k_2 vertices of highest degree in $B[(V_1 \setminus S_1), V_2]$.

Let us note that the algorithm above, since it runs for any value of k_1 and k_2, it will run for $(k_1, k_2) = (k, 0)$ and $(0, k)$. So, it will compute the optimum for the instances of [4], where the greedy algorithm attains the ratio $(e-1)/e$. Observe finally that, when $k \geqslant \min\{|V_1|, |V_2|\}$, then $\min\{|V_1|, |V_2|\}$ is an optimal solution since it covers the whole of E. This remark will be useful for some solutions in the sequel, for example in the completion of solution $\mathrm{SOL}_5(B)$.

3 A 0.821-Approximation for the Bipartite Max k-vertex Cover

Algorithm k-VC_ALGORITHM builds the following MAX k-VERTEX COVER-solutions:

$\mathrm{SOL}_1(\mathbf{B})$ and **$\mathrm{SOL}_2(\mathbf{B})$**, take, respectively, S_1 plus the k_2 remaining best vertices from V_2, and S_2 plus the k_1 remaining best vertices from V_1;

$\mathrm{SOL}_3(\mathbf{B})$ takes first $S_1 \cup X_1$ in the solution and completes it with the $(1 - \mu(1 - \nu))k_2$ best vertices from V_2;

$\mathrm{SOL}_4(\mathbf{B})$ takes S_2 and completes it either with vertices from V_2, or with vertices from both V_2 and V_1 (as specified in the next page);

$\mathrm{SOL}_5(\mathbf{B})$ takes a π-fraction of the best vertices in S_1 and X_1, $\pi \in (0, 1/2]$; then, solution is completed with the $k_1 + k_2 - \pi(2k_1 - k'_1)$ best vertices in V_2;

$\mathrm{SOL}_6(\mathbf{B})$ takes a λ-fraction of the best vertices in S_2 and X_2, $\lambda \in (0, (1+\mu)/(2-\xi)]$; then solution is completed with the $k_1 + k_2 - \lambda(2k_2 - k'_2)$ best vertices in V_1.

Let us note that the values of λ and π are parameters that we can fix. In what follows, we analyze solutions $\mathrm{SOL}_1(B) \ldots \mathrm{SOL}_6(B)$ computed by k-VC_ALGORITHM and give analytical expressions for their ratios. A fully detailed analysis of all these solutions is given in [7].

Solution $\mathrm{SOL}_1(\mathbf{B})$. The best k_2 vertices in V_2, provided that S_1 has already been chosen, cover at least the maximum of the following quantities:

$$
\begin{array}{ll}
\mathcal{A}_1 = J_1 + J_2 + J_3 + L_4 + L_7 + N_1 + N_2 + U_3 & \text{by } S_2 \\
\mathcal{A}_2 = I_1 + I_3 + I_5 + L_5 + L_8 + P_1 + P_4 + P_5 & \text{by } X_2 \\
\mathcal{A}_3 = L_4 + L_5 + L_6 + L_7 + L_8 + L_9 + N_1 + N_2 + P_2 + P_3 + P_4 + P_5 & \text{by } O_2
\end{array}
$$

So, the approximation ratio for $\mathrm{SOL}_1(B)$ satisfies:

$$
r_1 \geqslant \frac{\delta(S_1) + \max\left\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\right\}}{\mathrm{opt}(B)}
$$

Solution $\mathrm{SOL}_2(\mathbf{B})$. The best k_1 vertices in V_1, provided that S_2 has already been chosen, cover at least the maximum of the following quantities:

$$
\begin{array}{ll}
\mathcal{B}_1 = H_1 + H_2 + F_1 + F_2 + F_3 + L_2 + L_3 + U_2 & \text{by } S_1 \\
\mathcal{B}_2 = I_1 + I_2 + I_5 + I_6 + L_5 + L_6 + P_2 + P_4 & \text{by } X_1 \\
\mathcal{B}_3 = H_1 + H_2 + I_3 + I_4 + I_5 + I_6 + L_2 + L_3 + L_5 + L_6 + L_8 + L_9 & \text{by } O_1
\end{array}
$$

