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 over the years and we know that there is not any mathematical formulation that approximates it better than $2 - o(1)$. In other words, 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, by a combination of a well-known semidefinite programming formulation and a rounding procedure, along with satisfying new properties, we introduce an approximation algorithm for the vertex cover problem with a performance ratio of 1.999999 on arbitrary graphs, en route answering an open question about the unique games conjecture.

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 show that there is a $(2 - \mathcal{E})$-approximation ratio for the vertex cover problem, based on any feasible solution of it. Then, we fix the $\mathcal{E}$ value equal to $\mathcal{E} = 0.000001$ and we show that on arbitrary graphs a 1.999999-approximation ratio can be obtained by a combination of a well-known semidefinite programming (SDP) formulation and a rounding procedure.

The rest of the paper is structured as follows. Section 2 is about the vertex cover problem and introduces new properties about it. In section 3, we propose a rounding procedure along with using the satisfying properties to propose an algorithm with a performance ratio of 1.999999 on arbitrary graphs. Finally, Section 4 concludes the paper.

2. Performance ratio for a feasible solution of vertex cover problem

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, we calculate the performance ratios of vertex cover feasible solutions which lead to an approximation ratio of $2 - \mathcal{E}$, where the value of $\mathcal{E}$ is not constant and depends on the produced feasible solution. Then, in the next section, we fix the value of $\mathcal{E}$ equal to $\mathcal{E} = 0.000001$ to produce a 1.999999-approximation ratio for the vertex cover problem on arbitrary graphs.

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 solution of the graph G) and objective value $|V_{1G}|$, and for solving the problem, we use the well-known semidefinite programming (SDP) formulation as follows:

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

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

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

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

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

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

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

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

Note that, we know for sure that just by solving this SDP formulation or the other SDP formulations with additional constraints, we cannot approximate the vertex cover problem with a performance ratio better than $2 - o(1)$. In other words, in section 3, we are going to propose a randomized algorithm to classify the solution vectors of the SDP (1) relaxation to produce a suitable solution for the vertex cover problem with a performance ratio of 1.999999.

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

Proof. $z^* \geq \frac{(k+2)n}{2k}$ and $\frac{n}{Z^*} \leq \frac{2k}{k+2}$. Then, $\frac{|V_{1G}|}{Z^*} \leq \frac{n}{Z^*} \leq \frac{2k}{k+2} < 2$ ■

Theorem 2. Suppose that the vertex cover problem on G is hard ($z^* \geq \frac{n}{2}$) and we have produced a feasible solution $V_{1G} \cup V_{-1G}$ for which we have $|V_{1G}| \leq k|V_{-1G}|$. Then, we have an approximation ratio $\frac{|V_{1G}|}{Z^*} \leq \frac{2k}{k+1} < 2$.

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

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

In section 2 and based on the produced feasible solution, we could introduce a $(2 - \varepsilon)$-approximation ratio where Ɛ value was not a constant value. In this section, we fix the value of Ɛ equal to Ɛ = 0.000001 to produce a 1.999999-approximation ratio for the vertex cover problem on arbitrary graphs. To do this, we assume the following assumption about the solution of the SDP (1) relaxation.

Assumption 1. By solving the SDP (1) relaxation,

a) For less than $\frac{1}{1000000}n$ of vertices $j \in V$ and corresponding vectors we have $v_o^* v_j^* < 0$.
   Otherwise, we can produce $V_{-1G} = \{j \in V \mid v_o^* v_j^* < 0\}$ and $V_{1G} = V - V_{-1G}$, to have a feasible solution with $|V_{-1G}| \geq \frac{1}{1000000}n$ and $|V_{1G}| \leq \frac{999999}{1000000}n \leq 999999|V_{-1G}|$. Then, based on Theorem (2) we have an approximation ratio $\frac{|V_1G|}{Z^*} < \frac{2(999999)}{999999+1} = 1.999998 < 2$.

b) For less than $\frac{1}{100}n$ of vertices $j \in V$ and corresponding vectors we have $v_o^* v_j^* > 0.0004$.
   Otherwise, $z^* \geq \underbrace{\left(\frac{1+(-1)}{2} \times \frac{n}{1000000}\right)}_{v_o^* v_j^* < 0} + \underbrace{\left(\frac{1+0}{2} \times \frac{989999n}{1000000}\right)}_{0 \leq v_o^* v_j^* \leq 0.0004} + \underbrace{\left(\frac{1+0.0004}{2} \times \frac{n}{100}\right)}_{v_o^* v_j^* > 0.0004} = \frac{n}{2} + 0.0000015n.$

Then, based on Theorem (1) and for all feasible solutions $V = V_{1G} \cup V_{-1G}$, we have the approximation ratio $\frac{|V_{1G}|}{Z^*} \leq \frac{2(\frac{1}{0.0000015})}{\frac{1}{0.0000015}+2} \leq 1.999994 < 2$,

Definition 1. Let $\varepsilon = 0.0004$ and $G_\varepsilon = \{j \in V \big| 0 \leq v_o^* v_j^* \leq +\varepsilon\}$.

