On the inapproximability of maximum intersection problems

Min-Zheng Shieha, Shi-Chun Tsai b,∗, Ming-Chuan Yangb

a Information and Communications Technology Labs, National Chiao Tung University, Hsinchu, Taiwan
b Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan

Abstract

Given $u$ sets, we want to choose exactly $k$ sets such that the cardinality of their intersection is maximized. This is the so-called MAX-$k$-INTERSECT problem. We prove that MAX-$k$-INTERSECT cannot be approximated within an absolute error of $1/2n^{1-2\epsilon} + O(n^{1-3\epsilon})$ unless $P=NP$. This answers an open question about its hardness. We also give a correct proof of an inapproximable result by Clifford and Popa (2011) [3] by proving that MAX-INTERSECT problem is equivalent to the MAX-CLIQUE problem.

1. Introduction

The SET-COVER problem [1,6,14] is one of the well-known NP-complete problems. There are many related variants corresponding to different applications. For example, there are several closely related problems mentioned in the work of Clifford and Popa [3], such as hitting set [6], minimum sum set cover [4], maximum coverage [9], budgeted maximum coverage [10] and $k$-set cover [5]. Besides, in the area of privacy protection two similar disclosure control techniques called $k$-anonymity [2] and $k$-intersection [16] are also investigated. Nevertheless, even essentially equivalent problems may have different degree of hardness for approximation. In this paper we focus on two intersection problems.

To solve the MAX-INTERSECT problem, Clifford and Popa [3] proposed a concise reduction, but they didn’t apply Zuckerman’s theorem correctly. Our first result is to correct the flaw and to give a correct proof of their main result. The second result answers an open question raised in [3], i.e., we prove MAX-$k$-INTERSECT cannot be approximated within an absolute error of $1/2n^{1-2\epsilon} + O(n^{1-3\epsilon})$ unless $P=NP$.

Formally, we define the problems as follows.

Problem 1 (MAX-INTERSECT). (See [3].) Given $u$ sets $A_1,\ldots,A_u$ over a finite universe $U=\{1,\ldots,n\}$, the goal is to select exactly one set from each of $A_1,\ldots,A_u$ in order to maximize the size of the intersection of the sets.

Problem 2 (MAX-$k$-INTERSECT). (See [3].) Given $u$ sets $A_1,\ldots,A_u$ over a finite universe $U=\{1,\ldots,n\}$ and an integer $k \leq u$, where each $A_i$ is a set of subsets in a universe $U=\{1,\ldots,n\}$, the goal is to select exactly one set from each of $A_1,\ldots,A_u$ in order to maximize the size of the intersection of the sets.
of all \(A_1, \ldots, A_u\) are bounded by a constant. However, for general case, the greedy strategy is not efficient.

We will use the following problem to investigate the hardness of Problems 1 and 2.

**Definition 1** (MAX-Clique). Given an undirected simple graph, the goal is to find a subset of vertices with maximum cardinality such that nodes in this subset are pairwise adjacent.

The MAX-INTERSECT problem can be used to solve a typical production line problem [3]. Let the universe \([1, \ldots, n]\) be the collection of different types of devices produced by machines \(A_1, \ldots, A_u\). There are \(u\) production stages and one machine is responsible for one stage. Moreover, each machine has a finite set of settings, which involve in some types of devices. The goal is to maximize the total number of the device types by selecting a setting from each machine.

The MAX-\(k\)-INTERSECT problem can be used as a mathematical model of the disclosure control problem [16]. Let \(A_1, \ldots, A_u\) be \(u\) individuals, and each person has some of \(n\) attributes. In order to ensure that the disclosed data cannot be used to identify any individual, it is only allowed to reveal the attributes owned by at least \(k\) persons, where \(k\) is large enough to make sure the privacy-preserving. Now we want to know the maximum set of attributes that are owned by any combination of \(k\) individuals. Note \(k\)-intersection is similar to but different from the method of \(k\)-anonymity [16]. The MAX-\(k\)-INTERSECT problem can also be formulated in the following setting. Consider a production line that is restricted to operate with exactly \(k\) machines because of resource constraints such as electrical power and working capital. Let \(A_1, \ldots, A_u\) be different machines, and each is associated with some production items in the universe \([1, \ldots, n]\). The goal is to find a set of \(k\) machines which can maximize the number of produced items (i.e., the cardinality of the intersection).