So, the approximation ratio for $\mathrm{SOL}_2(B)$ satisfies:

$$r_2 \geqslant \frac{\delta\left(S_2\right) + \max\left\{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3\right\}}{\mathrm{opt}(B)}$$

Solution SOL$_3$(B). Taking first $S_1 \cup X_1$ in the solution, $k - (k_1 + k_1 - k_1') = k_1 + k_2 - 2k_1 + k_1' = k_2 - (k_1 - k_1') = (1 - \mu(1 - \nu))k_2$ vertices remain to be taken in V_2. The best such vertices will cover at least the maximum of the following quantities:

$$\mathcal{C}_1 = (1 - \mu(1 - \nu))\left(J_2 + N_2 + L_7 + U_3\right) \tag{9}$$

$$\mathcal{C}_2 = \frac{1 - \mu(1 - \nu)}{2 - \xi}\left(I_3 + J_2 + L_7 + L_8 + N_2 + P_1 + P_5 + U_3\right) \tag{10}$$

$$\mathcal{C}_3 = \frac{1 - \mu(1 - \nu)}{3 - 2\xi}\left(I_3 + J_2 + L_7 + L_8 + L_9 + N_2 + P_1 + P_3 + P_5 + U_3\right) \tag{11}$$

where (9) corresponds to a completion by the $(1 - \mu(1 - \nu))k_2$ best vertices of S_2, (10) corresponds to a completion by the $(1 - \mu(1 - \nu))k_2$ best vertices of $S_2 \cup X_2$, while (11) corresponds to a completion by the $(1 - \mu(1 - \nu))k_2$ best vertices of $S_2 \cup X_2 \cup \bar{O}_2$. The denominator $3 - 2\xi$ in (11) is due to the fact that, using the expression for $\bar{O}_2$, $|S_2 \cup X_2 \cup (O_2 \setminus (S_2 \cup X_2))| \leqslant (3 - 2\xi)k_2$. So, the approximation ratio for $\mathrm{SOL}_3(B)$ is:

$$r_3 \geqslant \frac{\delta\left(S_1\right) + \delta\left(X_1\right) + \max\left\{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3\right\}}{\mathrm{opt}(B)} \tag{12}$$

Solution SOL$_4$(B). Once S_2 is taken in the solution, $k_1 = \mu k_2$ are still to be taken. Completion can be done in the following ways:

1. **if** $k_1 \leqslant k_2 - k_2'$, i.e., $\mu \leqslant 1 - \xi$, completion can be done by vertices taken either from X_2, or from $X_2 \cup \bar{O}_2$; in the first case, the best vertices taken for completion will cover at least either a $\mu/(1-\xi)$ fraction of edges incident to X_2; in the second case, they will cover at least a $\mu/2(1-\xi)$ fraction of edges incident to $X_2 \cup \bar{O}_2$, i.e., at least $\mathcal{M}_1$ edges, where $\mathcal{M}_1$ is given by:

$$\max\left\{\frac{\mu}{1 - \xi}\delta\left(X_2\right), \frac{\mu}{2(1 - \xi)}\left(\delta\left(X_2\right) + F_2 + L_3 + L_6 + L_9 + P_2 + P_3\right)\right\} \tag{13}$$

2. **else**, completion can be done by taking the whole set X_2 and then the additional vertices taken:
   (a) either within the rest of V_2 covering, in particular, a $\min\{1, (\mu - 1 + \xi)/|\bar{O}_2|\} \geqslant \min\{1, (\mu - 1 + \xi)/(1 - \xi)\}$ fraction of edges incident to $\bar{O}_2$ (quantity $\mathcal{M}_2$ in (14)),
   (b) or in S_1 covering, in particular, a $(\mu - 1 + \xi)/\mu$ fraction of uncovered edges incident to S_1 (quantity $\mathcal{M}_3$ in (14)),
   (c) or in $S_1 \cup X_1$ covering, in particular, a $(\mu - 1 + \xi)/\mu(2 - \nu)$ fraction of uncovered edges incident to $S_1 \cup X_1$ (quantity $\mathcal{M}_4$ in (14)),

