New insight into introducing a $(2-\varepsilon)$-approximation ratio for minimum vertex cover problem

Majid Zohrehbandian

Department of Mathematics, Karaj Branch, Islamic Azad University, Karaj, Iran.

zohrebandian@yahoo.com

Abstract

Vertex cover problem is a famous combinatorial problem, which its complexity has been heavily studied. It is known that it is hard to approximate to within any constant factor better than 2, while a 2-approximation for it can be trivially obtained. In this paper, new properties and new techniques are introduced which lead to approximation ratios smaller than 2 on special graphs. Then, by introducing a modified graph and corresponding model along with satisfying the proposed assumptions, we propose new insight into solving this open problem and we introduce an approximation algorithm with a performance ratio of 1.999999 on arbitrary graphs.

Keywords: Discrete Optimization, Vertex Cover Problem, Complexity Theory, NP-Complete Problems.

MSC 2010: 90C35**,** 90C60.

1. Introduction

In complexity theory, the abbreviation NP refers to "nondeterministic polynomial", where a problem is in NP if we can quickly (in polynomial time) test whether a solution is correct. P and NP-complete problems are subsets of NP Problems. We can solve P problems in polynomial time while determining whether or not it is possible to solve NP-complete problems quickly (called the P vs NP problem) is one of the principal unsolved problems in Mathematics and Computer science.

Here, we consider the vertex cover problem which is a famous NP-complete problem. It cannot be approximated within a factor of 1.36 [1], unless $P = NP$, while a 2-approximation factor for it can be trivially obtained by taking all the vertices of a maximal matching in the graph. However, improving this simple 2-approximation algorithm has been a quite hard task [2,3].

In this paper, we introduce a $(2-\varepsilon)$-approximation ratio on special graphs, and then, we show that on arbitrary graphs a $(2-\varepsilon)$-approximation ratio can be obtained by solving a new semidefinite programming problem (SDP). The rest of the paper is structured as follows. Section 2 is about the vertex cover problem and introduces new properties and new techniques which lead to a $(2-\varepsilon)$-approximation ratio on special graphs. In section 3, we solve a new SDP model along with using the satisfying properties to propose an algorithm with a performance ratio smaller than 2 on arbitrary graphs. Finally, Section 4 concludes the paper.

2. Introducing a $(2-\varepsilon)$-approximation ratio on special graphs

In the mathematical discipline of graph theory, a vertex cover of a graph is a set of vertices such that each edge of the graph is incident to at least one vertex of the set. The problem of finding a minimum vertex cover is a typical example of an NP-complete optimization problem. In this section, new properties and new techniques are introduced which lead to approximation ratios smaller than 2 on special problems.

Let $G=(V,E)$ be an undirected graph on vertex set V and edge set E, where $|V|=n$. Throughout this paper, suppose that the vertex cover problem on G is hard and we have produced an arbitrary feasible solution for the problem, with vertex partitioning $V=V_{1G}\cup V_{-1G}$ (V_{1G} is a vertex cover of graph G) and objective value $|V_{1G}|$.

By defining the decision variables x_j and x_{ij} as follows:

$$x_j=\begin{cases}+1 & j\in V_{1G}^*\\ -1 & j\in V_{-1G}^*\end{cases}$$

$$x_{ij}=\begin{cases}+1 & i,j\in V_{1G}^* \quad or \quad i,j\in V_{-1G}^*\\ -1 & otherwise\end{cases}$$

And by considering the triangle inequalities, we can introduce the following integer linear programming (ILP) model for the minimum vertex cover problem:

$$(1) \qquad \min_{s.t.} \quad z^1=\sum_{1\le j\le n}\frac{1+x_j}{2}$$

$$\begin{array}{ll}
+x_i+x_j-x_{ij}=+1 & ij\in E, 1\le i<j\le n\\
+x_{ij}+x_{jk}+x_{ik}\ge -1 & 1\le i<j<k\le n\\
+x_{ij}-x_{jk}-x_{ik}\ge -1 & 1\le i<j<k\le n\\
-x_{ij}+x_{jk}-x_{ik}\ge -1 & 1\le i<j<k\le n\\
-x_{ij}-x_{jk}+x_{ik}\ge -1 & 1\le i<j<k\le n\\
x_j, x_{ij}\in\{-1,1\} & 1\le i<j\le n
\end{array}$$

