On Approximating Minimum Vertex Cover for Graphs with Perfect Matching*

Jianer Chen† and Iyad A. Kanj‡

Abstract

It has been a challenging open problem whether there is a polynomial time approximation algorithm for the Vertex Cover problem whose approximation ratio is bounded by a constant less than 2. In this paper, we study the Vertex Cover problem on graphs with perfect matching (shortly, VC-PM). We show that if the VC-PM problem has a polynomial time approximation algorithm with approximation ratio bounded by a constant less than 2, then so does the Vertex Cover problem on general graphs. Approximation algorithms for VC-PM are developed, which induce improvements over previously known algorithms on sparse graphs. For example, for graphs of average degree 5, the approximation ratio of our algorithm is 1.414, compared with the previously best ratio 1.615 by Halldórsson and Radhakrishnan.

Keywords. vertex cover, graph matching, approximation algorithm, inapproximability

1 Introduction

Approximation algorithms for NP-hard optimization problems have been a very active research in recent years. In particular, the study of approximability for certain famous NP-hard optimization problems has achieved great success. For example, now it is known that the polynomial time approximability for the Max-3Sat problem is exactly 8/7, based on the lower bound derived by Håstad [9], and the fact that a random truth assignment satisfies, on average, 7/8 of the optimal number of satisfiable clauses [13].

On the other hand, some other famous NP-hard optimization problems still resist stubbornly improvements. A well-known example is the Vertex Cover problem. A very simple approximation algorithm based on maximal matchings gives an approximation ratio 2 for the Vertex Cover problem. However, despite long time efforts, no significant progress has been made on this ratio bound. It has become an outstanding open problem whether there is a polynomial time approximation algorithm for the Vertex Cover problem whose approximation ratio is bounded by a constant less than 2. On the other hand, the best lower bound for the ratio is $10\sqrt{5}-21 \approx 1.360$, which was derived by Dinur and Safra [5].

Considerable efforts have been made on trying to improve the upper bound on the approximability for the Vertex Cover problem. Hochbaum [10] presented an algorithm of approximation ratio $2-2/d$ for graphs of degree bounded by d. Monien and Speckenmeyer [15] improved this bound to $1-(\log\log n)/(2\log n)$. The same bound was also achieved independently by Bar-Yehuda

*A preliminary version of this work was presented at *The Eleventh International Symposium on Algorithms and Computation* (ISAAC 2000), Taipei, Taiwan, December 18 – December 20, 2000, and published in *Lecture Notes in Computer Science 1969*, pp. 132-143, 2000.

†Department of Computer Science, Texas A&M University, College Station, TX 77843-3112. Email: chen@cs.tamu.edu. This work was supported in part by the National Science Foundation under the Grants CCR-0311590 and CCF-0430683.

‡School of CTI, DePaul University, 243 S. Wabash Avenue, Chicago, IL 60604. Email: ikanj@cs.depaul.edu. This work was supported in part by DePaul University Competitive Research Grant.

and Even [1], whose result is also applicable to the weighted vertex cover problem. However, no further progress has been made on this bound for the last one and a half decades. For the VERTEX COVER problem on sparse graphs, Berman and Fujito [2] presented an approximation algorithm for graphs of degree bounded by 3, whose approximation ratio is bounded by $7/6 + \epsilon$. Halldórsson [7] developed an algorithm of ratio $2 - \log d/d(1 + \epsilon)$ for graphs of degree bounded by d. On graphs of average degree $\bar{d}$, Hochbaum [10] has studied the approximation algorithms for the INDEPENDENT SET problem. Halldórsson and Radhakrishnan [8] further improved Hochbaum's algorithm. Under the assumption that a minimum vertex cover of the input graph contains at least half of the vertices in the graph, Halldórsson and Radhakrishnan's algorithm implies an algorithm of approximation ratio $(4\bar{d} + 1)/(2\bar{d} + 3)$ for the VERTEX COVER problem on graphs of average degree $\bar{d}$.

For general graphs, no polynomial time approximation algorithms have been developed for the VERTEX COVER problem whose approximation ratios are bounded by a constant less than 2. Hochbaum has once conjectured that no such approximation algorithm exists [10]. Motivated by these facts, we study in the current paper the VERTEX COVER problem on graphs with perfect matching (or shortly, the VC-PM problem). The VERTEX COVER problem and graph matching are closely related. In fact, the maximum matching problem is the dual problem of the minimum vertex cover problem when they are given in their integer linear programming forms. We first show that unless P = NP, the VC-PM problem cannot be approximated in polynomial time to a ratio $5\sqrt{5} - 10 - \epsilon \approx 1.180 - \epsilon$ for any constant $\epsilon > 0$. We then show that if the VC-PM problem has a polynomial time approximation algorithm with approximation ratio bounded by a constant less than 2, then so does the VERTEX COVER problem on general graphs. Approximation algorithms for the VC-PM problem are then investigated, with its close relation to the MAX-2SAT problem. A polynomial time approximation algorithm is developed for the VC-PM problem, which induces improvements over previous algorithms for the VERTEX COVER problem on sparse graphs. For example, for graphs of average degree 5, the approximation ratio of our algorithm is 1.414, compared with the previously best ratio 1.615 by Halldórsson and Radhakrishnan [8].

We note that after the publication of a preliminary version of this paper, there was a very recent improvement on these results by Chlebík and Chlebíková [3], who showed that the VERTEX COVER problem and the VC-PM problem have the same inapproximability threshold (Theorem 1, [3]).

2 Preliminaries

We briefly review the related terminologies and previous results used in this paper. Let $G = (V, E)$ be a graph. A *vertex cover* C for G is a set of vertices in G such that every edge in E has at least one endpoint in C. An *independent set* I in G is a set of vertices in G such that no two vertices in I are adjacent. It is easy to see that a set $C \subseteq V$ is a vertex cover for G if and only if the complement set $V - C$ is an independent set in G. The VERTEX COVER problem is to construct for a given graph a vertex cover of the minimum number of vertices, and the INDEPENDENT SET problem is to construct for a given graph an independent set of the maximum number of vertices. Both VERTEX COVER and INDEPENDENT SET are well-known NP-hard problems [12].