Recently, Xavier [17] proved another inapproximability result of MAX-\(k\)-INTERSECT: suppose \(\mathsf{NP} \neq \mathsf{BPTIME}(2^{\epsilon n})\) for a small constant \(\epsilon > 0\), then MAX-\(k\)-INTERSECT cannot have a polynomial time \((\epsilon n)^k\)-approximation algorithm, where \(n\) is the instance size and \(\epsilon\) depends only on \(\epsilon\). Note that \(\mathsf{NP} \not\subseteq \mathsf{BPTIME}(2^{\epsilon n})\) implies \(\mathsf{P} \neq \mathsf{NP}\), hence the assumption in [17] is much stronger. With the stronger assumption, their inapproximable gap [17] is larger than ours (Theorem 7).

Our main results are: (1) give a correct proof for the result claimed by Clifford and Popa [3] by showing that the hardness of approximation of MAX-INTERSECT and MAX-Clique are the same (Lemma 2); (2) it is \(\mathsf{NP}\)-hard to approximate MAX-\(k\)-INTERSECT within an absolute error of \(2^n^{1-2\epsilon} + 3n^{1-3\epsilon} - 1\). This paper is organized as follows. In Section 2, we introduce some notations and definitions. Section 3 shows the inapproximability results. Section 4 concludes the paper.

### 2. Preliminaries

Let \(G = (V, E)\) be an undirected simple graph, where \(V = V(G) = \{1, 2, \ldots, n\}\) is the set of vertices and the edge set \(E = E(G)\) is a subset of \([i, j] : i, j \in V\). For convenience, we denote \([1, 2, \ldots, n]\) as \([n]\). \(N(i)\) indicates the neighbor set of the vertex \(i \in V\), that is, \(N(i) = \{(i, j) : (i, j) \in E\}\). The cardinality of a set \(X\) is denoted as \(|X|\). Let \(\Pi\) be an optimization problem and \(\mathsf{OPT}\Pi\) denote the optimal solution set of \(\Pi\). Furthermore, if \(x\) is an instance of the problem \(\Pi\), then \(\mathsf{OPT}(x)\) means the corresponding optimal solution of \(x\). More precisely, we use the notation \(\mathsf{OPT}(\Pi(x))\). The measure of solutions used in this paper is the cardinality of a set, so we directly denote \([\mathsf{OPT}(\Pi)\] or \(|\mathsf{OPT}(x)|\) for the optimization problem \(\Pi\) or the instance \(x\), respectively. For example, if \(G\) is an instance of MAX-Clique, then \(\mathsf{OPT}(G)\) is a maximum clique in \(G\) and \(|\mathsf{OPT}(G)|\) is the maximum clique size of \(G\).

**Definition 2** (Absolute error). (See [1].) Given an optimization problem \(\Pi\), for any instance \(x\) and for any feasible solution \(y\) of \(x\), the absolute error of \(y\) with respect to \(x\) is defined as

\[
\delta(x, y) = |m^*(x) - m(x, y)|
\]

where \(m^*(x)\) denotes the measure of an optimal solution of instance \(x\) and \(m(x, y)\) denotes the measure of solution \(y\).

We say that an approximation algorithm \(A\) for an optimization problem \(\Pi\) is an absolute approximation algorithm if there exists a constant \(K\) such that, for any instance \(x\) of \(\Pi\), \(\delta(x, A(x)) \leq K\).

