Information Processing Letters 112 (2012) 723–727

Contents lists available at SciVerse ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

On the inapproximability of maximum intersection problems

Min-Zheng Shieh [a], Shi-Chun Tsai [b,*,1], Ming-Chuan Yang [b]

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

ARTICLE INFO

Article history:
Received 27 February 2012
Received in revised form 16 June 2012
Accepted 20 June 2012
Available online 23 June 2012
Communicated by J. Torán

Keywords:
Approximation algorithm
Theory of computation
Inapproximability
Maximum intersection
Disclosure control

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 $\frac{1}{2}n^{1-2\epsilon} + O(n^{1-3\epsilon})$ unless $\mathbf{P} = \mathbf{NP}$. 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 $\frac{1}{2}n^{1-2\epsilon} + O(n^{1-3\epsilon})$ unless $\mathbf{P} = \mathbf{NP}$.

Formally, we define the problems as follows.

Problem 1 *(MAX-INTERSECT).* (See [3].) Given u sets $A_1, \ldots, A_u$, 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.

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 $k \leqslant u$, where each $A_i \subseteq U$, the goal is to select exactly k sets from $A_1, \ldots, A_u$ to maximize their intersection size.

The **NP**-hardness of MAX-k-INTERSECT was first proved by Vinterbo [15], but the proposed reduction does not lead to an inapproximability result. Later, Vinterbo [16] gave a greedy algorithm which can approximate MAX-k-INTERSECT within some constant factor if the cardinality

* Corresponding author.
E-mail addresses: mzshieh@nctu.edu.tw (M.-Z. Shieh), sctsai@cs.nctu.edu.tw (S.-C. Tsai), mingchuan.cs96g@g2.nctu.edu.tw (M.-C. Yang).

1 The research was supported in part by the National Science Council of Taiwan under contracts NSC-97-2221-E-009-064-MY3 and NSC-98-2221-E-009-078-MY3.

0020-0190/$ – see front matter
http://dx.doi.org/10.1016/j.ipl.2012.06.014

of all $A_1, \dots, A_u$ 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 $\{1, \dots, n\}$ be the collection of different types of devices produced by machines $A_1, \dots, A_u$. 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 $A_1, \dots, A_u$ 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 possessed 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 [16,11]. 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 $A_1, \dots, A_u$ be different machines, and each is associated with some production items in the universe $\{1, \dots, n\}$. 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 **NP** is not a subset of $\mathbf{BPTIME}(2^{n^{\epsilon}})$ for a small constant $\epsilon > 0$, then MAX-k-INTERSECT cannot have a polynomial time $(n^{\epsilon'})$-approximation algorithm, where n is the instance size and ϵ' depends only on ϵ. Note that $\mathbf{NP} \not\subset \mathbf{BPTIME}(2^{n^{\epsilon}})$ implies $\mathbf{P} \neq \mathbf{NP}$, 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 **NP**-hard to approximate MAX-k-INTERSECT within an absolute error of $\frac{1}{2}n^{1-2\epsilon} + \frac{3}{8}n^{1-3\epsilon} - 1$. 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 $V = V(G) = \{1, 2, \dots, n\}$ 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 $\{1, 2, \dots, n\}$ as $[n]$. $N(i)$ indicates the neighbor set of the vertex $i \in V$, that is, $N(i) = \{j:\ (i, j) \in E\}$. The cardinality of a set X is denoted as $|X|$. Let Π be an optimization problem and OPT_{Π} denote the optimal solution set of Π. Furthermore, if x is an instance of the problem Π, then $OPT(x)$ means the corresponding optimal solution of x. More precisely, we use the notation $OPT_{\Pi}(x)$. The measure of solutions used in this paper is the cardinality of a set, so we directly denote $|OPT_{\Pi}|$ or $|OPT(x)|$ for the optimization problem Π or the instance x, respectively. For example, if G is an instance of MAX-CLIQUE, then $OPT(G)$ is a maximum clique in G and $|OPT(G)|$ is the maximum clique size of G.

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

$$D(x, y) = \left| m^*(x) - m(x, y) \right|$$

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 Π is an absolute approximation algorithm if there exists a constant K such that, for any instance x of Π, $D(x, A(x)) \leqslant K$.