A *matching* M in a graph $G = (V, E)$ is a set of edges in G such that no two edges in M share a common endpoint. A vertex is *matched* if it is an endpoint of an edge in M, and is *unmatched* otherwise. A matching M in G is *maximal* if no edge can be added to M to make a larger matching. A matching M is *maximum* if no matching in G is larger than M. The MAXIMUM MATCHING problem is to construct for a given graph a maximum matching. A graph G of n vertices has a *perfect matching* if G has a matching of $n/2$ edges. Given a matching M in a graph G, an *augmenting path* in G (with respect to M) is a simple path $\{u_0, u_1, \ldots, u_{2k+1}\}$ of odd length such that u_0 and u_{2k+1} are unmatched, and the edges $[u_{2i-1}, u_{2i}]$, $i = 1, \ldots, k$, are in the matching M. It is well-known that a matching M is maximum if and only if there is no augmenting path in G with respect to M. The Maximum Matching problem can be solved in time $O(m\sqrt{n})$ [14]. In particular, it can be tested in time $O(m\sqrt{n})$ whether a graph has a perfect matching.

The current paper will concentrate on the Vertex Cover problem on graphs with perfect matching. Formally, *the VC-PM problem* is, for a graph G with perfect matching, to construct a minimum vertex cover for G. It is not difficult to prove, via a standard reduction from the 4-Satisfiability problem, that the VC-PM problem is NP-hard.

Let $G = (V, E)$ be a graph. For a subset $V' \subseteq V$ of vertices in G, denote by $G(V')$ the subgraph induced by V'. That is, the vertex set of the subgraph $G(V')$ is V', and an edge e in G is in $G(V')$ if and only if both endpoints of e are in V'. We denote by $Opt(G)$ the size of the minimum vertex cover for the graph G. The importance of the following proposition, due to Nemhauser and Trotter [16], to the approximation of the Vertex Cover problem was first observed by Hochbaum [10].

Proposition 2.1 (NT-Theorem) *Given a graph G, there is an $O(m\sqrt{n})$ time algorithm that partitions the vertex set of G into three subsets I_0, C_0, and V_0 such that*
(1) *$Opt(G(V_0)) \geq |V_0|/2$; and*
(2) *for any vertex cover C of $G(V_0)$, $C \cup C_0$ is a vertex cover of G satisfying*

$$\frac{|C \cup C_0|}{Opt(G)} \leq \frac{|C|}{Opt(G(V_0))}.$$

According to the NT-Theorem, the approximation ratio on vertex cover for the graph $G(V_0)$ implies an equally good approximation ratio on vertex cover for the original graph G. Thus, we only need to concentrate on approximating vertex cover for the graph $G(V_0)$, for which the minimum vertex cover has a lower bound $|V_0|/2$.

We say that a graph G is *everywhere k-sparse* if for any subset V' of vertices in G, the number of edges in the induced subgraph $G(V')$ is bounded by $k|V'|$. For a graph G of n vertices and m edges, we define the *average degree* $\bar{d}$ of G by $\bar{d} = 2m/n$.

3 On the inapproximability of VC-PM

Theorem 3.1 *Unless $P = NP$, the* VC-PM *problem has no polynomial time approximation algorithm with ratio $5\sqrt{5} - 10 - \epsilon \approx 1.180 - \epsilon$ for any constant $\epsilon > 0$.*

Proof. Suppose to the contrary that there is a polynomial time approximation algorithm A_{pm} of ratio $r = 5\sqrt{5} - 10 - \epsilon$ for the VC-PM problem, where $\epsilon > 0$ is a constant. We show that this would imply a polynomial time approximation algorithm of ratio $10\sqrt{5} - 21 - \delta$ for the Vertex Cover problem on general graphs for some constant $\delta > 0$, which, using [5], would imply that P = NP.

Let G be a graph with n vertices. By the NT-Theorem, we can assume that $Opt(G) \geq n/2$. Construct a maximal matching M for G. Let I be the set of the unmatched vertices. Then I is an independent set in G. Let $s = |I|$. Since $Opt(G) \geq n/2$, we have $s \leq n/2$. Introduce a new clique Q of s vertices (Q is disjoint with G). Pair the vertices in Q and the vertices in I arbitrarily, and connect each pair by a new edge. Let the resulting graph be G_+. The graph G_+ has $n + s$ vertices, and has a perfect matching. Moreover, it is easy to verify that

$$Opt(G) + s - 1 \leq Opt(G_+) \leq Opt(G) + s. \tag{1}$$

Now apply the approximation algorithm A_{pm} on the graph G_+, we get a vertex cover C_+ for the graph G_+. By our assumption, $|C_+|/Opt(G_+) \leq r$. Remove all vertices in $C_+ \cap Q$ from C_+, we get a vertex cover C for the graph G. Since C_+ contains at least $s-1$ vertices and at most s vertices in Q, we have

$$|C| + s - 1 \leq |C_+| \leq |C| + s. \tag{2}$$

Consider the approximation ratio for the vertex cover C for the graph G:

$$\begin{aligned}
\frac{|C|}{Opt(G)} &\leq \frac{|C_+| - s + 1}{Opt(G_+) - s} = 1 + \frac{|C_+| - Opt(G_+) + 1}{Opt(G_+) - s} \\
&= 1 + \frac{(|C_+|/Opt(G_+)) - 1 + (1/Opt(G_+))}{1 - (s/Opt(G_+))} \\
&\leq 1 + \frac{r - 1 + (1/Opt(G_+))}{1 - (s/Opt(G_+))} \\
&\leq 1 + \frac{r - 1 + (1/Opt(G_+))}{1 - ((Opt(G_+) - n/2 + 1)/Opt(G_+)} \\
&= 1 + \frac{(r - 1 + (1/Opt(G_+))) \cdot Opt(G_+)}{(n/2) - 1} \\
&\leq 1 + \frac{(r - 1 + (1/Opt(G_+))) \cdot n}{(n/2) - 1} \\
&= 1 + \frac{(r - 1 + (1/Opt(G_+)))(2n)}{n - 2} \\
&= 1 + \frac{(r - 1 + (1/Opt(G_+)))(2n - 4)}{n - 2} + \frac{4(r - 1 + (1/Opt(G_+)))}{n - 2} \\
&= 1 + 2(r-1) + \frac{2}{Opt(G_+)} + \frac{4(r - 1 + (1/Opt(G_+)))}{n - 2}.
\end{aligned} \tag{3}$$