(d) or, finally, in $S_1 \cup X_1 \cup \bar{O}_1$ covering, in particular, a $(\mu-1+\xi)/\mu(3-2\nu)$ fraction of uncovered edges incident to this vertex-set (quantity $\mathcal{M}_5$ in (14));

in any case such a completion will cover a number of edges that is at least the maximum of the following quantities:

$$
\begin{aligned}
\mathcal{M}_2 &= \min\left\{1, \tfrac{\mu-1+\xi}{1-\xi}\right\}(F_2 + L_3 + L_6 + L_9 + P_2 + P_3) \\
\mathcal{M}_3 &= \tfrac{\mu-1+\xi}{\mu}(F_2 + H_2 + L_3 + U_2) \\
\mathcal{M}_4 &= \tfrac{\mu-1+\xi}{\mu(2-\nu)}(F_2 + H_2 + I_2 + I_6 + L_3 + L_6 + P_2 + U_2) \\
\mathcal{M}_5 &= \tfrac{\mu-1+\xi}{\mu(3-2\nu)}(F_2 + H_2 + I_2 + I_4 + I_6 + L_3 + L_6 + L_9 + P_2 + U_2)
\end{aligned}
\tag{14}
$$

Using (13) and (14), the following holds for the approximation ratio of $\text{SOL}_4(B)$:

$$
r_4 \geqslant \frac{\delta(S_2) + \begin{cases} \mathcal{M}_1 & \mu \leq 1-\xi \\ \delta(X_2) + \max\{\mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4, \mathcal{M}_5\} & \mu \geq 1-\xi \end{cases}}{\text{opt}(B)}
\tag{15}
$$

Vertical Separations – Solutions $\text{SOL}_5(\mathbf{B})$ and $\text{SOL}_6(\mathbf{B})$. For $i = 1, 2$, given a vertex subset $V' \subseteq V_i$, we call *vertical separation of V' with parameter* $c \in (0, 1/2]$, a partition of V' into two subsets such that one of them contains a c-fraction of the best (highest degree) vertices of V' (i.e., contains the $c|V'|$ best vertices of V'). Then, the following easy claim holds for a vertical separation of $V' \cup V''$ with parameter c.

Claim. Let $A(V')$ be a c-fraction of the best vertices in V' and $A(V'')$ the same in V''. Then $\delta(A(V')) + \delta(A(V'')) \geq c\delta(V' \cup V'')$.

Proof. Assume that in V' we have n' vertices. To form $A(V')$ we take the cn' vertices of V' with highest degree. The average degree of V' is $\delta(V')/n'$. The average degree of $A(V')$ is $\delta(A(V'))/(cn')$. But, from the selection of $A(V')$ as the cn' vertices with highest degree, we have that $\delta(A(V'))/(cn') \geq \delta(V')/n' \Rightarrow \delta(A(V')) \geq c\delta(V')$. Similarly for V'', i.e., $\delta(A(V'')) \geq c\delta(V'')$.

Solutions $\text{SOL}_5(B)$ and $\text{SOL}_6(B)$ are based upon vertical separations of $S_i \cup X_i$, $i = 1, 2$, with parameters π and λ, called π- and λ-vertical separations, respectively.

The idea behind vertical separation, is to handle the scenario when there is a "tiny" part of the solution (i.e. few in comparison to, let's say, k_1 vertices) that covers a large part of the solution and the "completion" of the solution done by the previous cases does not contribute more than a small fraction to the final solution. The vertical separation indeed tries to identify such a small part, and then continues the completion on the other side of the bipartition.

Solution $\text{SOL}_5(\mathbf{B})$. It consists of *separating $S_1 \cup X_1$ with parameter $\pi \in (0, 1/2]$, of taking a π-fraction of the best vertices of S_1 and a π-fraction of the best vertices of X_1 in the solution and of completing it with the adequate vertices from V_2.* A π-vertical separation of $S_1 \cup X_1$ introduces in the solution $\pi(2k_1 - k'_1) = \pi(2 - \nu)\mu k_2$ vertices of V_1, which are to be completed with $k - \pi(2 - \nu)\mu k_2 =$