**Definition 3** (\(r\)-Approximate algorithm). (See [1].) Given an optimization problem \(\Pi\) and an approximation algorithm \(A\) for \(\Pi\), define the performance ratio of \(A\) as

\[
\rho(x, A(x)) = \max \left(\frac{|\mathsf{OPT}(x)|}{|A(x)|}, \frac{|A(x)|}{|\mathsf{OPT}(x)|}\right)
\]

We say that \(A\) is a \(r\)-approximation algorithm for \(\Pi\) if, given any input instance \(x\) of \(\Pi\), the performance ratio \(\rho(x, A(x))\) of the approximation solution \(|A(x)|\) is bounded by \(r\), that is,

\[
\rho(x, A(x)) \leq r.
\]

Note that \(r \geq 1\), and equivalently we have that \(|A(x)| \geq \frac{1}{r} \cdot |\mathsf{OPT}(x)|\) if \(\Pi\) is a maximization problem.

**Definition 4** (Promise problem). (See [7].) A promise problem \(\Pi\) is a pair of non-intersecting sets, denoted \((\mathsf{Π}^\text{Yes}, \mathsf{Π}^\text{No})\); that is, \(\mathsf{Π}^\text{Yes}, \mathsf{Π}^\text{No} \subseteq \{0, 1\}^*\) and \(\mathsf{Π}^\text{Yes} \cap \mathsf{Π}^\text{No} = \emptyset\). The set \(\mathsf{Π}^\text{Yes} \cup \mathsf{Π}^\text{No}\) is called the promise.

An algorithm solves a promise problem if it distinguishes instances in \(\Pi^\text{Yes}\) from that in \(\Pi^\text{No}\).

**Definition 5** (Gap preserving reduction). (See [14].) Let \(\Pi_1\) and \(\Pi_2\) be some maximization problems. A gap preserving reduction from \(\Pi_1\) to \(\Pi_2\) comes with four parameters (functions) \(f_1, \alpha, f_2\) and \(\beta\). Given an instance \(x\) of \(\Pi_1\), the reduction computes in polynomial time an instance \(y\) of \(\Pi_2\) such that:
1. $|\text{OPT}_{\Pi_1}(x)| \geq f_1(x) \rightarrow |\text{OPT}_{\Pi_2}(y)| > f_2(y)$,
2. $|\text{OPT}_{\Pi_1}(x)| \leq \alpha(x) f_1(x) \rightarrow |\text{OPT}_{\Pi_2}(y)| \leq \beta(y) f_2(y)$.

Note the gaps $1/\alpha > 1$ and $1/\beta > 1$. Moreover, there are three other similar definitions.

For a maximization problem $\Pi$, let $\Pi_{\leq f}$ and $\Pi_{> f}$ be the languages of $\{x: |\text{OPT}(x)| \leq f(x)\}$ and $\{x: |\text{OPT}(x)| > f(x)\}$, respectively. A gap preserving reduction can be interpreted as a reduction which maps a promise problem $(\Pi_{\geq f_1}, \Pi_{\leq f_1})$ to another promise problem $(\Pi_{\geq f_2}, \Pi_{\leq f_2})$. Let us see how this reduction works. Observe that if $\Pi_2$ has a polynomial time algorithm $A_{\Pi_2}$ whose approximating factor is better than the gap $1/\beta$ (i.e. $\text{OPT}_{\Pi_2}/A_{\Pi_2} < 1/\beta$), then $A_{\Pi_2}$ solves the promise problem $(\Pi_{\geq f_2}, \Pi_{\leq f_2})$. Moreover, since $(\Pi_{\geq f_1}, \Pi_{\leq f_1})$ can be reduced to $(\Pi_{\geq f_2}, \Pi_{\leq f_2})$ efficiently, there is a polynomial time algorithm solving the promise problem $(\Pi_{\geq f_1}, \Pi_{\leq f_1})$. Conversely, if $(\Pi_{\geq f_1}, \Pi_{\leq f_1})$ is $\text{NP}$-hard, then so is $(\Pi_{\geq f_2}, \Pi_{\leq f_2})$.

