Inapproximability of the Minimum Biclique Edge Partition Problem

Hideaki OTSUKI†† and Tomio HIRATA††, Members

SUMMARY For a graph $G$, a biclique edge partition $S_{BP}(G)$ is a collection of bicliques (complete bipartite subgraphs) $B_i$ such that each edge of $G$ is contained in exactly one $B_i$. The Minimum Biclique Edge Partition Problem (MBEPP) asks for $S_{BP}(G)$ with the minimum size. It is known that MBEPP is NP-hard [1].

The Minimum Biclique Cover Problem (MBCP) is a graph covering problem that is NP-hard and its inapproximability has been investigated. A biclique cover of a graph $G$, $S_{BC}(G)$, is a collection of biclique subgraphs $B_i$ such that each edge of $G$ is contained in some $B_i$. MBCP asks for $S_{BC}(G)$ with the minimum size. Unless $P = NP$, MBCP does not have $O(n^{1/3})$-approximation algorithm [2], where $n$ is the number of vertices of $G$.

While MBCP was studied well, MBEPP has not been given attention. To the best of our knowledge, no lower bound for approximation of MBEPP is known. In this paper, we construct a gap preserving reduction [3] from Max E2-SAT to MBEPP, and show for arbitrary small $\epsilon > 0$, $(6053/6052 - \epsilon)$-approximation of MBEPP is NP-hard.

Note that “c-approximation of an optimization problem is NP-hard” means that if there exists a polynomial-time algorithm guaranteeing the output size within $c$ times the size of the optimal solution then $P = NP$.

1. Construction of an Instance of MBEPP

A Boolean expression $\varphi$ is in the Conjunctive Normal Form (CNF) if $\varphi$ is a conjunction of clauses and each clause is a disjunction of literals. For a given $\varphi$ in CNF, the Maximum Satisfiability Problem (MAX SAT) asks for an assignment that satisfies simultaneously the maximum number of clauses of $\varphi$. MAX 2-SAT is MAX SAT in which each clause has at most two literals. MAX E2-SAT is MAX 2-SAT in which each clause has exactly two literals of different variables. $k$-OCC-MAX 2-SAT ($k$-OCC-MAX E2-SAT) is MAX 2-SAT (MAX E2-SAT) in which each variable occurs exactly $k$ times in the expression.

Let $N$ be a positive integer, Berman and Karpinski [4] showed inapproximability of $3$-OCC-MAX 2-SAT as follows.

Theorem 2.1 ([5]): For any $\epsilon \in (0, 1/2)$, it is NP-hard to decide whether an instance of $3$-OCC-MAX 2-SAT with $2016N$ clauses has a truth assignment that satisfies at least $(2012 - \epsilon)N$ clauses, or at most $(2011 + \epsilon)N$.

In their proof, all clauses of an instance of $3$-OCC-MAX 2-SAT have exactly two literals [4]. So this theorem can be applied to $3$-OCC-MAX E2-SAT.

Let $\varphi$ be an instance of $3$-OCC-MAX E2-SAT and let $\sigma(\varphi)$ be the maximum number of clauses that can be satisfied simultaneously by an assignment. In this paper, we transform $\varphi$ into an instance $G = (V, E)$ of MBEPP as follows.

Suppose $\varphi$ has $n$ variables $x_i$ ($i = 1, \ldots, n$) and $m$ clauses $c_j$ ($j = 1, \ldots, m$). For $c_j = \alpha \vee \beta$, we call $\alpha$ ($\beta$) as the first (second) literal of $c_j$. Since each variable occurs exactly three times in $\varphi$, $3n = 2m$ holds. For each variable $x_i$, we construct $G_i$ as follows.

\[
V(G_i) = \{x_i^1, x_i^2, x_i^3, \bar{x}_i^1, \bar{x}_i^2, \bar{x}_i^3\}
\]
\[
E(G_i) = \{(x_i^1, \bar{x}_i^2), (x_i^2, \bar{x}_i^1), (x_i^3, \bar{x}_i^1), (\bar{x}_i^1, x_i^2), (\bar{x}_i^2, x_i^3), (\bar{x}_i^3, x_i^1)\}.
\]

Each $G_i$ is a cycle graph $C_6$ (Fig. 1 (a)). We denote by $V_x$ the set of vertices in these cycles, that is, $V_x = \{x_i^d, \bar{x}_i^d|1 \leq i \leq n, d = 1, 2, 3\}$. For each clause $c_j$, we create two vertices $y_j, z_j$ and an edge $e_j = (y_j, z_j)$. Let $V_{y_j} = \{y_j, z_j|1 \leq j \leq m\}$.