$(1 + \mu)k_2 - \pi(2 - \nu)\mu k_2 = (1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices from V_2. Observe that such a separation implies the cuts with corresponding cardinalities B, C, F_i, $i = 1, 2, 3$, H_1, H_2, I_1, I_2, I_5, I_6, J_1, J_3, L_j, $j = 1, \ldots, 6$, N_1, P_2, P_4, U_1 and U_2. Let us group these cuts in the following way:

$$\begin{aligned}
\Pi_1 &= C + J_1 + J_3 + U_1 \\
\Pi_2 &= B + L_1 + L_4 + N_1 \\
\Pi_3 &= F_3 + L_2 + L_5 + P_4 \\
\Pi_4 &= I_1 + I_5 + F_1 + H_1 \\
\Pi_5 &= F_2 + L_3 + L_6 + P_2 \\
\Pi_6 &= I_2 + I_6 + H_2 + U_2
\end{aligned} \tag{16}$$

We may also notice that group Π_1 refers to $S_2 \setminus O_2$, Π_2 refers to $S_2 \cap O_2$, Π_3 to $X_2 \cap O_2$, Π_5 to $\bar{O}_2$ and Π_4 to $X_2 \setminus O_2$. Assume that a $\pi_i < 1$ fraction of each group Π_i, $i = 1, \ldots 6$ contributes in the π vertical separation of $S_1 \cup X_1$. Then, a π-vertical separation of $S_1 \cup X_1$ will contribute with a value:

$$\sum_{i=1}^{6} \pi_i \Pi_i \geqslant \pi \sum_{i=1}^{6} \Pi_i \tag{17}$$

to $\mathrm{sol}_5(B)$. We now distinguish two cases.

Case 1: $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2 \geqslant k_2$, i.e., $1 - \mu(2\pi - 1) + \mu\nu\pi \geqslant 1$. Then we further distinguish the following two subcases *1.* and *2.*:
1. $\mu(1 - 2\pi) + \mu\nu\pi \leq 1 - \xi$; then, the partial solution induced by the π-vertical separation will be completed in such a way that the contribution of the completion is at least equal to $\max\{Z_i, i = 1, \ldots, 5\}$, where:
Z_1 refers to S_2 plus the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2 - k_2 = (\mu(1 - 2\pi) + \mu\nu\pi)k_2$ vertices of O_2 having a contribution of:

$$\begin{aligned}
Z_1 = \sum_{i=1}^{2} (1 - \pi_i)\, \Pi_i + (J_2 + L_7 + N_2 + U_3) + \frac{\mu(1 - 2\pi) + \mu\nu\pi}{1 - \xi} [(1 - \pi_3)\, \Pi_3 \\
+ (1 - \pi_5)\, \Pi_5 + (L_8 + L_9 + P_3 + P_5)]
\end{aligned} \tag{18}$$

Z_2 refers to S_2 plus the best $(\mu(1 - 2\pi) + \mu\nu\pi)k_2$ vertices of X_2 having a contribution of:

$$\begin{aligned}
Z_2 = \sum_{i=1}^{2} (1 - \pi_i)\, \Pi_i + (J_2 + L_7 + N_2 + U_3) \\
+ \frac{\mu(1 - 2\pi) + \mu\nu\pi}{1 - \xi} \left[\sum_{j=3}^{4} (1 - \pi_i)\, \Pi_i + (I_3 + L_8 + P_1 + P_5)\right]
\end{aligned}$$

Z_3 and Z_4 refer to the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices of $S_2 \cup X_2$ and of $S_2 \cup O_2$ having, respectively, contributions:

$$Z_3 = \frac{1 - \mu(2\pi - 1) + \mu\nu\pi}{2 - \xi} \left[\sum_{i=1}^{4} (1 - \pi_i)\, \Pi_i \right.$$
$$+ (I_3 + J_2 + L_7 + L_8 + N_2 + P_1 + P_5 + U_3)]$$

$$Z_4 = \frac{1 - \mu(2\pi - 1) + \mu\nu\pi}{2 - \xi} \left[\sum_{i=1}^{3} (1 - \pi_i)\, \Pi_i + (1 - \pi_5)\, \Pi_5 \right.$$
$$+ (J_2 + L_7 + L_8 + L_9 + N_2 + P_3 + P_5 + U_3)]$$