The first inequality follows from the relations (2) and (1), the second inequality is true due to the assumption $|C_+|/Opt(G_+) \leq r$. The third inequality follows from (1) since we have $s \leq Opt(G_+) - Opt(G) + 1$, and from our assumption $Opt(G) \geq n/2$. The fourth inequality is true because the vertices in Q plus the vertices in M obviously make a vertex cover for the graph G_+, so $Opt(G_+) \leq 2|M| + |Q| = n$.

Since $r = 5\sqrt{5} - 10 - \epsilon$, we have $1 + 2(r-1) = 10\sqrt{5} - 21 - 2\epsilon$. Now for n sufficiently large, and observing that $Opt(G_+) \geq n/2 + s - 1$, we conclude from (3) that $|C|/Opt(G) \leq 10\sqrt{5} - 21 - \delta$ for some constant $\delta > 0$. This, according to [5], would imply that P = NP. □

Similar results to those of Theorem 3.1 also hold for the VC-PM problem on graphs of bounded degree and on everywhere sparse graphs.

Theorem 3.2 *For any constant $\epsilon > 0$, there is a constant B such that unless unlikely consequences occur in complexity theory, there is no polynomial time approximation algorithm with ratio $13/12 - \epsilon$ for the* VC-PM *problem on graphs of degree bounded by B nor on everywhere B-sparse graphs.*

PROOF. By Clementi and Trevisan [4], for any $\epsilon > 0$, there is a constant B_ϵ such that unless P = NP, there is no polynomial time approximation algorithm with ratio $7/6 - \epsilon$ for the VERTEX-COVER problem on graphs of degree bounded by B_ϵ, and that unless NP = co-RP, there is no polynomial time approximation algorithm with ratio $7/6 - \epsilon$ for the VERTEX-COVER problem on everywhere B_ϵ-sparse graphs.

The proof now goes in exactly the same logic as that for Theorem 3.1 except that instead of introducing a single new clique Q of size s, here we introduce s/B_ϵ cliques of size B_ϵ. It is easy to see that if the original graph is of degree bounded by B_ϵ (resp. everywhere B_ϵ-sparse), then the new graph G_+ is of degree bounded by $B_\epsilon + 1$ (resp. everywhere $(B_\epsilon + 1)$-sparse). The analysis for the approximation ratio goes through with straightforward modifications. □

4 Vertex Cover and VC-PM

In this section, we study the relation between approximating the VC-PM problem and approximating the VERTEX COVER problem on general graphs. We show that in order to overcome the bound 2 approximability barrier of the VERTEX COVER problem for general graphs, it suffices to overcome this barrier for the VC-PM problem.

Theorem 4.1 *If the* VC-PM *problem has a polynomial time approximation ratio $r \leq 2$, then the* VERTEX COVER *problem on general graphs has a polynomial time approximation ratio $(r + 2)/2$.*

PROOF. Suppose that there is a polynomial time approximation algorithm A_{pm} of approximation ratio r for the VC-PM problem.

Given a graph $G = (V, E)$, where $|V| = n$, according to the NT-Theorem, we can assume $Opt(G) \geq n/2$. Pick any maximal matching M in G. Let V_M be the set of vertices that are endpoints of the edges in M, and let $I = V - V_M$. Since M is a maximal matching, I is an independent set and V_M is a vertex cover for G. Moreover, it is also easy to see that if C_M is a vertex cover for the graph $G(V_M)$, then $C_M \cup I$ is a vertex cover for the graph G.

Let $c = (2 - r)/4$. Consider the following algorithm: if $|I| \geq cn$, then return V_M, while if $|I| < cn$ then call the algorithm A_{pm} on the graph with perfect matching $G(V_M)$ to get a vertex cover C_M for the graph $G(V_M)$ and return $C_M \cup I$.

This algorithm obviously constructs a vertex cover C for the graph G. In case $|I| \geq cn$, we have

$$\frac{|C|}{Opt(G)} = \frac{|V_M|}{Opt(G)} = \frac{n - |I|}{Opt(G)} \leq \frac{n - cn}{n/2} = 2 - 2c = \frac{r + 2}{2}.$$

On the other hand in case $|I| < cn$, since $Opt(G(V_M)) \leq Opt(G)$, we have

$$\frac{|C|}{Opt(G)} = \frac{|C_M| + |I|}{Opt(G)} \leq \frac{|C_M|}{Opt(G(V_M))} + \frac{|I|}{Opt(G)} \leq r + \frac{cn}{n/2} = r + 2c = \frac{r + 2}{2}.$$

This completes the proof. □

In particular, if the VC-PM problem has a polynomial time approximation ratio $r < 2$ for a constant r, then the VERTEX COVER problem on general graphs has a polynomial time approximation ratio $(r + 2)/2 < 2$.

In the following, we present a better approximation algorithm for the VERTEX COVER problem via approximating the VC-PM problem, which will also be used when we consider approximating the VERTEX COVER problem on everywhere sparse graphs.

Let M be a maximum matching in a graph G. For each matched vertex u in M, we will denote by u' the partner of u in M (i.e., $[u, u'] \in M$). Note that the set I_M of the unmatched vertices is an independent set in G.

