A (1.999999)-approximation ratio for vertex cover problem

Majid Zohrehbandian *

Abstract

Until the authenticity of the unique games conjecture is proven, it can be thought to be false. Therefore, I request you, dear reader, to read this paper carefully, and if you don't find any mistake in it, think about the option that maybe unique games conjecture is not correct! If the unique games conjecture is true then it is impossible to produce a less than 2 approximation ratio for the vertex cover problem. Vertex cover problem is a famous combinatorial problem, which its complexity has been heavily studied over the years and while a 2-approximation for it can be trivially obtained, researchers have not been able to approximate it better than 2-$o(1)$. In this paper, by a combination of a new semidefinite programming formulation along with satisfying new proposed 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.

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-ε)-approximation ratio for the vertex cover problem, where the value of ε is not constant. Then, we fix the ε value equal to ε=0.000001 and we show that on arbitrary graphs, a 1.999999-approximation ratio can be obtained by solving a new semidefinite programming (SDP) formulation.

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, by using the satisfying properties, we propose a solution algorithm for VCP 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

*Department of Mathematics, Karaj Branch, Islamic Azad University, Karaj, Iran. zohrebandian@yahoo.com

ratios of VCP feasible solutions which lead to an approximation ratio of 2-ε, 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 ε=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, z^*_{VCP} is the optimal value for the vertex cover problem on G, where $z^*_{VCP} \geqslant \frac{n}{2}$ and we have produced a feasible solution for the problem with vertex partitioning $V = V_1 \cup V_0$ and objective value $|V_1|$. Moreover, we know that we can efficiently solve the following well-known SDP formulation as a relaxation for the VCP formulation, where, to produce the exact solution of the VCP problem, the last constraint should be transformed as $v_o v_j$,$v_i v_j \in \{0, +1\}$,

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

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

$$+v_i v_j + v_i v_k + v_j v_k \geqslant +1 \quad i,j,k \in V \cup \{o\}$$

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

$$v_i v_i = 1, \quad 0 \leqslant v_o v_j \leqslant +1 \quad 0 \leqslant v_i v_j \leqslant +1 \quad i,j \in V \cup \{o\}$$

Theorem 1. Although it is hard to exactly produce the VCP optimal value, let's assume that we have a lower bound on the VCP optimal value and we know $z^*_{VCP} \geqslant \frac{n}{2} + \frac{n}{k} = \frac{(k+2)n}{2k}$. Then, for all vertex cover feasible partitioning $V = V_1 \cup V_0$, we have the approximation ratio $\frac{|V_1|}{z^*_{VCP}} \leqslant \frac{2k}{k+2} < 2$.

Proof. If $z^*_{VCP} \geqslant \frac{(k+2)n}{2k}$ then $\frac{n}{z^*_{VCP}} \leqslant \frac{2k}{k+2}$. Hence, $\frac{|V_1|}{z^*_{VCP}} \leqslant \frac{n}{z^*_{VCP}} \leqslant \frac{2k}{k+2} < 2 \diamond$

Theorem 2. Suppose that the vertex cover problem on G is hard (i.e. $z^*_{VCP} \geqslant \frac{n}{2}$) and we have produced a VCP feasible partitioning $V = V_1 \cup V_0$, where $| V_1 | \leqslant \frac{kn}{k+1}$ and $| V_0 | \geqslant \frac{n}{k+1}$ (or $| V_1 | \leqslant k \mid V_0 |$). Then, based on such a solution we have an approximation ratio $\frac{|V_1|}{z^*_{VCP}} \leqslant \frac{2k}{k+1} < 2$.

Proof. If $| V_1 | \leqslant \frac{kn}{k+1}$ then $n \geqslant \frac{k+1}{k} \mid V_1 \mid$. Hence, $z^*_{VCP} \geqslant \frac{n}{2} \geqslant \frac{k+1}{2k} \mid V_1 \mid$ which concludes that $\frac{|V_1|}{z^*_{VCP}} \leqslant \frac{2k}{k+1} < 2 \diamond$

3 A (1.999999)-approximation algorithm on arbitrary graphs

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

Assumption 1. By solving the SDP (1) relaxation, both of the following conditions occur:

a) For less than 0.000001n of vertices $j \in V$ and corresponding vectors we have $v^*_o v^*_j < 0.5$.

b) For less than 0.01n of vertices $j \in V$ and corresponding vectors we have $v^*_o v^*_j > 0.5004$.

Theorem 3. If the solution of the SDP (1) relaxation does not meet the Assumption (1) then we can produce a solution with a performance ratio of 1.999999.