Z_5 refers to the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices of $S_2 \cup X_2 \cup \bar{O}_2$ having a contribution of:

$$Z_5 = \frac{1 - \mu(2\pi - 1) + \mu\nu\pi}{3 - 2\xi} \left[\sum_{i=1}^{5} (1 - \pi_i)\, \Pi_i \right.$$
$$+ (I_3 + J_2 + L_7 + L_8 + L_9 + N_2 + P_1 + P_3 + P_5 + U_3)]$$

2. $\mu(1 - 2\pi) + \mu\nu\pi \geq 1 - \xi$; in this case, the partial solution induced by the π-vertical separation will be completed in such a way that the contribution of the completion is at least $\max\{\Theta_i, i = 1, \ldots, 3\}$, where:
Θ_1 refers to $S_2 \cup X_2$ plus the best $(\mu(1 - 2\pi) + \mu\nu\pi - (1 - \xi))k_2$ vertices of $\bar{O}_2$, all this having a contribution of:

$$\Theta_1 = \sum_{i=1}^{4} (1 - \pi_i)\, \Pi_i + (I_3 + J_2 + L_7 + L_8 + N_2 + P_1 + P_5 + U_3)$$
$$+ \frac{\mu(1 - 2\pi) + \mu\nu\pi - (1 - \xi)}{1 - \xi} [(1 - \pi_5)\, \Pi_5 + L_9 + P_3]$$

Θ_2 refers to $S_2 \cup O_2$ plus the best $(\mu(1 - 2\pi) + \mu\nu\pi - (1 - \xi))k_2$ vertices of $X_2 \setminus O_2$, all this having a contribution of:

$$\Theta_2 = \sum_{i=1}^{3} (1 - \pi_i)\, \Pi_i$$
$$+ (1 - \pi_5)\, \Pi_5 + (J_2 + L_7 + L_8 + L_9 + N_2 + P_3 + P_5 + U_3)$$
$$+ \frac{\mu(1 - 2\pi) + \mu\nu\pi - (1 - \xi)}{1 - \xi} [(1 - \pi_4)\, \Pi_4 + I_3 + P_1]$$

Θ_3 refers to the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices of $S_2 \cup X_2 \cup \bar{O}_2$ having a contribution of:

$$\Theta_3 = \frac{1 - \mu(2\pi - 1) + \mu\nu\pi}{3 - 2\xi} \left[\sum_{i=1}^{5} (1 - \pi_i)\, \Pi_i \right.$$
$$+ (I_3 + J_2 + L_7 + L_8 + L_9 + N_2 + P_1 + P_3 + P_5 + U_3)]$$

Case 2: $1 - \mu(2\pi - 1) + \mu\nu\pi < 1$. The partial solution induced by the π-vertical separation will be completed in such a way that the contribution of the completion is at least equal to $\max\{\Phi_i, i = 1, \ldots, 5\}$, where:

Φ_1 refers to the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices in S_2 with a contribution:

$$\Phi_1 = (1 - \mu(2\pi - 1) + \mu\nu\pi) \left[\sum_{i=1}^{2} (1 - \pi_i) \, \Pi_i + (J_2 + L_7 + N_2 + U_3) \right]$$

Φ_2 refers to the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices in X_2 with a contribution:

$$\Phi_2 = \frac{1 - \mu(2\pi - 1) + \mu\nu\pi}{1 - \xi} \left[\sum_{i=3}^{4} (1 - \pi_i) \, \Pi_i + (I_3 + L_8 + P_1 + P_5) \right]$$

Φ_3 refers to the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices in O_2 with a contribution:

$$\Phi_3 = (1 - \mu(2\pi - 1) + \mu\nu\pi) \left[\sum_{i=2}^{3} (1 - \pi_i) \, \Pi_i + (1 - \pi_5) \, \Pi_5 \right.$$
$$\left. + (L_7 + L_8 + L_9 + N_2 + P_3 + P_5) \right]$$

Φ_4 refers to the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices in $S_2 \cup X_2$ with a contribution:

$$\Phi_4 = \frac{1 - \mu(2\pi - 1) + \mu\nu\pi}{2 - \xi} \left[\sum_{j=1}^{4} (1 - \pi_j) \, \Pi_j \right.$$
$$\left. + (I_3 + J_2 + L_7 + L_8 + N_2 + P_1 + P_5 + U_3) \right]$$