The following inapproximable gap was first shown in Håstad's work [8] under the assumption $\text{NP} \neq \text{ZPP}$. Then Zuckerman derandomized the reduction and proved the same gap under a weaker assumption $\text{P} \neq \text{NP}$.

**Theorem 1.** (See Zuckerman [18],) MAX-CLIQUE does not have a polynomial time $(n^{1-\epsilon})$-approximation for any $\epsilon > 0$, unless $\text{P} = \text{NP}$.

Take a closer look at Zuckerman's theorem. A critical step in his proof states that for any $\epsilon > 0$ it is $\text{NP}$-hard to distinguish the instance class with clique size at least $2^R$ from the class with clique size at most $2^{1+\epsilon}R$ vertices. Let $2^{1+\epsilon}R = n$ and $2^R = n^\xi$, then the above statement is equivalent to that the promise problem $(\Pi_{\geq n^{-1+\xi}}, \Pi_{\leq n^{\xi}})$ is $\text{NP}$-hard, where $\Pi = \text{MAX-CLIQUE}$. Hence, no polynomial time algorithm can guarantee a performance factor of $n^{1-\epsilon}/n^\xi = n^{1-2\xi}$ unless $\text{P} = \text{NP}$. Replacing $\xi$ with $\epsilon/2$ leads to the conclusion.

Note that $\epsilon$ (and hence $\xi$) is fixed positive number, although it can be arbitrarily small. Otherwise, for example, let $\xi = 1/n$ such that the promise problem $(\Pi_{\geq n^{1-\xi}}, \Pi_{\leq n^{\xi}})$ is equivalent to $(\Pi_{\geq n^{-1}}, \Pi_{\leq n^0})$. Then, it would imply that the promise MAX-CLIQUE problem $(\Pi_{\geq n^{-1}}, \Pi_{\leq n^0})$ is $\text{NP}$-hard, which is obviously not true. Besides, not all instances of an $\text{NP}$-hard problem $(\Pi_{\geq n^{-1}}, \Pi_{\leq n^0})$ are intractable.

**3. Inapproximability results**

3.1. **Hardness of MAX-INTERSECT**

Let $\Pi = \text{MAX-CLIQUE}$ and $\Phi = \text{MAX-INTERSECT}$. Clifford and Popa [3] proposed a gap preserving reduction from the promised problem $(\Pi_{\geq n^0}, \Pi_{\leq n^{1-\epsilon}})$ to $(\Phi_{\geq n^0}, \Phi_{\leq n^{1-\epsilon}})$. However, it is insufficient to prove their claimed inapproximable result. The reason is that Zuckerman's theorem is a worst-case statement and not all promise problems with a gap less than $n^{1-\epsilon}$ are $\text{NP}$-hard. In particular, for a simple graph with $n$ vertices, it takes only $O(n^2)$ time to distinguish the class of $n$-clique from the others. It is clear that the promised MAX-CLIQUE problem $(\Pi_{\geq n^0}, \Pi_{\leq n^{1-\epsilon}})$ is in $\text{P}$. Even though their reduction is valid, it does not imply the hardness of the promised MAX-INTERSECT problem $(\Phi_{\geq n^0}, \Phi_{\leq n^{1-\epsilon}})$. In fact, it is also in $\text{P}$ to distinguish the case $|\text{OPT}_{\text{MAX-INTERSECT}}| = n$ from the case $|\text{OPT}_{\text{MAX-INTERSECT}}| < n^{1-\epsilon}$. Besides, the annotations of Definition 5 and Theorem 1 show that the gap between instance classes plays an important role. In order to apply the inapproximability of MAX-CLIQUE, the gap of classes to be distinguished should be $n^{1-\epsilon}$. The gap in [3] was mistaken for $n^{1-\epsilon} = n^\epsilon$.