Here, triangle inequalities are as cutting plane inequalities, and by consideration of x_j's as x_{oj} and addition of the constraint $x\succcurlyeq 0$, we have the well known SDP formulation as follows:

$$(2) \qquad \min_{s.t.} \quad z^2 = \sum_{i \in V} \frac{1 + v_o v_i}{2}$$

$$+v_o v_i + v_o v_j - v_i v_j = 1 \qquad ij \in E$$

$$+v_i v_j + v_i v_k + v_j v_k \geq -1 \qquad i, j, k \in V \cup \{o\}$$

$$+v_i v_j - v_i v_k - v_j v_k \geq -1 \qquad i, j, k \in V \cup \{o\}$$

$$-v_i v_j + v_i v_k - v_j v_k \geq -1 \qquad i, j, k \in V \cup \{o\}$$

$$-v_i v_j - v_i v_k + v_j v_k \geq -1 \qquad i, j, k \in V \cup \{o\}$$

$$v_i v_i = 1 \qquad i \in V \cup \{o\}$$

$$v_i v_j \in \{-1, +1\} \qquad i, j \in V \cup \{o\}$$

Theorem 1. Suppose that $z^{2*} \geq \frac{n}{2} + \frac{n}{k} = \frac{(k+2)n}{2k}$. Then, for all feasible solutions $V = V_{1G} \cup V_{-1G}$ we have the approximation ratio $\frac{|V_{1G}|}{Z^{2*}} \leq \frac{2k}{k+2}$.

Proof. $\frac{|V_{1G}|}{Z^{2*}} \leq \frac{n}{Z^{2*}} \leq \frac{2k}{k+2} < 2$ ■

Assumption 1. From now on, we assume that $\frac{n}{2} \leq z^{2*} < \frac{n}{2} + \frac{9n}{2000000}$; Otherwise for all feasible solutions $V = V_{1G} \cup V_{-1G}$ we have the approximation ratio $\frac{|V_{1G}|}{Z^{2*}} \leq \frac{2 \times \frac{2000000}{9}}{\frac{2000000}{9} + 2} < 1.99999 < 2$.

Theorem 2. Suppose that we have a suitable feasible solution $V_{1G} \cup V_{-1G}$ for which we have $|V_{1G}| \leq k|V_{-1G}|$. Then, we have the approximation ratio $\frac{|V_{1G}|}{Z^{2*}} \leq \frac{2k}{k+1} < 2$.

Proof. $\exists t \leq k$, for which we have $|V_{1G}| = t|V_{-1G}| = t\frac{n}{t+1}$. Then, $z^{2*} \geq \frac{n}{2} = \frac{t+1}{2t}|V_{1G}|$ which concludes that $\frac{|V_{1G}|}{Z^{2*}} \leq \frac{2t}{t+1} \leq \frac{2k}{k+1}$ ■

Therefore, for bounded values of k, we have some approximation ratios smaller than 2. But, if $k \to \infty$ then $\frac{|V_{1G}|}{Z^{2*}} \to 2$.

Corollary 1. If $|V_{1G}| < \frac{k}{k+1}n$ then $|V_{1G}| < k|V_{-1G}|$ and $\frac{|V_{1G}|}{Z^*} < \frac{2k}{k+1} < 2$.

Assumption 2. We don't have a suitable feasible solution $\mathrm{V} = \mathrm{V}_1 \cup \mathrm{V}_{-1}$ for which $|\mathrm{V}_1| < \frac{999999}{1000000}n$; Otherwise, for this feasible solution we have the approximation ratio $\frac{|V_1|}{Z^{2*}} \leq \frac{2 \times 999999}{999999+1} < 1.999999 < 2$.