Φ_5 refers to the best $(1 - \mu(2\pi - 1) + \mu\nu\pi)k_2$ vertices in $S_2 \cup X_2 \cup \bar{O}_2$ with a contribution:

$$\Phi_5 = \frac{1 - \mu(2\pi - 1) + \mu\nu\pi}{3 - 2\xi} \left[\sum_{j=1}^{5} (1 - \pi_j) \, \Pi_j \right.$$
$$\left. + (I_3 + J_2 + L_7 + L_8 + L_9 + N_2 + P_1 + P_3 + P_5 + U_3) \right] \tag{19}$$

Setting $Z^* = \max\{Z_i : i = 1, \ldots 5\}$, $\Theta^* = \max\{\Theta_i : i = 1, 2, 3\}$ and $\Phi^* = \max\{\Phi_i : i = 1, \ldots 5\}$, and putting (16) and (17) together with expressions (18) to (19), we get the following lower bound for ratio r_5:

$$\frac{\sum_{i=1}^{6} \pi_i \Pi_i + \begin{cases} \begin{cases} Z^* \text{ if } \mu(1 - 2\pi) + \mu\nu\pi \leq 1 - \xi \\ \Theta^* \text{ if } \mu(1 - 2\pi) + \mu\nu\pi \geq 1 - \xi \end{cases} & \text{case: } 1 - \mu(2\pi - 1) + \mu\nu\pi \geq 1 \\ \Phi^* & \text{case: } 1 - \mu(2\pi - 1) + \mu\nu\pi < 1 \end{cases}}{\mathrm{opt}(B)}$$

$$\tag{20}$$

Solution $\mathrm{SOL_6(B)}$. In a complete analogy with $\mathrm{SOL_5}$, solution $\mathrm{SOL_6}(B)$ consists of *separating $S_2 \cup X_2$ with parameter $\lambda \in (0, 1/2]$*. It consists of *separating*

$S_2 \cup X_2$ with parameter λ, of taking a λ fraction of the best vertices of S_2 and X_2 in the solution and of completing it with the adequate vertices from V_1. Here, we need that $\lambda(k_2 + k_2 - k_2') \leqslant k \Rightarrow \lambda(2 - \xi)k_2 \leqslant (1 + \mu)k_2 \Rightarrow \lambda \leqslant {}^{(1+\mu)}/_{(2-\xi)} \Rightarrow \lambda \in (0, {}^{(1+\mu)}/_{(2-\xi)}]$.

A λ-vertical separation of $S_2 \cup X_2$ introduces in the solution $\lambda(2-\xi)k_2$ vertices of V_2, which are to be completed with $k - \lambda(2 - \xi)k_2 = (1+\mu)k_2 - \lambda(2-\xi)k_2 = (1 + \mu - \lambda(2 - \xi))k_2$ vertices from V_1.

Observe that such a separation implies the cuts with corresponding cardinalities B, C, F_1, F_3, H_1, I_1, I_3, I_5, J_i, $i = 1, 2, 3$, L_1, L_2, L_4, L_5, L_7, L_8, N_1, N_2, P_1, P_4, P_5, U_1 and U_3. We group these cuts in the following way:

$$
\begin{aligned}
\Lambda_1 &= B + F_1 + F_3 + U_1 \\
\Lambda_2 &= C + H_1 + L_1 + L_2 \\
\Lambda_3 &= J_3 + I_5 + L_4 + L_5 \\
\Lambda_4 &= I_1 + J_1 + N_1 + P_4 \\
\Lambda_5 &= I_3 + J_2 + L_7 + L_8 \\
\Lambda_6 &= N_2 + P_1 + P_5 + U_3
\end{aligned}
\tag{21}
$$

Group Λ_1 refers to $S_1 \setminus O_1$, Λ_2 to $S_1 \cap O_1$, Λ_3 to $X_1 \cap O_1$, Λ_5 to $\bar{O}_1$ and Λ_4 to $X_1 \setminus O_1$. Assume, as previously, that a $\lambda_i < 1$ fraction of each group Λ_i, $i = 1, \dots 6$ contributes in the λ vertical separation of $S_2 \cup X_2$. Then, a λ-vertical separation of $S_2 \cup X_2$ will contribute with a value:

$$
\sum_{i=1}^{6} \lambda_i \Lambda_i \geqslant \lambda \sum_{i=1}^{6} \Lambda_i
\tag{22}
$$

to $\mathrm{sol}_6(B)$. We again distinguish two cases.