For each $j = 1, \ldots, m$, we add edges as follows. We connect a vertex of $V_x$ and a vertex of $V_{y_j}$ by these edges. Let $x_i$ be a variable and suppose it appears in three clauses $c_{i_1}, c_{i_2}, c_{i_3}$. If $d = 1, 2, 3$, if the occurrence of $x_i$ is the first literal of $c_{i_d}$, we connect $y_{i_d}$ to either $x_i^d$ (if the literal is $x_i$) or $\bar{x}_i^d$ (if the literal is $\bar{x}_i$) by an edge. If the occurrence of $x_i$ is the second literal of $c_{i_d}$, we connect $z_{i_d}$ to either $x_i^d$ (if the literal is $x_i$) or $\bar{x}_i^d$ (if the literal is $\bar{x}_i$) by an edge. We denote by $e_{i_d}(e_{i_d})$ the added edge incident to $y_{i_d}(z_{i_d})$.

Note that if $x_i$ occurs all positive (all negative) in $\varphi$,
We denote two sets of graphs as follows.

\[ G = \{ G_1, G_2, \ldots, G_n \} \]

\[ \bar{G} = \{ \bar{G}_1, \bar{G}_2, \ldots, \bar{G}_n \} \]

The degree-three vertices appearing in \( G_i \) are not adjacent each other (Fig. 1 (b)). Otherwise, the degree-three vertices appear contiguously in \( G_i \) (Fig. 1 (c)). We summarize the construction of \( G \) as follows.

\[ V(G) = V_x \cup V_c \]

\[ E(G) = \bigcup_{i=1}^{n} E(G_i) \cup \bigcup_{j=1}^{m} E(G_{c_j}). \]

The following lemma holds.

**Lemma 2.2:** All biclique subgraphs of \( G \) are star graphs \( K_{1,t} \) (\( s \geq 1 \)).

**Proof:** \( G \) is constructed by cycle graphs \( C_6 \) connected each other by a path graph \( P_4 \). Thus, \( G \) has no \( K_{s,t} \) (\( s \geq 2 \), \( t \geq 2 \)) as a subgraph.

\[ \square \]

3. Inapproximability of MBEPP

In the sequel, we use “biclique” and “star graph” interchangeably. For a star graph \( K_{1,t} \) (\( s \geq 2 \)), the vertex of degree \( s \) is called its center. For each \( i \), we define six star graphs \( (K_{1,t}) \) as follows.

\[ S_i^1 = (\{x_i^1, x_i^2\}, (x_i^1, x_i^2)) \]

\[ S_i^2 = (\{x_i^1, x_i^2\}, (x_i^2, x_i^3)) \]

\[ S_i^3 = (\{x_i^1, x_i^2\}, (x_i^3, x_i^4)) \]

\[ \bar{S}_i^1 = (\{x_i^1, x_i^2\}, (\bar{x}_i^1, \bar{x}_i^2)) \]

\[ \bar{S}_i^2 = (\{x_i^1, x_i^2\}, (\bar{x}_i^2, x_i^3)) \]

\[ \bar{S}_i^3 = (\{x_i^1, x_i^2\}, (\bar{x}_i^3, x_i^4)). \]

We denote two sets of graphs as follows.

\[ S_i^T = \{ S_i^1, S_i^2, \bar{S}_i^3 \} \]

\[ S_i^F = \{ S_i^1, \bar{S}_i^2, S_i^3 \} \]

Each \( G_i \) (\( i = 1, \ldots, n \)) can be partitioned into the three bicliques of \( S_i^T \) or into the three bicliques of \( S_i^F \).

We denote by \( S(G) \) the size of an optimal solution of MBEPP for \( G \). We give the following lemma.

**Lemma 3.1:** If \( s(\varphi) > (1 - \epsilon)m \) then \( S(G) < (3 + \epsilon)m \).

**Proof:** Let \( \pi \) be an assignment that satisfies more than \((1 - \epsilon)m \) clauses of \( \varphi \). We show that \( \pi \) induces a solution, \( SOL'(G) \), of MBEPP that satisfies \( |SOL'(G)| < (3 + \epsilon)m \).