Up to now, we could introduce $(2-\varepsilon)$-approximation ratio on special graphs with suitable characteristics. In section 3, we are going to introduce such a ratio on arbitrary graphs, where we assume that we have $V = V_1 \cup V_{-1}$ as a feasible solution of the vertex cover problem on arbitrary graph G for which $|V_1| \geq 0.999999n$ and $\frac{n}{2} \leq z^{2*} < \frac{n}{2} + \frac{9n}{2000000}$.

3. A (1.999999)-Approximation algorithm for vertex cover problem

In section 2, we could introduce a $(2-\varepsilon)$-approximation ratio on graphs without the proposed assumptions. Here, we are going to introduce a 1.999999-approximation ratio on arbitrary graphs. To do this, we assume the following assumption.

Assumption 3. By solving the SDP relaxation (2),

a) For less than $\frac{1}{1000000}n$ of vertices $j \in V$ and corresponding vectors we have $v_o^* v_j^* < 0$; Otherwise based on these vertices, we have a feasible solution with $|V_{-1}| \geq \frac{1}{1000000}n$, $|V_1| \leq \frac{999999}{1000000}n$ and approximation ratio $\frac{|V_{1G}|}{Z^{2*}} < \frac{2(999999)}{999999+1} < 1.999999 < 2$.

b) For less than $\frac{1}{100}n$ of vertices $j \in V$ and corresponding vectors we have $v_o^* v_j^* > 0.001$. Otherwise, $z^{2*} \geq \underbrace{\left(\frac{1+(-1)}{2} \times \frac{n}{1000000}\right)}_{v_o^* v_j^* < 0} + \underbrace{\left(\frac{1+0}{2} \times \frac{989999n}{1000000}\right)}_{0 \leq v_o^* v_j^* \leq 0.001} + \underbrace{\left(\frac{1+0.001}{2} \times \frac{n}{100}\right)}_{v_o^* v_j^* > 0.001} = \frac{n}{2} + \frac{9n}{2000000}$ and for all feasible solutions, we have the approximation ratio $\frac{|V_{1G}|}{Z^{2*}} < \frac{2(\frac{2000000}{9})}{\frac{2000000}{9}+2} < 1.99999 < 2$.

Based on Assumption (3), for more than $\frac{9}{10}\mathrm{n} < \frac{989999}{1000000}n$ of vertices j ∈ V and corresponding vectors we have $0 \leq v_o^* v_j^* \leq 0.001$. Let $\varepsilon = 0.001$ and suppose that we have a graph $\mathrm{G(V, E)}$ with a feasible vertex cover solution $\mathrm{V_1 \cup V_{-1}}$, where $|\mathrm{V_1}| \geq \frac{999999\mathrm{n}}{1000000}$ and $\frac{\mathrm{n}}{2} \leq \mathrm{z}^{2*} < \frac{\mathrm{n}}{2} + \frac{9\mathrm{n}}{200000}$. Moreover, suppose that $\mathrm{G}_\varepsilon = \{\mathrm{j} \in \mathrm{V} \mid 0 \leq v_o^* v_j^* \leq +\varepsilon\}$, where $|\mathrm{G}_\varepsilon| \geq 0.9\mathrm{n}$.

Theorem 3. For a vertex $k \in V$ and the corresponding set $H_k = \{j \in G_\varepsilon;\ |v_k^* v_j^*| > \varepsilon\}$, the subgraph on H_k is a bipartite graph.