Definition Let M be a maximum matching in a triangle-free graph G, and let I_M be as above. Define the following sets.

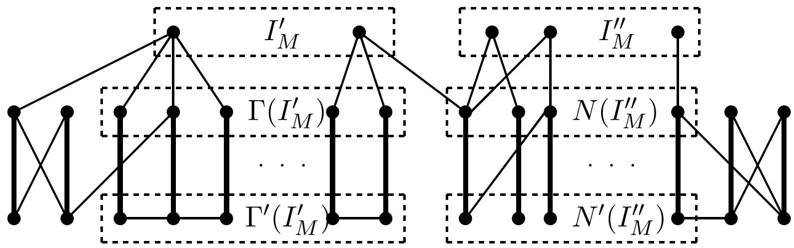

Figure 1: Illustration of the sets I'_M, I''_M, $\Gamma(I'_M)$, $\Gamma'(I'_M)$, $N(I''_M)$, and $N'(I''_M)$

- I'_M is the subset of unmatched vertices w in I_M such that there are two edges $[u, u']$ and $[v, v']$ in M and (w, u, u', v', v, w) is a 5-cycle.
- $I''_M = I_M - I'_M$.
- $\Gamma(I'_M)$ is the set of matched vertices u such that a vertex w in I'_M and two edges $[u, u']$ and $[v, v']$ in M make a 5-cycle (w, u, u', v', v, w).
- $N(I''_M)$ is the set of matched vertices u such that $[z, u]$ is an edge in G for some $z \in I''_M$.
- $\Gamma'(I'_M) = \{u' \mid [u, u'] \in M \text{ and } u \in \Gamma(I'_M)\}$.
- $N'(I''_M) = \{u' \mid [u, u'] \in M \text{ and } u \in N(I''_M)\}$.

See Figure 1 for illustration, where thicker lines are edges in the matching M.

We first list a number of properties for these sets. Note that since the matching M is maximum, there is no augmenting path in the graph G with respect to M.

Fact 1. The six sets I'_M, I''_M, $\Gamma(I'_M)$, $\Gamma'(I'_M)$, $N(I''_M)$, and $N'(I''_M)$ are mutually disjoint.

By the definitions, each of the sets I'_M, I''_M is mutually disjoint with all the other five sets. For a vertex u in $\Gamma(I'_M)$ with a 5-cycle (w, u, u', v', v, w), where $w \in I'_M$, we show that vertex u cannot be in any other set: (1) $u \in N(I''_M)$ would imply an augmenting path (w, v, v', u', u, z), for some $z \in I''_M$; (2) $u \in N'(I''_M)$ implies that $u' \in N(I''_M)$ and (w, u, u', z) would be an augmenting path, for some $z \in I''_M$; and (3) $u \in \Gamma'(I'_M)$ implies $u' \in \Gamma(I'_M)$ so (w, u, u', w') would be an augmenting path, for some $w' \in I'_M$ (note that $w' \neq w$ since the graph G is triangle-free). Summarizing all these, we conclude that the set $\Gamma(I'_M)$ is disjoint with all the other five sets. Similarly, we can show that each of the sets $\Gamma'(I'_M)$, $N(I''_M)$, and $N'(I''_M)$ is disjoint with all the other five sets.

Fact 2. Let $w_1 \in I'_M$ with a 5-cycle $C_1 = (w_1, u_1, u'_1, v'_1, v_1, w_1)$ and $w_2 \in I'_M$ with a 5-cycle $C_2 = (w_2, u_2, u'_2, v'_2, v_2, w_2)$. If $w_1 \neq w_2$ then the 5-cycles C_1 and C_2 are disjoint.

This is true because, for example, if $u_1 = u_2$, then there would be an augmenting path $(w_1, u_1, u'_1, v'_2, v_2, w_2)$.

Fact 3. $2|I'_M| \leq |\Gamma(I'_M)| = |\Gamma'(I'_M)|$, and $|N(I''_M)| = |N'(I''_M)|$.

The equalities $|\Gamma(I'_M)| = |\Gamma'(I'_M)|$ and $|N(I''_M)| = |N'(I''_M)|$ follow from the definitions and the disjointness of the sets, and $2|I'_M| \leq |\Gamma(I'_M)|$ follows from Fact 2.

Fact 4. The sets I''_M, $\Gamma(I'_M)$, and $N'(I''_M)$ are all independent sets in G.

The set I''_M is an independent set since I''_M is a subset of the independent set I_M. Let $u_1 \in \Gamma(I'_M)$ with a 5-cycles $C_1 = (w_1, u_1, u'_1, v'_1, v_1, w_1)$ and let $u_2 \in \Gamma(I'_M)$ with a 5-cycle $C_2 = (w_2, u_2, u'_2, v'_2, v_2, w_2)$, where $w_1, w_2 \in I'_M$. If $[u_1, u_2]$ is an edge, then $w_1 \neq w_2$ since the graph G is triangle-free. By Fact 2, the two cycles C_1 and C_2 are disjoint. Thus, there would be an augmenting path $(w_1, v_1, v'_1, u'_1, u_1, u_2, u'_2, v'_2, v_2, w_2)$. This contradiction shows that the set

> **Algorithm. VC-Apx.**
>
> Input: a triangle-free graph $G = (V, E)$ with $|V| = n$ and $Opt(G) \geq n/2$.
>
> Output: a vertex cover for the graph G.
>
> 1. construct a maximum matching M in G, let V_M be the set of matched vertices;
> 2. construct the sets I'_M, I''_M, $\Gamma(I'_M)$, $\Gamma'(I'_M)$, $N(I''_M)$, and $N'(I''_M)$ given in the Definition;
> 3. $c = (2 - r)/3 + |I'_M|/n$;
> 4. **if** $|I'_M| + |I''_M| + |N'(I''_M)| + |\Gamma(I'_M)| \geq cn$
>
>   **then return** $C_1 = V - (I''_M \cup N'(I''_M) \cup \Gamma(I'_M))$
>
>   **else** let S be the set of minimum cardinality among I''_M and $N(I''_M)$;
>
>   apply the algorithm A_{pm} to $G(V_M)$ and let C be the vertex cover returned by A_{pm};
>
>   **return** $C_2 = C \cup S \cup I'_M$.