We show the inapproximability result of MAX-INTERSECT by fixing the mistakes in their proof. Actually, we prove a stronger statement (Lemma 2). The reduction $f_\epsilon$ from MAX-CLIQUE to MAX-INTERSECT is defined as: for a given graph $G = (V, E)$ with $V = \{1, \ldots, n\}$, let $f_\epsilon(G)$ be a family of sets $A_1, A_2, \ldots, A_n$, where $A_i := \{N(i) \cup [i], V \setminus [i]\}$ for $i \in [n]$. It is easy to check that the mapping $f_\epsilon$ is a polynomial time reduction.

**Lemma 2.** Let $G$ be an instance of MAX-CLIQUE and $f_\epsilon(G)$ be the corresponding instance of MAX-INTERSECT, then

$$|\text{OPT}(G)| = |\text{OPT}(f_\epsilon(G))|.$$ 

**Proof.** By the reduction, we have to select either $N(i) \cup [i]$ or $V \setminus [i]$ from every $A_i$ to maximize their intersection size. We first prove that $|\text{OPT}(G)| \leq |\text{OPT}(f_\epsilon(G))|$, where $G$ is a maximum clique of $G$. Let $S_i = N(i) \cup [i]$ for $i \in \text{OPT}(G)$ and $S_i = V \setminus [i]$ for $i \notin \text{OPT}(G)$. Observe that for $i, j \in \text{OPT}(G)$ and $k \notin \text{OPT}(G)$, it is clear that $[i, j] \subseteq S_i$ and $[i, j] \subseteq S_j$. So for every $i \in V$, we have $\text{OPT}(G) \subseteq S_i$ and hence $\text{OPT}(G) \subseteq S_i \cap S_j \cap \cdots \cap S_n$. This concludes $|\text{OPT}(G)| \leq |S_i \cap S_j \cap \cdots \cap S_n| \leq |\text{OPT}(f_\epsilon(G))|$. Next we prove $|\text{OPT}(G)| \geq |\text{OPT}(f_\epsilon(G))|$. Assume for $i = 1, \ldots, n$, $S_i^* \subseteq A_i$ are the selected subsets which maximize the intersection cardinality. It means $|S_1^* \cap S_2^* \cap \cdots \cap S_n^*| = |\text{OPT}(f_\epsilon(G))|$. Denote $S_i^* \cap S_j^* \cap \cdots \cap S_n^*$ as $c^*(G)$ for short. We claim $c^*(G)$ is a clique. This is because if $i, j \in c^*(G)$ then immediately we know $i, j \in S_i^*$ and $i, j \in S_j^*$, which implies $S_i^* = N(i) \cup [i]$ and $S_j^* = N(j) \cup [j]$. Hence for all $[i, j] \subseteq c^*(G)$, $(i, j) \in E$, i.e., $c^*(G)$ must be a clique. Thus, we have $|c^*(G)| \leq |\text{OPT}(G)|$. By the definition of $c^*(G)$, we have that $|\text{OPT}(G)| \geq |c^*(G)| = |\text{OPT}(f_\epsilon(G))|$. □

Consequently, by Theorem 1 and the above lemma we obtain the result claimed in [3]:

**Theorem 3.** For any constant $\epsilon > 0$, the MAX-INTERSECT problem does not admit an $(n^{1-\epsilon})$-approximation unless $\text{P} = \text{NP}$.

**Proof.** If one can approximate this problem in polynomial time within $n^{1-\epsilon}$ factor, then MAX-CLIQUE can be approximated within a factor of $n^{1-\epsilon}$ by Lemma 2. It cannot be true unless $\text{P} = \text{NP}$. □
3.2. Hardness of MAX-k-INTERSECT