Case 1. $(1 + \mu - \lambda(2 - \xi))k_2 \geqslant \mu k_2$, i.e., $1 + \mu - \lambda(2 - \xi) \geqslant \mu$. Here we have the two following subcases.
(a) $1 - \lambda(2 - \xi) \leq (1 - \nu)\mu$; then, the partial solution induced by the λ-vertical separation will be completed in such a way that the contribution of the completion is at least equal to $\Upsilon^* = \max\{\Upsilon_i, i = 1, \dots, 5\}$, where: Υ_1 refers to S_1 plus the best $(1 - \lambda(2 - \xi))k_2$ vertices of X_1;
Υ_2 refers to S_1 plus the best $(1 - \lambda(2 - \xi))k_2$ vertices of O_1;
Υ_3 and Υ_4 refer to the best $(1 + \mu - \lambda(2 - \xi))k_2$ vertices of $S_1 \cup X_1$ and $S_1 \cup O_1$;
Υ_5 refers to the best $(1 + \mu - \lambda(2 - \xi))k_2$ vertices of $S_1 \cup X_1 \cup \bar{O}_1$. (b) $1 - \lambda(2 - \xi) \geq (1 - \nu)\mu$; in this case, the partial solution induced by the λ-vertical separation will be completed in such a way that the contribution of the completion is at least $\Psi^* = \max\{\Psi_i, i = 1, \dots, 3\}$, where:
Ψ_1 refers to $S_1 \cup X_1$ plus the best $(1 - \lambda(2 - \xi) - (1 - \nu))k_2$ vertices of $\bar{O}_1$;
Ψ_2 refers to $S_1 \cup O_1$ plus the best $(1 - \lambda(2 - \xi) - (1 - \nu))k_2$ vertices of $X_1 \setminus O_1$;
Ψ_3 refers to the best $(\mu + 1 - \lambda(2 - \xi))k_2$ vertices of $S_1 \cup X_1 \cup \bar{O}_1$.

Case 2. $1 + \mu - \lambda(2 - \xi) \leqslant \mu$. The partial solution induced by the λ-vertical separation will be completed in such a way that the contribution of the completion is at least equal to $\Omega^* = \max\{\Omega_i, i = 1, \dots, 5\}$, where:

Ω_1 refers to the best $(1 + \mu - \lambda(2 - \xi))k_2$ vertices in S_1;
Ω_2 refers to the best $(1 + \mu - \lambda(2 - \xi))k_2$ vertices in X_1;
Ω_3 refers to the best $(1 + \mu - \lambda(2 - \xi))k_2$ vertices in O_1;
Ω_4 refers to the best $(1 + \mu - \lambda(2 - \xi))k_2$ vertices in $S_1 \cup X_1$;
Ω_5 refers to the best $(1 + \mu - \lambda(2 - \xi))k_2$ vertices in $S_1 \cup X_1 \cup \bar{O}_1$.

Putting all this together we get:

$$r_6 \geqslant \frac{\sum_{i=1}^{6} \lambda_i \Lambda_i + \begin{cases} \begin{cases} \Upsilon^* \text{ if } 1 - \lambda(2 - \xi) \leq (1 - \nu)\mu \\ \Psi^* \text{ if } 1 - \lambda(2 - \xi) > (1 - \nu)\mu \end{cases} \text{ case: } \mu + 1 - \lambda(2 - \xi) \geq \mu \\ \Omega^* \qquad\qquad\qquad\qquad\qquad \text{ case: } \mu + 1 - \lambda(2 - \xi) < \mu \end{cases}}{\text{opt}(B)}$$

(23)

The complete study of solution $\text{SOL}_6(B)$ is deferred to [7].