Figure 2: The algorithm **VC-Apx**

$\Gamma(I'_M)$ must be an independent set. Finally, let $u'_1, u'_2 \in N'(I''_M)$, and $[u'_1, u'_2]$ is an edge in G, then $(z_1, u_1, u'_1, u'_2, u_2, z_2)$, where $z_1, z_2 \in I''_M$, would be an augmenting path (note $z_1 \neq z_2$ since otherwise $z_1 = z_2$ would be a vertex in I'_M). Thus, the set $N'(I''_M)$ must be an independent set.

Fact 5. The set $I''_M \cup N'(I''_M) \cup \Gamma(I'_M)$ is an independent set in G.

By Fact 4, it suffices to prove that there is no edge between the sets I''_M, $N'(I''_M)$, and $\Gamma(I'_M)$. Let z be a vertex in I''_M, u be a vertex in $\Gamma(I'_M)$ with a 5-cycle (w, u, u', v', v, w), where $w \in I'_M$, and x' be a vertex in $N'(I''_M)$ such that $[x, x'] \in M$ and $[x, y]$ is an edge where $y \in I''_M$. The edge $[u, x']$ would imply an augmenting path $(y, x, x', u, u', v', v, w)$. The edge $[z, u]$ implies $u \in N(I''_M)$ and the edge $[z, x']$ implies $x' \in N(I''_M)$, both contradicting the disjointness of the sets $\Gamma(I'_M)$, $N(I''_M)$ and $N'(I''_M)$ proved in Fact 1.

Theorem 4.2 *If the* VC-PM *problem has a polynomial time approximation algorithm of ratio* r*, then the* Vertex Cover *problem on general graphs has a polynomial time approximation algorithm of ratio* $\max\{1.5, (2r + 2)/3\}$.

Proof. Let A_{pm} be an approximation algorithm of ratio r for the VC-PM problem.

We first assume that the input graph G of n vertices is triangle-free and satisfies $Opt(G) \geq n/2$. Consider the algorithm **VC-Apx** given in Figure 2.

We analyze the approximation ratio for the algorithm **VC-Apx**.

If $|I'_M| + |I''_M| + |N'(I''_M)| + |\Gamma(I'_M)| \geq cn$, then the set $C_1 = V - (I''_M \cup N'(I''_M) \cup \Gamma(I'_M))$ is returned. By Fact 5, the set $I''_M \cup N'(I''_M) \cup \Gamma(I'_M)$ is an independent set in G so the set C_1 is a vertex cover for G. Moreover, from $|I'_M| + |I''_M| + |N'(I''_M)| + |\Gamma(I'_M)| \geq cn$, we have $|I''_M| + |N'(I''_M)| + |\Gamma(I'_M)| \geq cn - |I'_M|$, hence $|C_1| \leq n - (cn - |I'_M|)$. Since $Opt(G) \geq n/2$, we have:

$$\frac{|C_1|}{Opt(G)} \leq \frac{n - (cn - |I'_M|)}{n/2} = 2 - 2c + \frac{2|I'_M|}{n} = \frac{2r + 2}{3}.$$

Consider now the case $|I'_M| + |I''_M| + |N'(I''_M)| + |\Gamma(I'_M)| < cn$. Since by the definitions the vertices in I''_M are only adjacent to vertices in the set $N(I''_M)$, the set S, which is either I''_M or $N(I''_M)$, will cover all the edges incident on vertices in I''_M. Since C is a vertex cover for the induced subgraph $G(V_M)$, the set $C_2 = C \cup S \cup I'_M$ is a vertex cover for the original graph G. Now

$$|I'_M| + |I''_M| + |N'(I''_M)| + |\Gamma(I'_M)| \quad \geq \quad 3|I'_M| + |I''_M| + |N(I''_M)|$$

$$\geq \quad 3|I'_M| + 2\min\{|I''_M|, |N(I''_M)|\} = 3|I'_M| + 2|S|.$$

The first inequality has used the relations $|\Gamma(I'_M)| \geq 2|I'_M|$ and $|N'(I''_M)| = |N(I''_M)|$ in Fact 3. Thus, $3|I'_M| + 2|S| < cn$ and $|S| < (cn - 3|I'_M|)/2$. This gives

$$|C_2| = |C \cup S \cup I'_M| \leq |C| + |S| + |I'_M| \leq |C| + \frac{cn - 3|I'_M|}{2} + |I'_M| = |C| + \frac{cn - |I'_M|}{2}.$$

By our assumption, $|C|/Opt(G(V_M)) \leq r$. We also have $Opt(G) \geq Opt(G(V_M))$ and $Opt(G) \geq n/2$. We finally derive the ratio

$$\frac{|C_2|}{Opt(G)} \quad \leq \quad \frac{|C|}{Opt(G)} + \frac{cn - |I'_M|}{2Opt(G)} \leq \frac{|C|}{Opt(G(V_M))} + \frac{cn - |I'_M|}{n} \leq r + c - \frac{|I'_M|}{n} = \frac{2r+2}{3}.$$

This proves that the algorithm **VC-Apx** on a triangle-free graph G with $Opt(G) \geq n/2$ returns a vertex cover C for G satisfying $|C|/Opt(G) \leq (2r+2)/3$.