Next we consider the MAX-k-INTERSECT problem, which had already been proved to be \textbf{NP}-hard [15]. However, the original proof does not imply an inapproximability result. We give a new reduction that can be used to prove a non-trivial inapproximability result. The idea is inspired by a reduction from MAX-CLIQUE to Balanced Complete Bipartite Subgraph (BCBS) [6,12]. For an instance \( G = (V, E) \) of MAX-CLIQUE with \( V = \{1, 2, \ldots, n\} \) and \( E = \{e_1, e_2, \ldots, e_m\} \) define the universe \( U = V \), and for each \( e_j = (s_j, t_j) \in E \) define a corresponding subset \( A_j \) as \( U \setminus \{s_j, t_j\} \). Hence \( |A_j| = n - 2 \) for all \( j \in [m] \). With \( A_1, A_2, \ldots, A_m \), the goal of the MAX-k-INTERSECT problem is to select \( k \) sets from \( \{A_1, \ldots, A_m\} \), such that their intersection is maximized. Consider a positive constant \( \epsilon \) and let \( k = \left(\frac{n^{1-\epsilon}}{2}\right) \), where \( n = |V| \). The corresponding reduced instance is \( f_r(G) = (U, \{A_j\}_{j=1}^m, k) \). This mapping obviously can be done in polynomial time. To prove our inapproximability result, we will use a simple consequence of Turán’s theorem.

**Lemma 4.** (See [13].) If a simple graph \( G = (V, E) \) has no \((p + 1)\)-clique, then

\[
|E| \leq \frac{p - 1}{2p} |V|^2.
\]

Lemma 4 implies that if a graph has a small clique size, then the number of edges cannot be too large.

**Lemma 5.** Let \( G = (V, E) \) be an instance of MAX-CLIQUE and \( f_r(G) \) be the corresponding instance of MAX-k-INTERSECT defined above. Let \( |V| = n, |E| = m \) and a constant \( \epsilon > 0 \). Then for large enough \( n \), we have

1. \( \text{OPT}(G) \geq n^{1-\epsilon} \Rightarrow \text{OPT}(f_r(G)) \geq n - n^{1-\epsilon} \),
2. \( \text{OPT}(G) \leq n^{\epsilon} \Rightarrow \text{OPT}(f_r(G)) \leq n - (n^{1-\epsilon} + \frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1) \).

**Proof.** Note that we let \( k = \left(\frac{n^{1-\epsilon}}{2}\right) \). If \( \text{OPT}(G) \geq n^{1-\epsilon} \), then there is a complete subgraph \( C \subseteq G \) with \( n^{1-\epsilon} \) vertices. W.l.o.g. let \( C = (V', E') \), where \( V' = \{1, 2, \ldots, n^{1-\epsilon}\} \) and \( E' = \{e_1, \ldots, e_{\ell}\} \). According to the above reduction, we can select subsets \( A_1, A_2, \ldots, A_{\ell} \) which correspond to the clique edges \( e_1, \ldots, e_{\ell} \). Hence \( \bigcap_{j=1}^{\ell} A_j = U \setminus \{1, \ldots, n^{1-\epsilon}\} \), and we have \( \text{OPT}(f_r(G)) \geq |\bigcap_{j=1}^{\ell} A_j| = n - n^{1-\epsilon} \).

If \( \text{OPT}(G) \leq n^{\epsilon} \), then it implies that any \( k \)-edge induced subgraph of \( G \) does not contain an \( (n^{\epsilon} + 1) \)-clique. In order to estimate \( \text{OPT}(f_r(G)) \), we need to bound the minimum number of vertices associated with \( k \) edges. Suppose a \( k \)-edge and \( (n^{\epsilon} + 1) \)-clique free simple graph has \( x \) vertices. By Lemma 4, we have \( k \leq x^2(n^{\epsilon} - 1)/2n^{\epsilon} \). With the binomial series expansion (see Claim 1 later for details), we have

\[
x \geq (n^{1-\epsilon}) \left(1 + \frac{n^{\epsilon}}{n}\right)^{1/2} \left(1 + \frac{-1}{n^{\epsilon}}\right)^{-1/2}
\geq n^{1-\epsilon} + \frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1.
\]

Denote \( n^{1-\epsilon} + \frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1 \) as \( x^* \). By the above reduction, selecting exactly \( k \) sets from \( A_1, A_2, \ldots, A_m \) corresponds to selecting exactly \( k \) edges from the edge set \( E \). Since any \( k \) edges in such \( G \) associate with at least \( x^* \) vertices, the intersection of any \( k \) sets can have at most \( n - x^* \) elements. Hence \( \text{OPT}(f_r(G)) \leq n - x^* = n - (n^{1-\epsilon} + \frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1) \).