Proof. Let us divide the vertex set H_k as follows:

$$S = \{j \in H_k \mid\ v_k^* v_j^* < -\varepsilon\} \quad \text{and} \quad T = \{j \in H_k \mid\ v_k^* v_j^* > +\varepsilon\}$$

Then, it is sufficient to show that the sets S and T are null subgraphs. Based on the first constraint of the SDP model (2), we have $v_i^* v_j^* \leq -1 + 2\varepsilon\ \ ij \in E(G),\ i,j \in H_k$. Therefore, if $i,j \in S$ then the second constraint of the SDP model (2) is violated and if $i,j \in T$ then the third constraint is violated ■

Corollary 2. If $\exists k \in V{:}\ |H_k| \geq \frac{n}{1000}$ then we have a feasible solution $V_{1G} \cup V_{-1G}$, correspondingly, where $|V_{-1G}| = max\{|S|,|T|\} \geq \frac{n}{2000}$. Hence, $|V_{1G}| \leq 1999|V_{-1G}|$ and $\frac{|V_{1G}|}{z^{2*}} \leq \frac{2\times1999}{1999+1} = 1.999 < 2$.

Assumption 4. $\forall k \in V: |H_k| < \frac{n}{1000}$ and we can't produce a suitable feasible solution. In this case, for each vector v_k^*, it is almost perpendicular to most of the vectors v_j^*. Moreover, we can display that each vertex $k \in V$ has a long-distance (on G_ε) to most of the vertices of G_ε and this is our reason to introduce the following graph.

Definition 1. For each pair of vertices i and j of graph $G = (V, E)$, add two new vertices i_j and j_i and a path with distance 3 through these vertices to produce the corresponding graph $H_G = (V \cup V', E \cup E')$, where $V' = \{i_j, j_i | i, j \in V\}$ and $E' = \{(i, i_j), (i_j, j_i), (j_i, j) | i, j \in V\}$, $|V'| = 2\binom{|V|}{2}$ and $|E'| = 3\binom{|V|}{2}$.

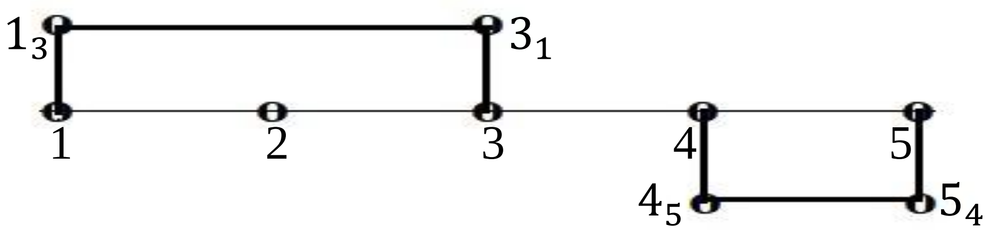

Figure 1. Addition of the new vertices to construct H_G.

Then, we introduce the following SDP model, which is almost similar to the SDP model (2). Here, the objective function is introduced so that for most of the pairs of vertices $i, j \in V$ we have $v_{i_j} v_{j_i} = -1$. In this manner, we will prove that for each vector v_k^* on G_ε, it is almost perpendicular only to a small number of the vectors v_j^* and as a result, we can use the Corollary (2).

$$(3) \qquad \min_{s.t.} \quad z^3 = \frac{|V|}{3} \sum_{i \in V} \frac{1 + v_o v_i}{2} + \sum_{\substack{i,j \in V \\ i<j}} \frac{-1 + v_{i_j} v_{j_i}}{2}$$