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

$$R\big(x, A(x)\big) = \max\left(\frac{|OPT_{\Pi}(x)|}{|A(x)|}, \frac{|A(x)|}{|OPT_{\Pi}(x)|} \right).$$

We say that A is an r-approximation algorithm for Π if, given any input instance x of Π, the performance ratio $R(x, A(x))$ of the approximation solution $|A(x)|$ is bounded by r, that is,

$$R\big(x, A(x)\big) \leqslant r.$$

Note that $r \geqslant 1$, and equivalently we have that $|A(x)| \geqslant \frac{1}{r} \cdot |OPT_{\Pi}(x)|$ if Π is a maximization problem.

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

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

Definition 5 *(Gap preserving reduction).* (See [14].) Let Π_1 and Π_2 be some maximization problems. A gap preserving reduction from Π_1 to Π_2 comes with four parameters (functions) f_1, α, f_2 and β. Given an instance x of Π_1, the reduction computes in polynomial time an instance y of Π_2 such that:

1. $|OPT_{\Pi_1}(x)| \geqslant f_1(x) \Rightarrow |OPT_{\Pi_2}(y)| \geqslant f_2(y)$,
2. $|OPT_{\Pi_1}(x)| \leqslant \alpha(|x|) f_1(x) \Rightarrow |OPT_{\Pi_2}(y)| \leqslant \beta(|y|) f_2(y)$.

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

For a maximization problem Π, let $\Pi^{\leqslant f}$ and $\Pi^{\geqslant f}$ be the languages of $\{x: |OPT(x)| \leqslant f(x)\}$ and $\{x: |OPT(x)| \geqslant f(x)\}$, respectively. A gap preserving reduction can be interpreted as a reduction which maps a promise problem $(\Pi_1^{\geqslant f_1}, \Pi_1^{\leqslant \alpha f_1})$ to another promise problem $(\Pi_2^{\geqslant f_2}, \Pi_2^{\leqslant \beta f_2})$. Let us see how this reduction works. Observe that if Π_2 has a polynomial time algorithm A_{Π_2} whose approximating factor is better than the gap $1/\beta$ (i.e. $OPT_{\Pi_2}/A_{\Pi_2} < 1/\beta$), then A_{Π_2} solves the promise problem $(\Pi_2^{\geqslant f_2}, \Pi_2^{\leqslant \beta f_2})$. Moreover, since $(\Pi_1^{\geqslant f_1}, \Pi_1^{\leqslant \alpha f_1})$ can be reduced to $(\Pi_2^{\geqslant f_2}, \Pi_2^{\leqslant \beta f_2})$ efficiently, there is a polynomial time algorithm solves the promise problem $(\Pi_1^{\geqslant f_1}, \Pi_1^{\leqslant \alpha f_1})$. Conversely, if $(\Pi_1^{\geqslant f_1}, \Pi_1^{\leqslant \alpha f_1})$ is **NP**-hard, then so is $(\Pi_2^{\geqslant f_2}, \Pi_2^{\leqslant \beta f_2})$.

The following inapproximable gap was first shown in Håstad's work [8] under the assumption **NP** ≠ **ZPP**. Then Zuckerman derandomized the reduction and proved the same gap under a weaker assumption **P** ≠ **NP**.

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

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

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

3. Inapproximability results

3.1. Hardness of MAX-INTERSECT

Let Π = MAX-CLIQUE and Φ = MAX-INTERSECT. Clifford and Popa [3] proposed a gap preserving reduction from the promise problem $(\Pi^{\geqslant n}, \Pi^{\leqslant n^{1-\epsilon}})$ to $(\Phi^{\geqslant n}, \Phi^{\leqslant n^{1-\epsilon}})$. 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 **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 $(\Pi^{\geqslant n}, \Pi^{\leqslant n^{1-\epsilon}})$ is in **P**. Even though their reduction is valid, it does not imply the hardness of the promised MAX-INTERSECT problem $(\Phi^{\geqslant n}, \Phi^{\leqslant n^{1-\epsilon}})$. In fact, it is also in **P** to distinguish the case $|OPT_{\text{MAX-INTERSECT}}| = n$ from the case $|OPT_{\text{MAX-INTERSECT}}| = n^{1-\epsilon}$. 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 $n/n^{1-\epsilon} = n^{\epsilon}$.