Let \( SOL'(G) \) be an empty set. For each \( x_i \), if \( \pi \) assigns TRUE (FALSE) to \( x_i \), we add \( S_i^T \) (\( S_i^F \)) to \( SOL'(G) \). Note that all edges of \( G_i \) have been partitioned by these \( 3m \) (\( 2m \)) bicliques.

Let \( c_j \) be an arbitrary clause of \( \varphi \). If the assignment \( \pi \) satisfies \( c_j \), there is at least one star graph \( K_{1,2} \) in \( SOL'(G) \) whose center is adjacent to \( y_j \) or \( z_j \). W.l.o.g., we assume that \( y_j \) is adjacent to the center of this \( K_{1,2} \). We replace this star graph \( K_{1,2} \) in \( SOL'(G) \) with a star graph \( K_{1,3} \) by adding \( e_{y_j} \). This manipulation does not increase \(|SOL'(G)|\). Furthermore, we add to \( SOL'(G) \) a star graph \( K_{1,2} \) consisting of \( e_j \) and \( e_{z_j} \).

If the assignment \( \pi \) does not satisfy \( c_j \), we add two star graphs to \( SOL'(G) \): \( K_{1,2} \) consisting of \( e_j \) and \( e_{y_j} \), and \( K_{1,1} \) (\( = e_{z_j} \)). The number of \( K_{1,1} \) in \( SOL'(G) \) is less than \( em \) because of the assumption. In \( SOL'(G) \), we have \( 2m \) star graphs, \( K_{1,2} \) or \( K_{1,3} \), whose centers are in \( V_x \), and \( m \) star graphs, \( K_{1,2} \), that have an edge \( e_j \). Thus, we have \(|SOL'(G)| < 2m + (1 - \epsilon)m + 2em = (3 + \epsilon)m \).

**Lemma 3.2:** If \( s(\varphi) \leq (1 - \epsilon)m \) then \( S(G) \geq (3 + \epsilon)m \).

**Proof:** We assume \( SOL(G) \) is a solution of MBEPP and \(|SOL(G)| < (3 + \epsilon)m \). We will show that there is an assignment that satisfies more than \((1 - \epsilon)m \) clauses of \( \varphi \). We construct a solution \( SOL'(G) \) that satisfies \(|SOL'(G)| \leq |SOL(G)| \) and then we show \( SOL'(G) \) induces an assignment that satisfies more than \((1 - \epsilon)m \) clauses of \( \varphi \).

Let \( SOL'(G) \) be an empty set. We denote by \( SC(G) \) the set of all bicliques in \( SOL(G) \) that have an edge \( e_j \) (\( j = 1, \ldots, m \)). Then \( |SC(G)| = m \). We add all bicliques in \( SC(G) \) to \( SOL'(G) \).

Next, we remove all edges of bicliques in \( SC(G) \) from \( G \). If there are singletons in the resulted graph, we remove all of them. Let \( G' \) be the resulted graph. \( G' \) consists of \( n \) connected components. Each of the connected components is \( G_i \) possibly with its incident edges. Note that for all \( j (= 1, \ldots, m) \), at least one edge \( e_{y_j} \) or \( e_{z_j} \) remains in \( G' \).

For each \( i (= 1, \ldots, m) \), we denote by \( G'_i \) a connected component of \( G' \) whose \( C_6 \) subgraph is \( G_i \). We denote by \( \mathcal{A} \) the set of all \( G'_i \) that has no contiguous degree-three vertices, and we denote by \( \mathcal{B} \) the set of all \( G'_i \) that has some contiguous degree-three vertices.

It is clear that each \( G'_i \in \mathcal{A} \) cannot be partitioned into less than three bicliques. For each \( G'_i \in \mathcal{A} \), we add three bicliques as shown in Fig. 2 (a), to \( SOL'(G) \) as follows. If some of \( x_i^d \) (\( d \in \{1, 2, 3\} \)) are the degree-three vertices, we add three bicliques (star graphs) whose centers are \( x_i^d \) to \( SOL'(G) \). Otherwise, we add three bicliques (star graphs) whose centers are \( x_i^d \) to \( SOL'(G) \).