Now we consider a general graph $G' = (V', E')$. We first remove all disjoint triangles from the graph G' (in an arbitrary order). Let the resulting graph be G and let V_Δ be the set of vertices of the removed triangles. Then, G is the triangle-free subgraph induced by the vertex set $V = V' - V_\Delta$. Now apply the NT-Theorem to the graph G and let I_0, C_0, and V_0 be the three vertex sets given in the NT-Theorem. Then the induced subgraph $G(V_0)$ is triangle-free and satisfies $Opt(G(V_0)) \geq |V_0|/2$. Thus, we can apply the algorithm **VC-Apx** to the graph $G(V_0)$. Let C be the vertex cover returned by the algorithm **VC-Apx** on the graph $G(V_0)$. By the discussion above, we have $|C|/Opt(V_0) \leq (2r+2)/3$. According to the NT-Theorem, $C_1 = C \cup C_0$ is a vertex cover for the graph G satisfying $|C_1|/Opt(G) \leq (2r+2)/3$.

Obviously, $C_2 = C_1 \cup V_\Delta$ is a vertex cover of the original graph G'. According to the Local-Ratio Theorem by Bar-Yehuda and Even [1], we have

$$\frac{|C_2|}{Opt(G')} \leq \max\left\{\frac{|V_\Delta|}{Opt(G(V_\Delta))}, \frac{|C_1|}{Opt(G)}\right\} \leq \max\left\{\frac{|V_\Delta|}{Opt(G(V_\Delta))}, \frac{2r+2}{3}\right\}.$$

The theorem follows from $|V_\Delta|/Opt(G(V_\Delta)) \leq 1.5$, because every vertex cover of $G(V_\Delta)$ contains at least two vertices from each triangle in $G(V_\Delta)$. □

We remark that it is possible to extend the method of Theorem 4.2 to first consider graphs that have neither triangles nor 5-cycles. However, when we work on general graphs, we need to apply the Local-Ratio Theorem of Bar-Yehuda and Even to eliminate all 5-cycles. This makes the resulting algorithm have approximation ratio at least $5/3 > 1.66$.

5 On approximating VC-PM

In this section, we study approximation algorithms for the VC-PM problem.

Recall that an instance of the Max-2Sat problem is a set F of clauses in which each clause is a disjunction of at most two literals, and we are looking for an assignment σ to the variables in F that satisfies the largest number of clauses in F. For an assignment σ to F, we denote by $|\sigma|$ the number of clauses satisfied by σ. Let $Opt(F)$ be the largest $|\sigma|$ among all assignments σ to F.

Theorem 5.1 *If the* MAX-2SAT *problem has a polynomial time approximation algorithm of ratio* r*, then the* VC-PM *problem has a polynomial time approximation algorithm of ratio* $1+(r-1)(2m+n)/(rn)$ *(n and m are the number of vertices and the number of edges in the graph, respectively).*

PROOF. Let $G = (V, E)$ be an instance of the VC-PM problem, where G has n vertices and m edges, and let M be a maximum matching in G. Define an instance F_G of the MAX-2SAT problem as follows:

$$F_G = \bigcup_{[u,v]\in M} \{(x_u \vee x_v), (\bar{x}_u \vee \bar{x}_v)\} \cup \bigcup_{[u,v]\in E-M} \{(x_u \vee x_v)\}.$$

The set F_G consists of $m + n/2$ clauses on the Boolean variable set $X_G = \{x_u \mid u \in V\}$.

Suppose that C is a minimum vertex cover for the graph G. Since C contains at least one endpoint of each edge in the matching M, the edges in M can be classified into two sets M_1 and M_2 such that each edge in M_1 has exactly one endpoint in C and each edge in M_2 has both endpoints in C. Thus, $|C| = |M_1| + 2|M_2|$. Define an assignment σ_C for F_G such that $\sigma_C(x_u) =$ TRUE if and only if $u \in C$. Then σ_C satisfies all m clauses of the form $(x_u \vee x_v)$ in F_G. For each edge $[u, v]$ in M_1, σ_C satisfies both corresponding clauses $(x_u \vee x_v)$ and $(\bar{x}_u \vee \bar{x}_v)$, while for each edge $[u, v]$ in M_2, σ_C satisfies exactly one of the corresponding clauses $(x_u \vee x_v)$ and $(\bar{x}_u \vee \bar{x}_v)$ (i.e., the clause $(x_u \vee x_v)$). In conclusion, there are exactly $|M_2|$ clauses in F_G that are not satisfied by the assignment σ_C. Since the clause set F_G totally contains $m + n/2$ clauses, the number of clauses satisfied by the assignment σ_C is (note that $n/2 = |M| = |M_1| + |M_2|$)

$$\begin{aligned} m + n/2 - |M_2| &= m + n - (n/2 + |M_2|) = m + n - (|M_1| + 2|M_2|) \\ &= m + n - |C| = m + n - Opt(G). \end{aligned}$$

Thus, there is an assignment that satisfies $m + n - Opt(G)$ clauses in F_G, and $Opt(F_G) \geq m + n - Opt(G)$.

Now let σ be an assignment to the clause set F_G. We first modify the assignment σ as follows. If we have both $\sigma(x_u) =$ FALSE and $\sigma(x_v) =$ FALSE for an edge $[u, v] \in E$, then we modify σ by setting $\sigma(x_u) =$ TRUE, where u is arbitrarily picked from the two endpoints of the edge $[u, v]$. We claim that this change does not decrease the value $|\sigma|$. In fact, there is only one clause $(\bar{x}_u \vee \bar{x}_w)$ in F_G that contains the literal $\bar{x}_u$, where $[u, w]$ is an edge in the matching M. Therefore, converting from $\sigma(x_u) =$ FALSE to $\sigma(x_u) =$ TRUE can make at most one satisfied clause become unsatisfied. On the other hand, the unsatisfied clause $(x_u \vee x_v)$ becomes satisfied after this change. Therefore, the value $|\sigma|$ is not decreased.