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_r from MAX-CLIQUE to MAX-INTERSECT is defined as: for a given graph $G = (V, E)$ with $V = \{1, \ldots, n\}$, let $f_r(G)$ 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_r is a polynomial time reduction.

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

$$\bigl|OPT(G)\bigr| = \bigl|OPT\bigl(f_r(G)\bigr)\bigr|.$$

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 $|OPT(G)| \leqslant |OPT(f_r(G))|$, where $OPT(G)$ is a maximum clique of G. Let $S_i = N(i) \cup \{i\}$ for $i \in OPT(G)$ and $S_i = V \setminus \{i\}$ for $i \notin OPT(G)$. Observe that for $i, j \in OPT(G)$ and $k \notin OPT(G)$, it is clear that $\{i, j\} \subseteq S_i$, $\{i, j\} \subseteq S_j$, and $\{i, j\} \subseteq S_k$. So for every $i \in V$, we have $OPT(G) \subseteq S_i$ and hence $OPT(G) \subseteq S_1 \cap S_2 \cap \cdots \cap S_n$. This concludes $|OPT(G)| \leqslant |S_1 \cap S_2 \cap \cdots \cap S_n| \leqslant |OPT(f_r(G))|$.

Next we prove $|OPT(G)| \geqslant |OPT(f_r(G))|$. Assume for $i = 1, \ldots, n$, $S_i^* \in A_i$ are the selected subsets which maximize the intersection cardinality. It means $|S_1^* \cap S_2^* \cap \cdots \cap S_n^*| = |OPT(f_r(G))|$. Denote $S_1^* \cap S_2^* \cap \cdots \cap S_n^*$ 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 $\{i, j\} \subset c^*(G)$, $(i, j) \in E$, i.e., $c^*(G)$ must be a clique. Thus, we have $|c^*(G)| \leqslant |OPT(G)|$. By the definition of $c^*(G)$, we have that $|OPT(G)| \geqslant |c^*(G)| = |OPT(f_r(G))|$. □

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* **P** = **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 **P** = **NP**. □

3.2. Hardness of MAX-k-INTERSECT

Next we consider the MAX-k-INTERSECT problem, which had been already proved to be **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 ϵ and let $k = \binom{n^{1-\epsilon}}{2}$, where $n = |V|$. The corresponding reduced instance is $f_r(G) = \langle U, \{A_j\}_{j=1}^m, k\rangle$. This mapping obviously can be done in polynomial time. To prove our inapproximable 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*

$$|E| \leqslant \frac{p-1}{2p}|V|^2.$$

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. $|OPT(G)| \geqslant n^{1-\epsilon} \Rightarrow |OPT(f_r(G))| \geqslant n - n^{1-\epsilon}$,
2. $|OPT(G)| \leqslant n^{\epsilon} \Rightarrow |OPT(f_r(G))| \leqslant n - (n^{1-\epsilon} + \frac{1}{2}n^{1-2\epsilon} + \frac{3}{8}n^{1-3\epsilon} - 1)$.

Proof. Note that we let $k = \binom{n^{1-\epsilon}}{2}$. If $|OPT(G)| \geqslant n^{1-\epsilon}$, then there is a complete subgraph $C \subset G$ with $n^{1-\epsilon}$ vertices. W.l.o.g. let $C = (V', E')$, where $V' = \{1, 2, \ldots, n^{1-\epsilon}\}$ and $E' = \{e_1, \ldots, e_k\}$. According to the above reduction, we can select subsets $A_1, A_2, \ldots, A_k$ which correspond to the clique edges $e_1, \ldots, e_k$. Hence $\bigcap_{j=1}^k A_j = U \setminus \{1, \ldots, n^{1-\epsilon}\}$, and we have $|OPT(f_r(G))| \geqslant |\bigcap_{j=1}^k A_i| = n - n^{1-\epsilon}$.