$$+v_o v_i + v_o v_j - v_i v_j = 1 \qquad ij \in E \cup E'$$

$$+v_i v_j + v_i v_k + v_j v_k \geq -1 \qquad i, j, k \in V \cup V' \cup \{o\}$$

$$+v_i v_j - v_i v_k - v_j v_k \geq -1 \qquad i, j, k \in V \cup V' \cup \{o\}$$

$$-v_i v_j + v_i v_k - v_j v_k \geq -1 \qquad i, j, k \in V \cup V' \cup \{o\}$$

$$-v_i v_j - v_i v_k + v_j v_k \geq -1 \qquad i, j, k \in V \cup V' \cup \{o\}$$

$$v_i v_i = 1 \qquad i \in V \cup V' \cup \{o\}$$

$$v_i v_j \in \{-1, +1\} \qquad i, j \in V \cup V' \cup \{o\}$$

Corollary 3. For feasible solutions $\bar{V}$ and $\hat{V}$ of vertex cover problem, where $|\bar{V}_1| = n$ and $|\hat{V}_1| \approx \frac{n}{2}$, we have $z^3(\hat{V}) \approx \frac{n^2}{6} - \left(\underbrace{\frac{n^2}{8}}_{i,j \in V_1^*} + \underbrace{\frac{n^2}{4}}_{i \in V_1^*; j \in V_o^*,} \right) = -\frac{10n^2}{48} < z^3(\bar{V}) \approx \frac{n^2}{3} - \frac{n^2}{2} = -\frac{n^2}{6}$. In other words, on G with the Assumption (3), after solving the SDP (3) relaxation we have $\sum_{i \in V} \frac{1 + v_o^* v_i^*}{2} \approx \frac{\mathrm{n}}{2}$.

Corollary 4. Let $n = 2$. By solving the SDP (3) relaxation on $V \cup V' \cup \{o\} = \{i, j, i_j, j_i, o\}$, where $i, j \in G_\varepsilon$ and almost perpendicular to each other, the second part of the objective function is almost equal to -0.75. To display this, we solve the following SDP relaxation on CVX Professional package (A system for disciplined convex programming, © 2005-2014 CVX Research, Inc., Austin, TX. http://cvxr.com) which is implemented in MATLAB. Here $i = 1, j = 2, i_j = 3, j_i = 4, o = 5$.

```
n = 5;
cvx_begin
    variable V( n, n)
    variable X( n, n ) symmetric

    for k = 1 : n,
        X(k,k) == 1;
    end
    0 <= X(5,1) <= 0.001;
    0 <= X(5,2) <= 0.001;
    -0.001 <= X(1,2) <= 0.001;
    X(5,1) + X(5,3) - X(1,3) == 1;
    X(5,2) + X(5,4) - X(2,4) == 1;
    X(5,3) + X(5,4) - X(3,4) == 1;
    for i = 1 : n,
        for j = i+1 : n,
            -1 <= X(i,j) <= 1;
            for k = j+1 : n,
                X(i,j) + X(i,k) + X(j,k) >= -1;
                X(i,j) - X(i,k) - X(j,k) >= -1;
               -X(i,j) + X(i,k) - X(j,k) >= -1;
               -X(i,j) - X(i,k) + X(j,k) >= -1;
            end
        end
    end
    X == semidefinite(n);
    minimize( (-1+X(3,4))/2 );
cvx_end
V = chol(X);

fprintf('Matrix X is:\n');
disp(X)
fprintf('Matrix V is:\n');
disp(V)
```

Theorem 4. In the optimal solution of the SDP (3) relaxation, there is not any vertex $i \in G_\varepsilon$ for which the corresponding vector is perpendicular to more than $\frac{3n}{4}$ of the other vertices of G_ε.

Proof. Based on the induction on n. It is true for $n = 2$. Suppose that it is true for $n - 1$ and we should prove it for a graph $G = (V, E)$ and its modification H_G, where $|V| = n$. Suppose that in the optimal solution of the SDP (3) relaxation we have a vertex $i \in G_\varepsilon$ for which the corresponding vector is perpendicular to more than $\frac{3n}{4}$ of the vertices of G_ε. By removing the vertex i and all vertices in V' which had been introduced based on the vertex i (without changing the objective coefficients), we have a feasible solution on $H_{G'} = H_{G_{V-\{i\}}}$, where

$$z^3_{relaxed}(G') \leq z^{3*}_{relaxed}(G) - \underbrace{0.5 \times \frac{n}{3}}_{removing\ i} + \underbrace{0.75 \times \frac{3n}{4}}_{\substack{removing\ i_k, k_i \in V' \\ v_i v_k \approx o}} + \underbrace{1 \times (\frac{n}{4} - 1)}_{\substack{removing\ i_k, k_i \in V' \\ |v_i v_k| > \varepsilon}} = z^{3*}_{relaxed}(G) + \frac{31n}{48} - 1.$$