To avoid the vacuous cases in Lemma 5, we consider \( n \geq n_\epsilon \) only, where \( \epsilon \) is a proper fixed number and \( n_\epsilon = \min\{n \in \mathbb{N} : n - n^{1-\epsilon} - \frac{1}{2} n^{1-2\epsilon} - \frac{3}{8} n^{1-3\epsilon} - 1 \geq 1\} \).

**Lemma 6.** Let \( \epsilon \) be any fixed number with \( 0 < \epsilon < \frac{1}{2} \). For any polynomial time approximation algorithm \( A \), there exists at least one instance \( y(\epsilon, A) \) of MAX-k-INTERSECT with the instance size \( n \geq n_\epsilon \) such that \( \text{OPT}(y(\epsilon, A)) - A(y(\epsilon, A)) \geq \frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1 \).

**Proof.** Suppose that a polynomial time approximation algorithm \( A \) guarantees \( \text{OPT}(y) - A(y) \geq \frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1 \) for any instance \( y \), i.e. \( A(y) > \text{OPT}(y) - (\frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1) \).

This implies that if \( \text{OPT}(y) \geq n - n^{1-\epsilon} \) then \( A(y) > n_\epsilon \), where \( n_\epsilon = n - n^{1-\epsilon} - (\frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1) \). Now consider a graph \( G = (V, E) \) with \( |V| = n \), as an instance of MAX-CLIQUE. By Lemma 5 we know that \( A(f_r(G)) > n_\epsilon \) if \( \text{OPT}(G) \geq n^{1-\epsilon} \) and \( A(f_r(G)) \leq n_\epsilon \) if \( \text{OPT}(G) \leq n^{\epsilon} \). Hence, we can apply the polynomial time reduction \( f_r \) and algorithm \( A \) to distinguish an instance \( G \) with \( \text{OPT}(G) \geq n^{1-\epsilon} \) from another instance \( G' \) with \( \text{OPT}(G') \leq n^{\epsilon} \) in polynomial time. However, it is impossible unless \( P = \text{NP} \).

Lemma 5 and Lemma 6 directly lead to an inapproximability result of the MAX-k-INTERSECT problem.

**Theorem 7.** For any constant \( 0 < \epsilon < \frac{1}{2} \), the MAX-k-INTERSECT problem of a universe size \( n \geq n_\epsilon \) cannot be approximated in polynomial time within an absolute error of \( \frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1 \) unless \( P = \text{NP} \).

We prove the claim used in the proof of Lemma 5 as follows.

**Claim 1.** If \( k \leq n^2(n^{\epsilon} - 1)/2n^{\epsilon} \), then for large enough \( n \) we have

\[
v \geq n^{1-\epsilon} + \frac{1}{2} n^{1-2\epsilon} + \frac{3}{8} n^{1-3\epsilon} - 1.
\]

**Proof.** It is clear that

\[
v = \left(\frac{n^{1-\epsilon}}{2}\right) \geq \frac{n^{1-\epsilon}(n^{1-\epsilon} - 1)}{2},
\]

i.e.,

\[
v \geq \left(\frac{(n^{1-\epsilon})(n^{1-\epsilon} - 1)}{2}\right)^{1/2} = \left(\frac{n^{1-\epsilon}}{2}\right)^{1-\epsilon} \left(1 + \frac{-1}{n^{\epsilon}}\right)^{-1/2}.
\]
According to the binomial series: for a real number $d$ and $|x| < 1$,

$$(1 + x)^d = 1 + dx + \frac{d(d - 1)}{2!}x^2 + \frac{d(d - 1)(d - 2)}{3!}x^3 + \ldots.$$ 

