New Insight Into Introducing a (2-Ɛ)-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 it is known that **there is not any mathematical programming formulation that approximates it better than $2 - o(1)$**, 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 randomized 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 to 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 (VCP) 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 $(2 - \varepsilon)$-approximation ratio for the vertex cover problem, based on any VCP feasible solution, **where the value of ε is not constant and depends on** the produced feasible solution. **Then, we fix the ε value equal to $\varepsilon = \mathbf{0.000001}$** 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 randomized 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 randomized 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 based on a VCP feasible solution

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 VCP feasible solutions which lead to an approximation ratio of $2 - \varepsilon$, where the value of ε is not constant and depends on the produced feasible solution. Then, in the next section, we fix the value of ε equal to $\varepsilon = 0.000001$ 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 $V = V_{1G} \cup V_{-1G}$ and objective value $|V_{1G}|$, and for solving the problem, we use the relaxation of the well-known semidefinite programming (SDP) formulation as follows:

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

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

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

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

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

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

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

Note that, we know for sure that just by solving this SDP relaxation 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. Although it is hard to exactly solve the SDP formulation (1), let's assume that we know $z^* \geq \frac{n}{2}+\frac{n}{k}=\frac{(k+2)n}{2k}$. Then, for all vertex cover feasible partitioning $V=V_{1G}\cup V_{-1G}$, we have the approximation ratio $\frac{|V_{1G}|}{Z^*}\leq\frac{2k}{k+2}<2$.

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

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

Proof. If $|V_{1G}|\leq\frac{kn}{k+1}$ then $n\geq\frac{k+1}{k}|V_{1G}|$. Hence, $z^*\geq\frac{n}{2}\geq\frac{k+1}{2k}|V_{1G}|$ which concludes that $\frac{|V_{1G}|}{Z^*}\leq\frac{2k}{k+1}<2$ ■

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

In section 2 and based on a 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 $V_{-1G} = \{j \in V | v_o^* v_j^* < 0\}$ 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 $\frac{|V_1 G|}{Z^*} < \frac{2(999999)}{999999+1} = 1.999998 < 2$.

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, $z^* \geq \underbrace{\left(\frac{1+(-1)}{2} \times \frac{n}{1000000}\right)}_{v_o^* v_j^* < 0} + \underbrace{\left(\frac{1+0}{2} \times \frac{989999n}{1000000}\right)}_{0 \leq v_o^* v_j^* \leq 0.0004} + \underbrace{\left(\frac{1+0.0004}{2} \times \frac{n}{100}\right)}_{v_o^* v_j^* > 0.0004} = \frac{n}{2} + 0.0000015n.$

Note that, the third summation is the minimum value on the vertices $j \in V$ with $v_o^* v_j^* > 0.0004$, where against Assumption (1.b) we have more than $\frac{1}{100}n$ of vertices $j \in V$ with $v_o^* v_j^* > 0.0004$. Moreover, due to Assumption (1.a) the first summation is the minimum value on the vertices $j \in V$ with $v_o^* v_j^* < 0$. Therefore, based on Theorem (1) and for all VCP 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 $G_\varepsilon = \{j \in V \mid 0 \leq v_o^* v_j^* \leq +\varepsilon\}$.

Based on Assumption (1), after solving the SDP (1) relaxation,

- If the solution of the SDP (1) relaxation does not meet the Assumption (1) then we have a performance ratio of $max\{1.999994, 1.999998\} = 1.999998 < 1.999999$,
- Otherwise (if the solution of the SDP (1) relaxation meets the Assumption (1)), for more than $\frac{989999}{1000000}n$ of vertices j ∈ V, we have $0 \leq v_o^* v_j^* \leq +\varepsilon$; i.e. $|G_\varepsilon| \geq 0.989999n$.

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, to produce a performance ratio of 1.999999, it is necessary to introduce a suitable feasible solution based on G_ε.

Theorem 3. By solving the SDP (1) relaxation and for any vector v_k^*, the induced subgraph on $F_k = \{j \in G_\varepsilon;\ |v_k^* v_j^*| > \varepsilon = 0.0004\}$ is a bipartite graph.

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

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

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 $i, j \in F_k \subseteq G_\varepsilon$ then we have $v_i^* v_j^* \leq -1 + 2\varepsilon$.