If $|OPT(G)| \leqslant n^{\epsilon}$, then it implies that any k-edge induced subgraph of G does not contain an $(n^{\epsilon}+1)$-clique. In order to estimate $|OPT(f_r(G))|$, 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 $k \leqslant x^2(n^{\epsilon}-1)/2n^{\epsilon}$. With the binomial series expansion (see Claim 1 later for details), we have

$$\begin{aligned} x &\geqslant \left(n^{1-\epsilon}\right)\left(1 + \frac{-n^{\epsilon}}{n}\right)^{1/2}\left(1 + \frac{-1}{n^{\epsilon}}\right)^{-1/2} \\ &\geqslant n^{1-\epsilon} + \frac{1}{2}n^{1-2\epsilon} + \frac{3}{8}n^{1-3\epsilon} - 1. \end{aligned}$$

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 $|OPT(f_r(G))| \leqslant n - x^* = n - (n^{1-\epsilon} + \frac{1}{2}n^{1-2\epsilon} + \frac{3}{8}n^{1-3\epsilon} - 1)$. □

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

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

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

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

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

Theorem 7. *For any constant $0 < \epsilon < \frac{1}{3}$, the MAX-k-INTERSECT problem of a universe size $n \geqslant n_{\epsilon}$ 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** = **NP**.

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

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

$$v \geqslant n^{1-\epsilon} + \frac{1}{2}n^{1-2\epsilon} + \frac{3}{8}n^{1-3\epsilon} - 1.$$

Proof. It is clear that

$$v^2\frac{(n^{\epsilon}-1)}{2n^{\epsilon}} \geqslant k = \frac{n^{1-\epsilon}(n^{1-\epsilon}-1)}{2},$$

i.e.

$$\begin{aligned} v &\geqslant \left[\left(n^{1-\epsilon}\right)\left(n^{1-\epsilon}-1\right)n^{\epsilon}\left(n^{\epsilon}-1\right)^{-1}\right]^{1/2} \\ &= \left[\left(n^{1-\epsilon}\right)\left(n^{1-\epsilon}\left(1 + \frac{-1}{n^{1-\epsilon}}\right)\right)\right. \end{aligned}$$

$$\times n^{\epsilon}\left(n^{\epsilon}\left(1+\frac{-1}{n^{\epsilon}}\right)\right)^{-1}\Bigg]^{1/2}$$
$$= \left(n^{1-\epsilon}\right)\left(1+\frac{-n^{\epsilon}}{n}\right)^{1/2}\left(1+\frac{-1}{n^{\epsilon}}\right)^{-1/2}.$$

According to the binomial series: for a real number d and $|x| < 1$,

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

Hence,

$$\left(1+\frac{-n^{\epsilon}}{n}\right)^{1/2}$$
$$= 1 + \left(\frac{1}{2}\right)\left(\frac{-n^{\epsilon}}{n}\right) + \frac{(1/2)(1/2-1)}{2!}\left(\frac{-n^{\epsilon}}{n}\right)^2 + \frac{(1/2)(1/2-1)(1/2-2)}{3!}\left(\frac{-n^{\epsilon}}{n}\right)^3 + \cdots$$
$$= 1 - \frac{1}{2}\left(\frac{n^{\epsilon}}{n}\right) - \frac{1}{8}\left(\frac{n^{\epsilon}}{n}\right)^2 - \frac{1}{16}\left(\frac{n^{\epsilon}}{n}\right)^3 - \frac{5}{128}\left(\frac{n^{\epsilon}}{n}\right)^4 - \cdots$$
$$> 1 - \frac{1}{2}\left(\frac{n^{\epsilon}}{n}\right) - \frac{1}{8}\left(\frac{n^{\epsilon}}{n}\right)^2 - 1\cdot\left(\frac{n^{\epsilon}}{n}\right)^3, \quad \text{for } n > (25/24)^{\frac{1}{1-\epsilon}},$$

$$\left(1+\frac{-1}{n^{\epsilon}}\right)^{-1/2}$$
$$= 1 + \left(\frac{-1}{2}\right)\left(\frac{-1}{n^{\epsilon}}\right) + \frac{(-1/2)(-1/2-1)}{2!}\left(\frac{-1}{n^{\epsilon}}\right)^2 + \frac{(-1/2)(-1/2-1)(-1/2-2)}{3!}\left(\frac{-1}{n^{\epsilon}}\right)^3 + \cdots$$
$$> 1 + \frac{1}{2}\left(\frac{1}{n^{\epsilon}}\right) + \frac{3}{8}\left(\frac{1}{n^{\epsilon}}\right)^2 + \frac{5}{16}\left(\frac{1}{n^{\epsilon}}\right)^3, \quad \text{for } n > 0.$$

