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:

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

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

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

\[
(1) \quad \min_{x} \quad z^1 = \sum_{1 \leq i < j \leq n} \frac{1 + x_{ij}}{2}

\text{s.t.} \\
+ x_i + x_j - x_{ij} = +1 \quad i, j \in E, 1 \leq i < j \leq n \\
+ x_{ij} + x_{jk} + x_{ik} \geq -1 \quad 1 \leq i < j < k \leq n \\
+ x_{ij} - x_{jk} - x_{ik} \geq -1 \quad 1 \leq i < j < k \leq n \\
- x_{ij} + x_{jk} - x_{ik} \geq -1 \quad 1 \leq i < j < k \leq n \\
- x_{ij} - x_{jk} + x_{ik} \geq -1 \quad 1 \leq i < j < k \leq n \\
x_j, x_{ij} \in \{-1,1\} \quad 1 \leq i < j \leq n
\]

Here, triangle inequalities are as cutting plane inequalities, and by consideration of \(x_j\)'s as \(x_{aj}\) and addition of the constraint \(x \geq 0\), we have the well known SDP formulation as follows:
Theorem 1. Suppose that $z^{*} \geq \frac{n}{2} + \frac{n}{k} = \frac{(k+2)n}{2k}$. Then, for all feasible solutions $V = V_{1G} \cup V_{-1G}$ we have the approximation ratio $\frac{|V_{1G}|}{z^{*}} \leq \frac{2k}{k+2}$.

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

Assumption 1. From now on, we assume that $\frac{n}{2} \leq z^{*} < \frac{n}{2} + \frac{9n}{2000000}$; Otherwise for all feasible solutions $V = V_{1G} \cup V_{-1G}$ we have the approximation ratio $\frac{|V_{1G}|}{z^{*}} \leq \frac{2\times2000000}{9} < 1.99999 < 2$.

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 $\frac{|V_{1G}|}{z^{*}} \leq \frac{2k}{k+1} < 2$.

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

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

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 $V = V_{1} \cup V_{-1}$ for which $|V_{1}| < \frac{99999}{100000} n$;

Otherwise, for this feasible solution we have the approximation ratio $\frac{|V_{1}|}{z^{*}} \leq \frac{2\times99999}{99999+1} < 1.99999 < 2$.

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 $|V_1| \geq 0.99999n$ 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 $(2 - \epsilon)$-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 $v_0^*v_j^* < 0$; 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 $v_0^*v_j^* > 0.001$. Otherwise, $z^{2*} \geq \frac{(1+(-1)}{2} \times \frac{n}{1000000} + \frac{(1+0)}{2} \times \frac{999999n}{1000000} + \frac{(1+0.001)}{2} \times \frac{n}{100}$ for all feasible solutions, we have the approximation ratio $\frac{|V_{1G}|}{z^{2*}} < \frac{2(999999)}{999999+1} < 1.999999 < 2$.

Based on Assumption (3), for more than $\frac{9}{10}n < \frac{999999}{1000000}n$ of vertices $j \in V$ and corresponding vectors we have $0 \leq v_0^*v_j^* \leq 0.001$. Let $\epsilon = 0.001$ and suppose that we have a graph $G(V, E)$ with a feasible vertex cover solution $V_1 \cup V_{-1}$, where $|V_1| \geq \frac{999999n}{1000000}$ and $\frac{n}{2} \leq z^{2*} < \frac{n}{2} + \frac{9n}{2000000}$. Moreover, suppose that $G_\epsilon = \{j \in V | 0 \leq v_0^*v_j^* \leq +\epsilon\}$, where $|G_\epsilon| \geq 0.9n$.

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

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

$S = \{j \in H_k | v_k^*v_j^* < -\epsilon\}$ and $T = \{j \in H_k | v_k^*v_j^* > +\epsilon\}$

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 $v_i^*v_j^* \leq -1 + 2\epsilon$ $ij \in E(G)$, $i, j \in H_k$. 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 \( G_e \)) to most of the vertices of \( G_e \) 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 \( H = (V \cup V', E \cup E') \), 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 \( H \).](image)

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 \( v_i^* v_j^* = -1 \). In this manner, we will prove that for each vector \( v_k^* \) on \( G_e \), it is almost perpendicular only to a small number of the vectors \( v_j^* \) and as a result, we can use the Corollary (2).