Let σ' be the resulting assignment, then $|\sigma'| \geq |\sigma|$, and for each edge $[u, v] \in E$, σ' assigns at least one of the variables x_u and x_v the value TRUE. Now we let C_σ be a set of vertices in G as follows: a vertex u is in C_σ if and only if $\sigma'(x_u) =$ TRUE. Since for each edge $[u, v] \in E$, we have either $\sigma'(x_u) =$ TRUE or $\sigma'(x_v) =$ TRUE (or both), the set C_σ is a vertex cover for the graph G. By the construction of the assignment σ', the m clauses of the form $(x_u \vee x_v)$ in F_G are all satisfied by σ'. Therefore, exactly $|\sigma'| - m$ clauses of the form $(\bar{x}_u \vee \bar{x}_v)$ in F_G are satisfied by σ'. Since these $|\sigma'| - m$ clauses correspond to $|\sigma'| - m$ disjoint edges in the matching M, we conclude that the assignment σ' has assigned exactly $|\sigma'| - m$ Boolean variables in X_G the value FALSE. Thus, exactly $|\sigma'| - m$ vertices in G are not in C_σ and the vertex cover C_σ contains exactly $n - (|\sigma'| - m) = m + n - |\sigma'| \leq m + n - |\sigma|$ vertices. It follows that $Opt(G) \leq m + n - |\sigma|$, which gives $|\sigma| \leq m + n - Opt(G)$. Since σ was arbitrarily chosen, this shows that an optimal assignment satisfies at most $m + n - Opt(G)$ clauses in F_G, and $Opt(F_G) \leq m + n - Opt(G)$.

We conclude from the above discussion that $Opt(F_G) = m + n - Opt(G)$.

The above construction suggests a polynomial time approximation algorithm for the VC-PM problem via a polynomial time approximation algorithm for the Max-2Sat problem, as follows. Given an instance G of VC-PM, we first construct a perfect matching M in G and define F_G as above. We then use the polynomial time approximation algorithm for Max-2Sat to construct a truth assignment σ to F_G, and we return the vertex cover C_σ of G as defined above. Clearly, this is a polynomial time approximation algorithm for VC-PM.

If the approximation algorithm for Max-2Sat has ratio r, then we have $Opt(F_G)/|\sigma| \leq r$. According to the above discussion, the vertex cover C_σ contains at most $m + n - |\sigma| \leq m + n - Opt(F_G)/r$ vertices in the graph G. Noting that $Opt(F_G) \leq m + n/2$ we have:

$$\begin{aligned}
\frac{|C_\sigma|}{Opt(G)} &\leq \frac{m + n - Opt(F_G)/r}{m + n - Opt(F_G)} = \frac{r(m+n) - Opt(F_G)}{r(m + n - Opt(F_G))} \\
&= 1 + \frac{(r-1)Opt(F_G)}{r(m + n - Opt(F_G))} \leq 1 + \frac{(r-1)(m + n/2)}{r(m + n - (m + n/2))} \\
&= 1 + \frac{(r-1)(2m+n)}{rn}.
\end{aligned}$$

This shows that the polynomial time approximation algorithm for VC-PM has ratio $1 + (r - 1)(2m + n)/(rn)$, and proves the theorem. □

Based on the currently best approximation algorithm for the Max-2Sat problem, Theorem 5.1 describes an approximation algorithm for the VC-PM problem.

Corollary 5.2 *There is a polynomial time approximation algorithm* **VCPM-Apx** *of ratio* $1 + 0.069(2m + n)/n$ *for the* VC-PM *problem on graphs of* n *vertices and* m *edges.*

Proof. The polynomial time approximation algorithm **VCPM-Apx** for the VC-PM problem is obtained from the polynomial time approximation algorithm of ratio 1.074 for the Max-2Sat problem developed by Feige and Goemans [6], and the reduction described in Theorem 5.1. □

Corollary 5.2 gives the best approximation algorithm for the VC-PM problem on sparse graphs. Based on a greedy strategy, Halldórsson and Radhakrishnan [8] proposed an approximation algorithm of ratio $(2\bar{d} + 3)/5$ for the Independent Set problem on graphs of average degree $\bar{d}$. This result, plus the assumption that a minimum vertex cover for the graph G contains at least half of the vertices in G, gives us a polynomial time approximation algorithm of ratio $(4\bar{d}+1)/(2\bar{d}+3)$ for the Vertex Cover problem on graphs of average degree $\bar{d}$. This is also currently the best result for the VC-PM problem on graphs of average degree $\bar{d}$.

According to Corollary 5.2, our algorithm **VCPM-Apx** improves the above approximation ratio by Halldórsson and Radhakrishnan on the VC-PM problem when the average degree $\bar{d}$ is not larger than 10. The approximation ratios of our algorithm and Halldórsson and Radhakrishnan's are compared in the table in Figure 3.

Strictly speaking, neither of Halldórsson and Radhakrishnan's algorithm and the algorithm **VCPM-Apx** is applicable directly to general graphs of average degree $\bar{d}$. Halldórsson and Radhakrishnan's algorithm requires that the minimum vertex cover of the input graph contain at least half of the vertices in the graph, while our algorithm requires that the input graph have a perfect matching. Note that the condition in Halldórsson and Radhakrishnan's algorithm, namely that the minimum vertex cover should contain at least half of the vertices in the input graph, can be obtained by first applying the NT-Theorem to the input graph. Unfortunately, applying the NT-Theorem does not preserve the average degree of a graph.

Avg-degree	2	3	4	5	6	7	8	9	10
VCPM-Apx	1.207	1.276	1.345	1.414	1.483	1.552	1.621	1.690	1.759
H&R [8]	1.286	1.444	1.545	1.615	1.666	1.706	1.737	1.762	1.783

Figure 3: The approximation ratios of **VCPM-Apx** and the algorithm in [8]

A graph class for which both Halldórsson and Radhakrishnan's algorithm and the algorithms we developed in this paper are applicable, is the class of everywhere sparse graphs. Note that a graph of degree bounded by d is an everywhere $(d/2)$-sparse graph, and an everywhere k-sparse graph has average degree bounded by $2k$. According to Clementi and Trevisan [4], unless NP = co-RP, for each fixed $\epsilon > 0$, there is a constant k such that there is no polynomial time approximation algorithm of ratio $7/6 - \epsilon$ for the Vertex Cover problem on everywhere k-sparse graphs.