Proof. If the solution of the SDP (1) relaxation does not meet the Assumption (1.a), then we can introduce $V_0 = \{j \in V \mid v_o^* v_j^* < 0.5\}$ and $V_1 = V - V_0$, to have a feasible solution with $\mid V_0 \mid \geqslant 0.000001n$ and $\mid V_1 \mid \leqslant 0.999999n \leqslant 999999 \mid V_0 \mid$. Then, for such a solution and based on Theorem (2), we have an approximation ratio $\frac{|V_1|}{z^*_{VCP}} < \frac{2(999999)}{999999+1} = 1.999998 < 1.999999$.

Otherwise, if the solution of the SDP (1) relaxation meets the Assumption (1.a) but it does not meet the Assumption (1.b) then we have

$$
\begin{aligned}
z^*_{VCP} \geqslant z^*_{SDP(1)} &\geqslant (0)(0.000001n)_{\{s.t.\ v_o^* v_j^* < 0.5\}} \\
&+(0.5)(0.989999n)_{\{s.t.\ 0.5 \leqslant v_o^* v_j^* \leqslant 0.5004\}} \\
&+(0.5004)(0.01n)_{\{s.t.\ v_o^* v_j^* > 0.5004\}} \\
&= \tfrac{n}{2} + 0.0000035n.
\end{aligned}
$$

Note that, due to the correctness of Assumption (1.a) we have less than 0.000001n of vertices $j \in V$ with $v_o^* v_j^* < 0.5$ and due to the incorrectness of Assumption (1.b) we have more than 0.01n of vertices $j \in V$ with $v_o^* v_j^* > 0.5004$. Therefore, in the above inequality, the first summation is the lower bound on the vertices $j \in V$ with $v_o^* v_j^* < 0.5$ and the third summation is the lower bound on only 0.01n of the vertices $j \in V$ with $v_o^* v_j^* > 0.5004$ (only 0.01n of the vertices with $v_o^* v_j^* > 0.5004$ are considered in third summation and beyond the 0.01n of such vertices are considered in second summation). Moreover, the second summation is the lower bound on the other vertices; i.e. the vertices $j \in V$ with $0.5 \leqslant v_o^* v_j^* \leqslant 0.5004$ or the vertices $j \in V$ with $v_o^* v_j^* > 0.5004$ and beyond the 0.01n of such vertices which have been considered in third summation.

Therefore, based on the above lower bound on z^*_{VCP} value and based on Theorem (1), for all VCP feasible solutions $V = V_1 \cup V_0$, we have the approximation ratio $\frac{|V_1|}{z^*_{VCP}} \leqslant \frac{2(\frac{1}{0.0000035})}{\frac{1}{0.0000035}+2} < 1.999999 \diamond$

Definition 1. Let ε=0.0004 and $G_\varepsilon = \{j \in V \mid 0.5 \leqslant v_o^* v_j^* \leqslant 0.5 + \varepsilon\}$.

Based on Theorem (3), after solving the SDP (1) relaxation,

$\diamond$ If the solution of the SDP (1) relaxation does not meet the Assumption (1) then we have a performance ratio of 1.999999,

$\diamond$ Otherwise (if the solution of the SDP (1) relaxation meets the Assumption (1)), for more than 0.989999n of vertices $j \in V$, we have $0.5 \leqslant v_o^* v_j^* \leqslant 0.5 + \varepsilon$; i.e. $\mid G_\varepsilon \mid \geqslant 0.989999n$. Moreover, for each edge ij in G_ε, we have $v_i^* v_j^* \simeq 0$; i.e. the corresponding vectors of each edge in G_ε are almost perpendicular to each other.

Therefore, to produce a VCP performance ratio of 1.999999 on arbitrary graphs, we need a solution for the SDP (1) relaxation which does not meet the Assumption (1). To do this, we introduce a new SDP model based on the SDP (1) formulation, as follows.

Let $G2 = (V_{new}, E_{new})$ be a new graph based on the connection of two copies of graph G ($G' = G" = G$), where each vertex in G' (one copy of G) is connected to all vertices of $G"$ (the other copy of G). Then, based on SDP model (1), we introduce a new SDP (2) relaxation model as follows:

$$(2)\ min_{s.t.} z = \sum_{i \in V_{new}} v_o v_i$$

$$SDP\ (1)\ constraints\ on\ G'\ and\ G"\ and\ common\ vector\ v_0$$

$$+v_o v_i + v_o v_j - v_i v_j = 1 \quad i \in V'\ ,\ j \in V"$$

$$-1 \leqslant v_i v_j \leqslant +1 \quad i \in V'\ ,\ j \in V"$$

Lemma 1. $z^*(SDP\ (2)) \geqslant 2z^*(SDP\ (1))$

Note that, the VCP feasible solutions on G and $G2$ are corresponding to each other and $z^*_{VCP}(G2) = 2z^*_{VCP}(G)$. In other words, for each VCP feasible partitioning $V = V_1 \cup V_0$, we have $V_{1new} = V'_1 \cup V"_1$ and there are two opposite vector sets V'_0, $V"_0$ where $V_1 = V'_1 = V"_1$ and $V_0 = -V'_0 = V"_0$.