It is clear that each \( G'_i \in \mathcal{B} \) cannot be partitioned into less than four bicliques. For each \( G'_i \in \mathcal{B} \), we add four bicliques as follows. If there are three degree-three vertices in \( G'_i \), we denote these contiguous vertices by \( v_1, v_2, v_3 \) in this order as shown in Fig. 1 (c). We add to \( SOL'(G) \) four bicliques; one star graph \( K_{1,1} \) that is an edge connecting \( v_2 \) and a vertex of \( e_j \), two star graphs \( K_{1,3} \) whose center vertices are \( v_1 \) and \( v_3 \) and one star graph \( K_{1,2} \) for the remaining part (Fig. 2 (b)).
If there are only two degree-three vertices in \( G' \), we denote these two contiguous vertices by \( v_1 \) and \( v_2 \). Then we add to \( \text{SOL}'(G) \) four bicliques; one star graph \( K_{1,1} \) that is an edge connecting \( v_2 \) and a vertex in \( e_j \), one star graph \( K_{1,3} \) whose center vertex is \( v_1 \) and two star graphs \( K_{1,2} \) for the remaining part.

\( \text{SOL}'(G) \) has the same subset \( \text{SC}(G) \) of \( \text{SOL}(G) \), and the remaining part is partitioned into the optimal number of bicliques. So it is clear that \( \text{SOL}'(G) \) is a biclique partition of \( G \) and \( |\text{SOL}'(G)| \leq |\text{SOL}(G)| \) holds.

Let \( |B| = \epsilon m \), then \( |A| = n - \epsilon m \) and

\[
|\text{SOL}'(G)| = |\text{SC}(G)| + 3|A| + 4|B| = (3 + \epsilon')m.
\]

From the assumption \( |\text{SOL}(G)| < (3 + \epsilon)m \), we have \( (3 + \epsilon')m < (3 + \epsilon)m \), and \( |B| < \epsilon m \) holds.

We induce an assignment \( \pi' \) from \( \text{SOL}'(G) \) as follows. For each \( G'_i \in \mathcal{A} \), if some of \( x'_i (\bar{x}'_i) \) are degree-three vertices, we assign TRUE (FALSE) to \( x_i \). If there is no degree-three vertex in \( G'_i \), we assign FALSE to \( x_i \). For each \( G'_i \in \mathcal{B} \), if the degree-three vertex \( v_1 \) is \( x'_i (\bar{x}'_i) \) for some \( d \in \{1, 2, 3\} \), we assign TRUE (FALSE) to \( x_i \).

Note that under this assignment \( \pi' \) the literals associating degree-three vertices denoted by \( v_2 \) are FALSE and the other literals are TRUE. Therefore, if \( c_j \) is not satisfied by \( \pi' \), at least one endpoint of \( e_j \) must be adjacent to \( v_2 \) in some \( G'_i \in \mathcal{B} \). The number of vertices denoted by \( v_2 \) in \( G \) is exactly the size of \( \mathcal{B} \). Since \( |\mathcal{B}| < \epsilon m \), the number of clauses not satisfied by \( \pi' \) is less than \( \epsilon m \), and thus \( \pi' \) satisfies more than \( (1 - \epsilon)m \) clauses in \( \varphi \).

\[
\text{Theorem 3.3:} \quad (6053/6052 - \epsilon)\text{-approximation of MBEPP is NP-hard, for arbitrary small } \epsilon > 0.
\]

**Proof:** From Theorem 2.1, it is NP-hard to decide whether \( s(\varphi) > (2016N - 4N - \epsilon N) \) or \( s(\varphi) \leq (2016N - 5N + \epsilon N) \). Let \( m = 2016N, \epsilon_1 m = (4 + \epsilon)N, \epsilon_2 m = (5 - \epsilon)N \). From Lemma 3.1, if \( s(\varphi) > (1 - \epsilon_1)m \) then \( S(G) < (3 + \epsilon_1)m = 3 \cdot 2016N + (4 + \epsilon)N \). From Lemma 3.2, if \( s(\varphi) \leq (1 - \epsilon_2)m \) then \( S(G) \geq (3 + \epsilon_2)m = 3 \cdot 2016N + (5 - \epsilon)N \). Therefore, for any \( \epsilon \), it is NP hard to decide whether \( S(G) < (6052 + \epsilon)N \) or \( S(G) \geq (6053 - \epsilon)N \).

**References**