4   Results and Discussion

To analyze the performance guarantee of k-VC_ALGORITHM, we set up a non-linear program and solved it to optimality. Here, we interpret the cardinalities of the edge-sets $B, C, F_i, \dots,$ as *variables*, the expressions in (8) as *constraints* and the *objective function* is $\min Z (\equiv \max_{j=1}^{6} r_j)$. In other words, we try to find a value assignments to the set of variables such that the maximum among all the six ratios defined is minimized. This value would give us the desired approximation guarantee of k-VC_ALGORITHM.

Towards this goal, we set up a GRG (Generalized Reduced Gradient [9]) program. The reasons this method is selected are presented in [7], as well as a more detailed description of the implementation. GRG is a generalization of the classical *Reduced Gradient* method [10] for solving (concave) quadratic problems so that it can handle higher degree polynomials and incorporate non-linear constraints.

As the values of parameters π and λ decrease, the approximation guarantee increases. The maximum of these ratios is attained for $\pi = \lambda = 10^{-5}$. For these values, the corresponding values of ratios $r_1 \div r_6$ computed for them are the following:

$$r_1 = 0.81806$$
$$r_2 = 0.81797$$
$$r_3 = 0.79280$$
$$r_4 = 0.79657$$
$$\mathbf{r_5 = 0.82104}$$
$$r_6 = 0.82103$$

These results correspond to the cycle that outputs the *minimum* value for the approximation factor and this is 0.821, given by solution SOL_5.

To conclude, let us note that the formulation of the non-linear program we developed for bounding the ratio below, could provide useful insights for

problem's understanding and could be applied for solving the problem on other graph-classes. Finally, since the overall algorithm chooses the best among a certain number of solutions it is easily parallelizable.

Remark. As we note in [7], the GRG solver does not guarantee the global optimal solution. The 0.821 guarantee is the minimum value that the solver returns after several runs from different initial starting points. However, successive re-executions of the algorithm, starting from this minimum value, were unable to find another point with smaller value. In each one of these successive re-runs, we tested the algorithm on 1000 random different starting points (which is greater than the estimation of the number of local minima) and the solver did not find value worse that the reported one.

References

1. Apollonio, N., Simeone, B.: The maximum vertex coverage problem on bipartite graphs. Discrete Appl. Math. **165**, 37–48 (2014)
2. Caskurlu, B., Mkrtchyan, V., Parekh, O., Subramani, K.: On partial vertex cover and budgeted maximum coverage problems in bipartite graphs. In: Diaz, J., Lanese, I., Sangiorgi, D. (eds.) TCS 2014. LNCS, vol. 8705, pp. 13–26. Springer, Heidelberg (2014)
3. Hochbaum, D.S., Pathria, A.: Analysis of the greedy approach in problems of maximum k-coverage. Naval Res. Logistics **45**, 615–627 (1998)
4. Badanidiyuru, A., Kleinberg, R., Lee, H.: Approximating low-dimensional coverage problems. In: Dey, T.K., Whitesides, S. (eds.) SoCG 2012, pp. 161–170. ACM, Chapel Hill (2012)
5. Ageev, A.A., Sviridenko, M.I.: Approximation algorithms for maximum coverage and max cut with given sizes of parts. In: Cornuéjols, G., Burkard, R.E., Woeginger, G.J. (eds.) IPCO 1999. LNCS, vol. 1610, p. 17. Springer, Heidelberg (1999)
6. Petrank, E.: The hardness of approximation: gap location. Comput. Complex. **4**, 133–157 (1994)
7. Bonnet, E., Escoffier, B., Paschos, V.T., Stamoulis, G.: A 0.821-ratio purely combinatorial algorithm for maximum k-vertex cover in bipartite graphs. CoRR arXiv:1409.6952v2 (2015)
8. Feige, U., Karpinski, M., Langberg, M.: Improved approximation of max-cut on graphs of bounded degree. J. Algorithms **43**, 201–219 (2002)
9. Abadie, J., Carpentier, J.: Generalization of the wolfe reduced gradient method to the case of non-linear constraints. In: Abadie, J., Carpentier, J. (eds.) Optimization. Academic Publishers (1969)
10. Frank, M., Wolfe, P.: An algorithm for quadratic programming. Naval Res. Logistics Q. **3**, 95–110 (1956)