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

Now, if $ij \in E(S)$ then the second constraint of the SDP model (1) is violated; i.e. We have:

$$\underbrace{\underbrace{+v_i v_j}_{v_i^* v_j^* \leq -1+2\varepsilon} \quad \underbrace{+v_i v_k}_{v_k^* v_i^* < -\varepsilon} \quad \underbrace{+v_j v_k}_{v_k^* v_j^* < -\varepsilon}}_{< -1} \geq -1$$

Likewise, if $ij \in E(T)$ then the third constraint of the SDP model (1) is violated; i.e. We have:

$$\underbrace{\underbrace{+v_i v_j}_{v_i^* v_j^* \leq -1+2\varepsilon} \quad \underbrace{-v_i v_k}_{-v_k^* v_i^* < -\varepsilon} \quad \underbrace{-v_j v_k}_{-v_k^* v_j^* < -\varepsilon}}_{< -1} \geq -1$$

■

Corollary 1. By solving the SDP (1) relaxation and for any vector v_k^*, if $|F_k| \geq \frac{n}{1000000}$, 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 $\frac{|V_{1G}|}{z^*} \leq \frac{2 \times 1999999}{1999999+1} = 1.999999 < 2$; i.e. We can produce a performance ratio of 1.999999

Assumption 2. By solving the SDP (1) relaxation, for all vector v_k^*, we have $|F_k| < \frac{n}{1000000}$; i.e. For each vector v_k^*, it is almost orthogonal to most (almost all) of the vectors in G_ε .

Theorem 4. 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, which is a contradiction; i.e. The triangle inequality satisfies and we have:

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

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

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

But, this means that $1.9995 \leq \sqrt{3.9984} \leq 2\sqrt{0.9994} \leq 1.9994$, 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 2. 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 Assumptions (1,2) then we have a performance ratio of 1.999999. Otherwise (the solution meets Assumptions (1,2)), it is sufficient to produce a normalized random vector w, where $|H_w| \geq \frac{n}{1000000}$.

Therefore, we should prove that the probability of introducing such a random vector w is acceptable; e.g. Its probability is more than 0.5.

Theorem 5. Let w be a normalized random vector, then for any optimal vector v_j^* $(j \in G_\varepsilon)$, we have $\Pr\left(\left|wv_j^*\right| \leq 0.5003\right) < 0.60933$.

Proof. Let $w = w' + w"$, where w' is the orthogonal projection of vector w onto the $v_o^*ov_j^*$ plane and $w"$ is the projection of w onto the normal vector of that plane (suppose that the vector v_j^* is on the ox axis and then the vector v_o^* is nearly on the oy axis). Hence, $\left|wv_j^*\right| = \left|(w'+w")v_j^*\right| = \left|w'v_j^*\right|$.

Moreover, suppose that the vector w' lies in the first quadrant and with an angle of $0 \leq \theta \leq \frac{\pi}{2}$ to the vector v_j^*. Then, we have $\left|w'v_j^*\right| > 0.5003$ if and only if the vector w' lies in the white region and above the function $f(\theta)$, where $f(\theta) = \begin{cases} \frac{0.5003}{\cos\theta} & 0 \leq \theta \leq \cos^{-1}(0.5003) \\ 1 & \cos^{-1}(0.5003) \leq \theta \leq \frac{\pi}{2} \end{cases}$; See Figure 1.

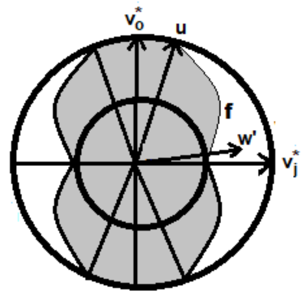

Figure 1. The $v_o^* O v_j^*$ plane, where the radius of the smaller circle is 0.5003, $uv_j^* = 0.5003$, $\widehat{uOv_j^*} = \cos^{-1}(0.5003)$, $f(\theta) = \frac{0.5003}{\cos\theta}$ $(0 \le \theta \le \cos^{-1}(0.5003))$, and the gray region is symmetric concerning the ox axis, the oy axis, and the origin.

In other words, if w' lies in the white region then we have $1 \ge |w'| > f(\theta)$. Hence, $|w'| > \frac{0.5003}{\cos\theta}$ and $|wv_j^*| = |w'v_j^*| = |w'||v_j^*||\cos\theta| = |w'||\cos\theta| > 0.5003.$

Therefore, $\Pr\left(|wv_j^*| \le 0.5003\right) = \frac{S}{\frac{\pi}{4}}$, where S is the area of the gray region in the first quadrant. Moreover, $S = \int_0^{\cos^{-1}(0.5003)} \frac{1}{2} \left(\frac{0.5003}{\cos\theta}\right)^2 d\theta + \int_{\cos^{-1}(0.5003)}^{\frac{\pi}{2}} \frac{1}{2} d\theta.$

$$S = 0.5\left((0.5003)^2 \tan\left(\cos^{-1}(0.5003)\right) + \left(\frac{\pi}{2} - \cos^{-1}(0.5003)\right)\right) < 0.478567 .$$

And we have $\Pr\left(|wv_j^*| \le 0.5003\right) < \frac{4(0.478567)}{\pi} < 0.60933$ ■

Therefore, if the solution of the SDP (1) relaxation meets Assumptions (1,2) then by introducing a normalized random vector w, the probability of the event $|wv_j^*| > 0.5003$ at each observation $j \in G_\varepsilon$ is at least 0.39067; i.e. we have a binomial distribution for $|G_\varepsilon| \ge 0.989999n$ observations, where the probability of the event is at least 0.39067 at each observation.