Now, we are going to prove that by solving SDP (2) relaxation, it is not possible to produce a solution which meets Assumption (1) on both graphs G' and $G"$ unless $G2_\varepsilon$ is bipartite on both parts V'_ε and $V"_\varepsilon$.

Theorem 4. For 4 normalized vectors v_1, v_2, v_3, v_4 which are perpendicular to each other, there exist exactly one normalized vector v where $v.v_i = 0.5 \ \ i = 1, 2, 3, 4$. Such a vector v satisfies the equation $v = 0.5(v_1 + v_2 + v_3 + v_4)$.

Proof.

$v_1.v_2 = 0$ and then we have $\mid v_1 + v_2 \mid = \sqrt{\mid v_1 \mid^2 + \mid v_2 \mid^2} = \sqrt{2}$.
$v_3.v_4 = 0$ and then we have $\mid v_3 + v_4 \mid = \sqrt{\mid v_3 \mid^2 + \mid v_4 \mid^2} = \sqrt{2}$.
$(v_1 + v_2).(v_3 + v_4) = 0$ and then we have

$$\mid v_1 + v_2 + v_3 + v_4 \mid = \sqrt{\mid v_1 + v_2 \mid^2 + \mid v_3 + v_4 \mid^2} = 2.$$

Finally, we have $(v_1 + v_2 + v_3 + v_4).v = 2$. Hence, $\mid v_1 + v_2 + v_3 + v_4 \mid . \mid v \mid .cos(\theta) = 2$ and this concludes that $\theta = 0$ and $v = 0.5(v_1 + v_2 + v_3 + v_4)$ $\diamond$

Corollary 1. For 4 normalized vectors v_1, v_2, v_3, v_4 which are almost perpendicular to each other, a normalized vector v where $v.v_i \simeq 0.5 \ \ i = 1, 2, 3, 4$, satisfies the equation $v \simeq 0.5(v_1 + v_2 + v_3 + v_4)$.

Theorem 5. By solving SDP (2) relaxation on $G2$, it is not possible to have an optimal solution which meets Assumption (1) on both graphs G' and $G"$ unless $G2_\varepsilon$ is bipartite on both parts V'_ε and $V"_\varepsilon$.

Proof. Suppose that we have an optimal solution which meets Assumption (1) on both graphs G' and $G"$. Therefore, for an edge ab in G'_ε and an edge cd in $G"_\varepsilon$ (a complete subgraph of $G2$ on four vertices a, b, c, d) we have 4 normalized vectors v_a, v_b, v_c, v_d which are almost perpendicular to each other.

Moreover, we have a normalized vector v_o where $v_o v_h \simeq 0.5 \ \ h = a, b, c, d$. Hence, based on Corollary (1) we have $v_o \simeq 0.5(v_a + v_b + v_c + v_d)$. This means that for each edge ij in G'_ε we have $v_o \simeq 0.5(v_i + v_j + v_c + v_d)$, and for each edge ij in $G"_\varepsilon$ we have $v_o \simeq 0.5(v_a + v_b + v_i + v_j)$.

Therefore, for each edge ij in G'_ε we have $v_i + v_j \simeq 2v_o - v_c - v_d = U$, and for each edge ij in $G"_\varepsilon$ we have $v_i + v_j \simeq 2v_o - v_a - v_b = W$, where, due to almost perpendicular property of v_i and v_j we have $\mid U \mid \simeq \mid W \mid \simeq \sqrt{\mid v_i \mid^2 + \mid v_j \mid^2} = \sqrt{2}$.

Now, suppose that we have an odd cycle on t vertices $1, 2, ..., t$, in G'_ε, where $t = 2k+1$ is an odd number. Then, by addition of the vectors in this cycle, we have $S = (v_1 + v_2) + (v_2 + v_3) + ... + (v_t + v_1) \simeq tU$.