Hence,

$$\left(1 + \frac{-n^e}{n}\right)^{1/2} = 1 + \frac{1}{2}\left(-\frac{n^e}{n}\right) + \frac{1/2(1/2 - 1)}{2!}\left(-\frac{n^e}{n}\right)^2 + \frac{(1/2)(1/2 - 1)(1/2 - 2)}{3!}\left(-\frac{n^e}{n}\right)^3 + \ldots$$

$$= 1 - \frac{1}{2}\left(\frac{n^e}{n}\right) - \frac{1}{8}\left(\frac{n^e}{n}\right)^2 - \frac{1}{16}\left(\frac{n^e}{n}\right)^3 - \ldots$$

$$> 1 - \frac{1}{2}\left(\frac{n^e}{n}\right) - \frac{1}{8}\left(\frac{n^e}{n}\right)^2 - \frac{1}{16}\left(\frac{n^e}{n}\right)^3, \quad \text{for } n > \frac{25}{24}$$

$$\left(1 + \frac{-1}{n^e}\right)^{-1/2} = 1 + \frac{-1}{2}\left(\frac{1}{n^e}\right) + \frac{-1/2(-1/2 - 1)}{2!}\left(\frac{1}{n^e}\right)^2 + \frac{-1/2(-1/2 - 1)(-1/2 - 2)}{3!}\left(\frac{1}{n^e}\right)^3 + \ldots$$

$$> 1 + \frac{1}{2}\left(\frac{1}{n^e}\right) + \frac{3}{8}\left(\frac{1}{n^e}\right)^2 + \frac{5}{16}\left(\frac{1}{n^e}\right)^3, \quad \text{for } n > 0.$$ 

Combine these two:

$$v > (n^{1-e})\left[1 - \frac{1}{2}\left(\frac{n^e}{n}\right) - \frac{1}{8}\left(\frac{n^e}{n}\right)^2 - \left(\frac{n^e}{n}\right)^3\right]$$

$$\times \left[1 + \frac{1}{2}\left(\frac{1}{n^e}\right) + \frac{3}{8}\left(\frac{1}{n^e}\right)^2 + \frac{5}{16}\left(\frac{1}{n^e}\right)^3\right]$$

$$= (n^{1-e})\left[1 + \frac{1}{2}n^{-e} + \frac{3}{8}n^{-2e} + \frac{5}{16}n^{-3e} - \frac{1}{2}n^{-1+e} - \frac{1}{4}n^{-2+e} - \frac{3}{64}n^{-2-\epsilon} - \frac{5}{128}n^{-2-\epsilon} - n^{-3+3\epsilon}\right]$$

$$= n^{1-e} + \frac{1}{2}n^{-1+e} + \frac{3}{8}n^{-1-2e} + \frac{5}{16}n^{-1-3e} - \frac{1}{2}$$

$$- \frac{1}{4}n^{-e} - O(n^{-2e})$$

$$\geq n^{1-e} + \frac{1}{2}n^{-1+e} + \frac{3}{8}n^{-1-3e} - 1,$$

for $n > \max\left\{\frac{25}{24}, \frac{16}{5}\right\}$. □

4. Conclusions

We give a correct proof to show that the hard-ness of approximating MAX-INTERSECT is exactly the same as MAX-CLIQUE. We also prove that MAX-$k$-INTERSECT cannot be approximated within an absolute error of $\frac{1}{8}n^{1-2e} + \frac{1}{2}n^{1-3e} - 1$ unless $P = \text{NP}$. It would be interesting to find a stronger inapproximable result for MAX-$k$-INTERSECT or design an efficient approximation algorithms for both problems.

References