But, by setting $v_i = -v_{i_k} = +v_{k_i} = +v_o^* \ \ k \in V - \{i\}$, we have a feasible solution on H_G with the objective value

$$z^3_{relaxed}(G) = z^3_{relaxed}(G') + \underbrace{1 \times \frac{n}{3}}_{adding\ i} - \underbrace{1 \times (n-1)}_{adding\ i_k, k_i \in V'} \leq z^{3*}_{relaxed}(G) + (\frac{31n}{48} - 1) + \left(-\frac{2n}{3} + 1\right)$$

or $z^3_{relaxed}(G) \leq z^{3*}_{relaxed}(G) - \frac{n}{48}$, which is a contradiction about the optimality of $z^{3*}_{relaxed}(G)$ ■

Corollary 5. If we have a feasible vertex cover with $|V_{1G}| \leq \frac{n}{2} + \frac{9n}{200000}$, then in the optimal solution of SDP (3) relaxation $\exists k \in G_\varepsilon$: $|H_k| \geq \frac{n}{4} - \underbrace{\frac{n}{10}}_{vertices\ i \in V - G_\varepsilon} = \frac{3n}{20}$. Hence, based on the optimal solution of SDP (3) relaxation we can produce a suitable feasible solution $V_{1G} \cup V_{-1G}$, correspondingly, where $|V_{-1G}| \geq \frac{3n}{40}$. Hence, $|V_{1G}| \leq \frac{37}{3}\ |V_{-1G}|$ and $\frac{|V_{1G}|}{z^*} \leq \frac{2 \times \frac{37}{3}}{\frac{37}{3} + 1} = 1.85 < 2$.

Now, we can introduce an algorithm to produce an approximation ratio $\rho \leq 1.999999$.

Zohrehbandian Algorithm (To produce a vertex cover solution with a factor $\rho \leq 1.999999$)

Step 1. Solve the SDP (3) relaxation.

Step 2. If for more than $\frac{n}{1000000}$ of vertices $j \in V$ and corresponding vectors we have $v_o^* v_j^* < 0$ then produce the suitable solution $V_{1G} \cup V_{-1G}$, correspondingly, where $V_{-1G} = \{j | v_o^* v_j^* < 0\}$. Therefore, based on the Assumption (3.a) we have $\frac{|V_{1G}|}{z^{2*}} \leq 1.999999$. Otherwise, go to Step 3.

Step 3. Based on the optimal solution of Step 1, produce H_k's. If $\exists k \in V$: $|H_k| \geq \frac{n}{1000}$ then produce the suitable solution $V_{1G} \cup V_{-1G}$, correspondingly, where $|V_{-1G}| = \max\{|S|, |T|\}$. Therefore, based on the Corollary (2) we have $\frac{|V_{1G}|}{z^{2*}} \leq 1.999$. Otherwise, go to Step 4.

Step 4. The optimal value of the vertex cover problem is greater than $\frac{|V|}{2} + \frac{9|V|}{2000000}$ and based on Assumption (1) for all feasible solutions, we have $\frac{|V_{1G}|}{Z^{2*}} \leq \frac{n}{Z^{*2}} \leq 1.99999$.

4. Conclusions

One of the open problems about the vertex cover problem is the possibility of introducing an approximation algorithm within any constant factor better than 2. Here, we proposed a new algorithm to introduce a 1.999999-approximation algorithm for the vertex cover problem on arbitrary graphs.

References

1. Dinur, I., & Safra, S. (2005). On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162, 439-485.

2. Khot, S., (2002, May). On the power of unique 2-Prover 1-Round games. Proceeding of 34th ACM Symposium on Theory of Computing, STOC.

3. Khot, S., & Regev, O. (2008). Vertex cover might be hard to approximate to within 2-Ɛ. *Journal of Computer and System Sciences*, 74, 335-349.