Based on Assumption (1), after solving the SDP (1) relaxation, we have a performance ratio of $max\{1.999994, 1.999998\} < 1.999999$ (if the solution of the SDP (1) relaxation does not meet the Assumption (1)) or for more than $\frac{989999}{1000000}n$ of vertices $j \in V$ and corresponding vectors we have $0 \leq v_o^* v_j^* \leq +\varepsilon$; i.e. $|G_\varepsilon| \geq 0.989999n$ (if the solution of the SDP (1) relaxation meets the Assumption (1)).

Note that, the induced subgraph on G_ε is a triangle-free graph and we know that almost all triangle-free graphs are bipartite. However, if the solution of the SDP (1) relaxation meets the Assumption (1), it is necessary to introduce a suitable feasible solution based on G_ε to produce a performance ratio of 1.999999.

Theorem 3. For any normalized vector w, the induced subgraph on $H_w = \{j \in G_\varepsilon;\ |wv_j^*| > 0.5003\}$ is a bipartite graph.

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

$$S = \{j \in H_w |\ wv_j^* < -0.5003\} \quad \text{and} \quad T = \{j \in H_w |\ wv_j^* > +0.5003\}$$

Then, it is sufficient to show that the sets S and T are null subgraphs. For each edge $ij \in E(G)$ and based on the first constraint of the SDP model (1), if $i, j \in H_w \subseteq G_\varepsilon$ then we have $v_i^* v_j^* \leq -1 + 2\varepsilon$.

$$+ \underbrace{v_o v_i}_{0 \leq v_o^* v_i^* \leq +\varepsilon} + \underbrace{v_o v_j}_{0 \leq v_o^* v_j^* \leq +\varepsilon} - v_i v_j = 1 \qquad ij \in E, \quad i, j \in H_w \subseteq G_\varepsilon$$

Moreover, the triangle inequality could not be violated between vectors $v_i^* - v_j^*,\ w - v_j^*$ and $w - v_i^*$. But, if $ij \in E(T)$ then the triangle inequality between these vectors is violated; i.e.

$$\|v_i^* - v_j^*\| \leq \|w - v_i^*\| + \|w - v_j^*\|$$

$$\sqrt{2 - 2v_i^* v_j^*} \leq \sqrt{2 - 2wv_i^*} + \sqrt{2 - 2wv_j^*}$$

$$\sqrt{2 - 2(-1 + 2(0.0004))} \leq \sqrt{2 - 2v_i^* v_j^*} \leq \sqrt{2 - 2wv_i^*} + \sqrt{2 - 2wv_j^*} \leq 2\sqrt{2 - 2(0.5003)}$$

Therefore, we have $1.9995999 \leq \sqrt{3.9984} \leq 2\sqrt{0.9994} \leq 1.9994$, which is a contradiction.

Likewise, if $ij \in E(S)$ then the triangle inequality between vectors $v_i^* - v_j^*,\ u - v_j^*$ and $u - v_i^*$ is violated, where $u = -w$ ∎

Corollary 1. By introducing a normalized random vector w, where $|H_w| \geq \frac{n}{1000000}$, we can produce a feasible solution $V_{1G} \cup V_{-1G}$, correspondingly, where $|V_{-1G}| = max\{|S|, |T|\} \geq \frac{n}{2000000}$. Hence, based on Theorem (2), we have $|V_{1G}| \leq \frac{1999999n}{2000000} \leq 1999999|V_{-1G}|$ and $\frac{|V_{1G}|}{z^*} \leq \frac{2\times1999999}{1999999+1} = 1.999999 < 2$.

In other words, to produce a performance ratio of 1.999999, we should solve the SDP (1) relaxation. Then, if the solution of the SDP (1) relaxation does not meet the Assumption (1), we have a performance ratio of 1.999999 and if the solution of the SDP (1) relaxation meets the Assumption (1), it is sufficient to produce a normalized random vector w, where $|H_w| \geq \frac{n}{1000000}$.

Theorem 4. Let u, w be two normalized random vectors, then for any normalized vector v_j^*, we have $\Pr\left(|uv_j^*| \leq 0.5003 \,\&\, |wv_j^*| \leq 0.5003\right) < 0.60933$.