Note that, if $v_i^* v_j^* \approx -0.5$ (the angle between two vectors is almost equal to 120^o) then we have $\Pr\left(|wv_i^*| > 0.5003 \;\&\; |wv_j^*| > 0.5003\right) \approx 0$. But, here, the joint probability distribution on all possible pairs of observations is not zero; i.e. We have many vectors v_j^* $(j \in G_\varepsilon)$ with this characteristic that $|wv_j^*| > 0.5003$. Because almost all pairs of vectors in G_ε are almost perpendicular to each other. See Figure 2, where $|wv_i^*| > 0.5003$ and $|wv_j^*| > 0.5003$ if and only if w' lies in the white region.

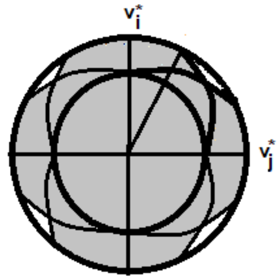

Figure 2. The $v_i^* O v_j^*$ plane, where $\Pr(|wv_i^*| > 0.5003\ \&\ |wv_j^*| > 0.5003) = (0.39067)^2 > 0.15$.

However, we know that the binomial distribution for a given value of $p = 0.39067$ and increasing values of 0.989999n converges to the normal distribution, where for large values of n the binomial distribution may be approximated by the normal distribution with mean $m = 0.39067\text{n}$ and variance $\sigma^2 = 0.39067(1 - 0.39067)\text{n}$. Therefore, the probability that we have $|H_w| \geq 0.000001n$ is approximated as follows:

$$P_N(0.000001n) = \frac{1}{\sqrt{2\pi\sigma}} \int_{0.000001n}^{+\infty} e^{\frac{-(t-m)^2}{2\sigma^2}} dt > \frac{1}{\sqrt{2\pi\sigma}} \int_{0.39067n}^{+\infty} e^{\frac{-(t-m)^2}{2\sigma^2}} dt = P_N(0.39067n) = 0.5.$$

Corollary 3. If the solution of the SDP (1) relaxation meets the Assumptions (1,2), then we have $|G_\varepsilon| \geq 0.989999\text{n}$ and by introducing a normalized random vector w, with a probability of 0.5, the bipartite graph H_w has more than $\frac{n}{1000000}$ vertices. Hence, based on Corollary (2) we have an approximation ratio $\frac{|V_{1G}|}{z^*} \leq 1.999999 < 2$.

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 a suitable solution $V_{1G} \cup V_{-1G}$, correspondingly, where $V_{-1G} = \{j | v_o^* v_j^* < 0\}$. Therefore, the solution does not meet Assumption (1.a) and we have $\frac{|V_{1G}|}{z^*} \leq 1.999999$. 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 the solution does not meet Assumption (1.b) and we have $z^* \geq \frac{n}{2} + 0.0000015n$. Hence, it is sufficient to produce an arbitrary feasible solution, and for all feasible solutions $V = V_{1G} \cup V_{-1G}$ we have $\frac{|V_{1G}|}{z^*} \leq 1.999999$. Otherwise, go to Step 4.

Step 4. We have $|G_{0.0004}| \geq 0.989999n$. Then, if there exists an optimal vector v_k^* for which we have $|F_k| \geq \frac{n}{1000000}$, then we can produce a suitable feasible solution $V_{1G} \cup V_{-1G}$, correspondingly, where $|V_{-1G}| \geq \frac{n}{2000000}$. Therefore, based on Corollary (1) we have $\frac{|V_{1G}|}{z^*} \leq 1.999999 < 2$. Otherwise, go to Step 5.

Step 5. Introduce a normalized random vector w, and produce H_w. Based on Corollary (3), with a probability of 0.5, the set H_w has more than $\frac{n}{1000000}$ vertices and based on it, we can produce a suitable feasible solution $V_{1G} \cup V_{-1G}$, correspondingly, where $|V_{-1G}| \geq \frac{n}{2000000}$. Therefore, based on Corollary (2) we have $\frac{|V_{1G}|}{z^*} \leq 1.999999$. Otherwise, repeat Step 5.

Corollary 4. 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 ratio 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.