But, the above summation can do as $S = 2(v_1 + v_2 + v_3 + ... + v_{t-2} + v_{t-1} + v_t)$ to produce the following results, which all of them must be $\simeq tU$.

$$S = 2((v_1 + v_2) + (v_3 + v_4) + ... + (v_{t-2} + v_{t-1}) + v_t) \simeq 2(kU + v_t) = (t-1)U + 2v_t$$
$$S = 2((v_2 + v_3) + (v_4 + v_5) + ... + (v_{t-1} + v_t) + v_1) \simeq (t-1)U + 2v_1$$
$$S = 2((v_3 + v_4) + (v_5 + v_6) + ... + (v_t + v_1) + v_2) \simeq (t-1)U + 2v_2$$
$$...$$
$$S = 2((v_t + v_1) + (v_2 + v_3) + ... + (v_{t-3} + v_{t-2}) + v_{t-1}) \simeq (t-1)U + 2v_{t-1}$$
$$-----------------------------$$
$$v_1 \simeq v_2 \simeq ... \simeq v_{t-1} \simeq v_t \simeq 0.5U$$

Hence $\mid U \mid \simeq 2 \mid v_1 \mid \simeq 2 \neq \sqrt{2}$ and this is a contradiction; e.g. $v_1 v_2 \simeq (0.5U).(0.5U) \neq 0$. Therefore, there is not any odd cycle in G'_ε, and similarly, there is not any odd cycle in $G"_\varepsilon$. Therefore, if the optimal solution of SDP (2) relaxation on $G2$ meets the Assumption (1) on both graphs G' and $G"$, then both of the subgraphs G'_ε and $G"_\varepsilon$ are bipartite $\diamond$

In other words, to produce a performance ratio of 1.999999, we should solve the SDP (2) relaxation. Then, if the solution of the SDP (2) relaxation does not meet Assumptions (1), we have a performance ratio of 1.999999. Otherwise, VCP problem on G'_ε is simple and produce a performance ratio of 1.999999. Therefore, our algorithm to produce an approximation ratio 1.999999 on arbitrary graphs is as follows:

Mahdis Algorithm (To produce a vertex cover solution with a ratio factor less than 1.999999)
Step 1. Solve the SDP (2) relaxation.
Step 2. If for more than 0.000001n of vertices $j \in V'$ and corresponding vectors we have $v_o^* v_j^* < 0.5$, then produce a suitable solution $V_1 \cup V_0$, correspondingly, where $V_0 = \{j \in V' \mid v_o^* v_j^* < 0.5\}$ and $V_1 = V' - V_0$. Hence, the solution does not meet Assumption (1.a) and we have $\frac{|V_1|}{z^*_{VCP}} \leqslant 1.999999$. Otherwise, go to Step 3.
Step 3. If for more than 0.000001n of vertices $j \in V"$ and corresponding vectors we have $v_o^* v_j^* < 0.5$, then produce a suitable solution $V_1 \cup V_0$, correspondingly, where $V_0 = \{j \in V" \mid v_o^* v_j^* < 0.5\}$ and $V_1 = V" - V_0$. Hence, the solution does not meet Assumption (1.a) and we have $\frac{|V_1|}{z^*_{VCP}} \leqslant 1.999999$. Otherwise, go to Step 4.
Step 4. If for more than 0.01n of vertices $j \in V'$ and corresponding vectors we have $v_o^* v_j^* > 0.5004$, then for all feasible solutions $V = V_1 \cup V_0$ we have $\frac{|V_1|}{z^*_{VCP}} \leqslant 1.999999$ and it is sufficient to produce an arbitrary VCP feasible solution. Otherwise, go to Step 5.
Step 5. If for more than 0.01n of vertices $j \in V"$ and corresponding vectors we have $v_o^* v_j^* > 0.5004$, then for all feasible solutions $V = V_1 \cup V_0$ we have $\frac{|V_1|}{z^*_{VCP}} \leqslant 1.999999$ and it is sufficient to produce an arbitrary VCP feasible solution. Otherwise, go to Step 6.
Step 6. Based on Theorem (5), the solution does not meet the Assumption (1) and then $G2_\varepsilon$ is bipartite on both parts V'_ε and $V"$, where $\mid V'_\varepsilon \mid, \mid V"_\varepsilon \mid \geqslant 0.989999n$. Solve VCP problem on bipartite subgraph G'_ε to produce a feasible solution $V_1 \cup V_0$ for which we have $\frac{|V_1|}{z^*_{VCP}} \leqslant 1.999999$.

Corollary 2. 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 lead to the conclusion that the unique games conjecture is not true.

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 $2-\varepsilon$. Journal of Computer and System Sciences, 74, 335-349.