Theorem 5.3 *Halldórsson and Radhakrishnan's algorithm has approximation ratio* $(8k+1)/(4k+3)$ *for the* Vertex Cover *problem on everywhere k-sparse graphs.*

Proof. Given an everywhere k-sparse graph G, we first apply the NT-theorem to G to obtain the sets C_0, I_0, and V_0. Now the induced subgraph $G(V_0)$ is still everywhere k-sparse and satisfies $Opt(G(V_0)) \geq |V_0|/2$. Thus, Halldórsson and Radhakrishnan's algorithm can be directly applied to the graph $G(V_0)$ to get a vertex cover C_1 for $G(V_0)$. Since an everywhere k-sparse graph has average degree $\bar{d}$ bounded by $2k$, we have

$$\frac{|C_1|}{Opt(G(V_0))} \leq \frac{4\bar{d}+1}{2\bar{d}+3} \leq \frac{8k+1}{4k+3}.$$

Now the theorem follows from the NT-Theorem. □

Combining the algorithms **VC-Apx** and **VCPM-Apx** developed in the current paper, we have the following theorem.

Theorem 5.4 *There is a polynomial time approximation algorithm whose ratio is bounded by* $\max\{1.5, 0.092k + 1.379\}$ *for the* Vertex Cover *problem on everywhere k-sparse graphs.*

Proof. The NT-Theorem preserves everywhere k-sparseness, and thus we can assume that the input graph G of n vertices and m edges, which is everywhere k-sparse, satisfies $Opt(G) \geq n/2$.

Let M be a maximum matching in G, and let V_M be the set of matched vertices. Let n_M and m_M be the number of vertices and number of edges in the graph $G(V_M)$, respectively. The subgraph $G(V_M)$ is still everywhere k-sparse, and $m_M \leq kn_M$. Since the graph $G(V_M)$ has a perfect matching, we can use the algorithm **VCPM-Apx** in Corollary 5.2 to construct a vertex cover for $G(V_M)$. According to Corollary 5.2, the ratio r of **VCPM-Apx** satisfies:

$$r \leq 1 + \frac{0.069(2m_M + n_M)}{n_M} \leq 1.069 + 0.138k.$$

Now the theorem follows directly from Theorem 4.2. □

For certain values of k, the approximation ratio in Theorem 5.4 is better than the one in Theorem 5.3 derived from Halldórsson and Radhakrishnan's algorithm [8]. For example, for $k = 2.5$, 3, and 3.5, Theorem 5.3 gives ratios 1.615, 1.667, and 1.705, respectively, while Theorem 5.4 gives ratios 1.609, 1.655, and 1.701, respectively.

References

[1] R. Bar-Yehuda and S. Even, A local-ratio theorem for approximating the weighted vertex cover problem, *Annals of Discrete Mathematics* **25**, (1985), pp. 27-46.

[2] P. Berman and T. Fujito, On approximation properties of the independent set problem for low degree graphs, *Theory Comput. Systems* **32**, (1999), pp. 115-132.

[3] M. Chlebík and J. Chlebíková, On approximation hardness of the minimum 2SAT-DELETION problem, *Proceedings of the 29th International Symposium on Mathematical Foundations of Computer Science, Lecture Notes in Computer Science* **3153**, (2004), pp. 263-273.

[4] A. Clementi and L. Trevisan, Improved non-approximability results for minimum vertex cover with density constraints, *Theoretical Computer Science* **225**, (1999), pp. 113-128.

[5] I. Dinur and S. Safra, The importance of being biased, *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, (2002), pp. 33-42.

[6] U. Feige and M. Goemans, Approximating the value of two prover proof system, with applications to MAX-2SAT and MAX-DCUT, *Proceedings of the 3rd Israel Symposium on Theory of Computing and Systems*, (1995), pp. 182-189.

[7] M. Halldórsson, Approximating discrete collections via local improvements, *Proceedings of the 6th Annual ACM-SIAM Symposium on Discrete Algorithms*, (1995), pp. 160-169.

[8] M. Halldórsson and J. Radhakrishnan, Greed is good: approximating independent sets in sparse and bounded-degree graphs, *Algorithmica* **18**, (1997), pp. 145-163.

[9] J. Håstad, Some optimal inapproximability results, *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, (1997), pp. 1-10.

[10] D. Hochbaum, Efficient bounds for the stable set, vertex cover and set packing problems, *Discrete Applied Mathematics* **6**, (1983), pp. 243-254.

[11] D. Hochbaum, Approximating covering and packing problems: set cover, vertex cover, independent set, and related problems, In: Hochbaum, D. (ed.): *Approximation Algorithms for NP-hard Problems*, PWS Publishing Company, Boston, (1997), pp. 94-143.

[12] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-completeness*, Freeman, San Francisco, 1979.

[13] H. Karloff and U. Zwick, A 7/8-approximation algorithm for MAX-3SAT? *Proceedings of the 30th IEEE Symposium on the Foundation of Computer Science*, (1997), pp. 406-415.

[14] S. Micali and V. Vazirani, An $O(\sqrt{|V|} \cdot |E|)$ algorithm for finding maximum matching in general graphs, *Proceedings of the 21st IEEE Symposium on the Foundation of Computer Science*, (1980), pp. 17-27.

[15] B. Monien and E. Speckenmeyer, Ramsey numbers and an approximation algorithm for the vertex cover problem, *Acta Informatica* **22**, (1985), pp. 115-123.

[16] G. L. Nemhauser and L. E. Trotter, Vertex packing: structural properties and algorithms, *Mathematical Programming* **8**, (1975), pp. 232-248.