\[
\min_{s.t.} z^3 = \frac{|V|}{3} \sum_{i \in V} \frac{1 + v_i v_i}{2} + \sum_{\substack{i < j \in V \atop E'}} \frac{-1 + v_i v_j}{2} \\
+ v_o v_i + v_o v_j - v_i v_j = 1 \quad ij \in E \cup E' \\
+ v_i v_j + v_i v_k + v_j v_k \geq -1 \quad i, j, k \in V \cup V' \cup \{o\} \\
+ v_i v_j - v_i v_k - v_j v_k \geq -1 \quad i, j, k \in V \cup V' \cup \{o\} \\
- v_i v_j + v_i v_k - v_j v_k \geq -1 \quad i, j, k \in V \cup V' \cup \{o\} \\
- v_i v_j - v_i v_k + v_j v_k \geq -1 \quad i, j, k \in V \cup V' \cup \{o\} \\
v_i v_i = 1 \quad i \in V \cup V' \cup \{o\} \\
v_i v_j \in \{-1, +1\} \quad i, j \in V \cup V' \cup \{o\}
\]
Corollary 3. For feasible solutions \( \bar{V} \) and \( \hat{V} \) of vertex cover problem, where \( |V_1| = n \) and \( |V_1| \approx \frac{n}{2} \), we have \( z^3(\bar{V}) \approx \frac{n^2}{6} - \left( \frac{n^2}{8} \sum_{i,j \in V_1^*} \right) + \left( \frac{n^2}{4} \sum_{i \in V_1^*, j \in V_0^*} \right) = -\frac{10n^2}{48} \). In other words, on \( G \) with the Assumption (3), after solving the SDP (3) relaxation we have \( \sum_{i \in V} \frac{1 + v_i^* v_i^T}{2} \approx \frac{n}{2} \).

Corollary 4. Let \( n = 2 \). By solving the SDP (3) relaxation on \( V \cup V' \cup \{o\} = \{i, j, i_0, j_0, o\} \), where \( i, j \in G_e \) 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 \( i = 1, j = 2, i_0 = 3, j_0 = 4, o = 5 \).

```matlab
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(5,1) <= 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:
');
disp(X)
fprintf('Matrix V is:
');
disp(V)
```
Theorem 4. In the optimal solution of the SDP (3) relaxation, there is not any vertex \( i \in G_e \) for which the corresponding vector is perpendicular to more than \( \frac{3n}{4} \) of the other vertices of \( G_e \).

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 \( i \in G_e \) for which the corresponding vector is perpendicular to more than \( \frac{3n}{4} \) of the other vertices of \( G_e \). 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 \( H_{G_i} = H_{G^i-V} \), where

\[
3^3 \left( G' \right) \leq 3^3 \left( G \right) - 0.5 \times \frac{n}{3} + 0.75 \times \frac{3n}{4} + 1 \times \frac{n - 1}{4} \left( \frac{3n}{4} - 1 \right) = 3^3 \left( G \right) + \frac{n}{48} - 1.
\]

But, by setting \( v_i = -v_{i_k} = +v_{k_i} = +v_0 \ \text{k \in V - \{i\}} \), we have a feasible solution on \( H_G \) with the objective value

\[
3^3 \left( G \right) = 3^3 \left( G' \right) + \frac{n}{3} - 1 \times \frac{n - 1}{4} \left( \frac{3n}{4} - 1 \right) = 3^3 \left( G \right) + \frac{3n - 1}{48} + \frac{2n}{3} + 1
\]

or \( 3^3 \left( G \right) \leq 3^3 \left( G \right) - \frac{n}{48} \), which is a contradiction about the optimality of \( 3^3 \left( G \right) \).

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 \( \exists k \in G_e \colon |H_k| \geq \frac{n}{4} - \frac{n}{1000000} = \frac{3n}{20} \). Hence, based on the optimal solution of SDP (3) relaxation we can produce a suitable feasible solution \( V_{1G} \cup V_{-1G} \), correspondingly, where \( |V_{-1G}| \geq \frac{n}{40} \). Hence, \( |V_{1G}| \leq \frac{27}{3} |V_{-1G}| \) and \( \frac{|V_{1G}|}{z^*} \leq \frac{27}{3} + 1 = 1.85 < 2 \).

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 \( v_i^* v_j^* < 0 \) then produce the suitable solution \( V_{1G} \cup V_{-1G} \), correspondingly, where \( V_{-1G} = \{ |v_i^* v_j^* < 0 \} \). Therefore, based on the Assumption (3.a) we have \( \frac{|V_{1G}|}{z^*} \leq 1.999999 \). Otherwise, go to Step 3.

Step 3. Based on the optimal solution of Step 1, produce \( H_k \) s. If \( \exists k \in V \colon |H_k| \geq \frac{n}{1000} \) 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 \( \frac{|V_{1G}|}{z^*} \leq 1.999999 \). 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_{16}|}{Z^2} \leq \frac{n}{Z^2} \leq 1.999999.

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