Proof. Let $v_j^* = v_j' + v_j''$, where v_j' is the projection of vector v_j^* onto the $u - w$ plane (suppose that the vector u is on the ox axis) and v_j'' is the projection of v_j^* onto the normal vector of that plane. Then, $|uv_j^*| = |u(v_j' + v_j'')| = |uv_j'|$ and we have $|uv_j'| \leq 0.5003$ if and only if the vector v_j' is projected on the gray region in Figure (1). In other words (suppose that the vector v_j' is projected onto the first quadrant and with an angle of $0 \leq \theta \leq \frac{\pi}{2}$ concerning the vector u), we have:

If $|v_j'| > f(\theta) = \frac{0.5003}{\cos\theta}$ (white region) then $|uv_j^*| = |uv_j'| = |v_j'||\cos\theta| > 0.5003$,

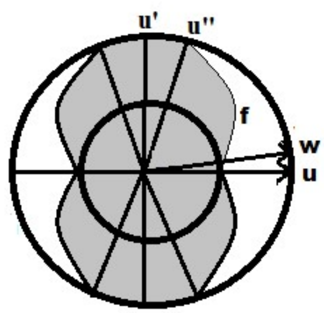

Figure 1. $u - w$ plane, where the radius of the smaller circle is 0.5003,
$uu' = 0$, $uu'' = 0.5003$, $\widehat{uou''} = \cos^{-1}(0.5003)$, $f(\theta) = \frac{0.5003}{\cos\theta}$ $(0 \leq \theta \leq \cos^{-1}(0.5003))$,
and the gray region is symmetric concerning the ox axis, the oy axis, and the origin.

Note that, the maximum area of the region S, where S is the area of the common gray region between two vectors u and w, is produced based on the $|uw| \cong 1$ condition (and the minimum value is produced when $|uw| \cong 0$).

Therefore, $\Pr\left(|uv_j^*| \leq 0.5003 \,\&\, |wv_j^*| \leq 0.5003\right) = \frac{S}{\pi}$, where

$$S \leq 4\left(\int_0^{\cos^{-1}(0.5003)} \frac{1}{2}\,(\frac{0.5003}{\cos\theta})^2 d\theta + \int_{\cos^{-1}(0.5003)}^{\frac{\pi}{2}} \frac{1}{2} d\theta\right)$$

$$= 2\left((0.5003)^2 \tan\left(\cos^{-1}(0.5003)\right) + \left(\frac{\pi}{2} - \cos^{-1}(0.5003)\right)\right) < 2(0.957132).$$

Hence, $\Pr\left(|uv_j^*| \leq 0.5003 \;\&\; |wv_j^*| \leq 0.5003\right) < \frac{2(0.957132)}{\pi} < 0.60933$ ∎

Note that, if $|uw| < 1$ then we have smaller values for this probability. Therefore, by introducing two normalized random vectors u, w, we have $|uv_j^*| \leq 0.5003$ and $|wv_j^*| \leq 0.5003$ for at most $0.60933n$ of the vectors v_j^* , the optimal solution of the SDP (1) relaxation. Hence, for at least $0.39067n$ of the vectors v_j^* we have $|uv_j^*| > 0.5003$ or $|wv_j^*| > 0.5003$.

Corollary 2. If the solution of the SDP (1) relaxation meets the Assumption (1), then we have $|\mathrm{G}_\varepsilon| \geq 0.989999\mathrm{n}$ and one of the bipartite graphs H_u or H_w has more than $\frac{0.39067|\mathrm{G}_\varepsilon|}{2} = \frac{0.39067\times 0.989999\mathrm{n}}{2} > \frac{n}{1000000}$ of vertices which produces a null subgraph V_{-1G} with more than $\frac{0.39067\times 0.98999}{4} > \frac{n}{2000000}$ of the vertices and based on the Corollary (1) we have an approximation ratio $\frac{|V_{1G}|}{z^{2*}} \leq 1.999999 < 2$.

Now, we can introduce our 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 (1) 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 $\mathrm{V}_{1\mathrm{G}} \cup \mathrm{V}_{-1\mathrm{G}}$, correspondingly, where $\mathrm{V}_{-1\mathrm{G}} = \{\mathrm{j} | v_o^* v_j^* < 0\}$. Therefore, based on the Assumption (1. a) we have $\frac{|\mathrm{V}_{1\mathrm{G}}|}{\mathrm{z}^*} \leq 1.999999$. Otherwise, go to Step 3.

Step 3. If for more than $\frac{1}{100}n$ of vertices $j \in V$ and corresponding vectors we have $v_o^* v_j^* > 0.0004$, then $z^* \geq \frac{n}{2} + 0.0000015n$ and it is sufficient to produce an arbitrary feasible solution. Therefore, based on Assumption (1. b) for all feasible solutions $V = V_{1G} \cup V_{-1G}$ we have $\frac{|\mathrm{V}_{1\mathrm{G}}|}{\mathrm{z}^*} \leq 1.999999$. Otherwise, go to Step 4.

Step 4. We have $|\mathrm{G}_{0.0004}| \geq 0.989999\mathrm{n}$. Then, introduce two normalized random vectors u and w, and produce H_u and H_w. One of these bipartite graphs has more than $\frac{n}{1000000}$ vertices and based on it, we can produce a suitable solution $V_{1G} \cup V_{-1G}$, correspondingly, where $|V_{-1G}| \geq \frac{n}{2000000}$. Therefore, based on the Corollary (2) we have $\frac{|V_{1G}|}{z^*} \leq 1.999999$.

Corollary 3. Based on the proposed 1.999999-approximation algorithm for the vertex cover problem, the unique games conjecture is not true.

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, and this may lead to the conclusion that $P = NP$.

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). 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.