Combine these two:

$$v > \left(n^{1-\epsilon}\right)\left[1 - \frac{1}{2}\left(\frac{n^{\epsilon}}{n}\right) - \frac{1}{8}\left(\frac{n^{\epsilon}}{n}\right)^2 - \left(\frac{n^{\epsilon}}{n}\right)^3\right] \times \left[1 + \frac{1}{2}\left(\frac{1}{n^{\epsilon}}\right) + \frac{3}{8}\left(\frac{1}{n^{\epsilon}}\right)^2 + \frac{5}{16}\left(\frac{1}{n^{\epsilon}}\right)^3\right]$$
$$= \left(n^{1-\epsilon}\right)\Bigg[1 + \frac{1}{2}n^{-\epsilon} + \frac{3}{8}n^{-2\epsilon} + \frac{5}{16}n^{-3\epsilon} - \frac{1}{2}n^{-1+\epsilon} - \frac{1}{4}n^{-1} - \frac{3}{16}n^{-1-\epsilon} - \frac{5}{32}n^{-1-2\epsilon} - \frac{1}{8}n^{-2+2\epsilon} - \frac{1}{16}n^{-2+\epsilon} - \frac{3}{64}n^{-2} - \frac{5}{128}n^{-2-\epsilon} - n^{-3+3\epsilon} - \frac{1}{2}n^{-3+2\epsilon} - \frac{3}{8}n^{-3+\epsilon} - \frac{5}{16}n^{-3}\Bigg]$$
$$= n^{1-\epsilon} + \frac{1}{2}n^{1-2\epsilon} + \frac{3}{8}n^{1-3\epsilon} + \frac{5}{16}n^{1-4\epsilon} - \frac{1}{2} - \frac{1}{4}n^{-\epsilon} - O\left(n^{-2\epsilon}\right)$$
$$\geqslant n^{1-\epsilon} + \frac{1}{2}n^{1-2\epsilon} + \frac{3}{8}n^{1-3\epsilon} - 1,$$
$$\text{for } n > \max\left\{(25/24)^{\frac{1}{1-\epsilon}}, (16/5)^{\frac{1}{1-3\epsilon}}\right\}. \qquad \square$$

4. Conclusions

We give a correct proof to show that the hardness 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 $\frac{1}{2}n^{1-2\epsilon} + \frac{3}{8}n^{1-3\epsilon} - 1$ unless $\mathbf{P} = \mathbf{NP}$. It would be interesting to find a stronger inapproximable result for MAX-k-INTERSECT or design an efficient approximation algorithms for both problems.

References

[1] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, M. Protasi, Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties, second corrected ed., Springer-Verlag, Berlin/Heidelberg, 2003.

[2] P. Bonizzoni, G.D. Vedova, R. Dondi, Anonymizing binary and small tables is hard to approximate, Journal of Combinatorial Optimization 22 (1) (2011) 97–119.

[3] R. Clifford, A. Popa, Maximum subset intersection, Information Processing Letters 111 (2011) 323–325.

[4] U. Feige, L. Lovász, P. Tetali, Approximating min sum set cover, Algorithmica 40 (4) (2004) 219–234.

[5] R. Gandhi, S. Khuller, A. Srinivasan, Approximation algorithms for partial covering problems, Journal of Algorithms 53 (1) (2004) 55–84.

[6] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of **NP**-Completeness, W.H. Freeman, 1979.

[7] O. Goldreich, On promise problems (a survey in memory of Shimon Even), Electronic Colloquium on Computational Complexity (ECCC) (2005) TR05–TR018.

[8] J. Håstad, Clique is hard to approximate within $n^{1-\epsilon}$, Acta Mathematica 182 (1999) 105–142.

[9] J. Håstad, Some optimal inapproximability results, Journal of the ACM 48 (2001) 798–859.

[10] S. Khuller, A. Moss, J. Naor, The budgeted maximum coverage problem, Information Processing Letters 70 (1) (1999) 39–45.

[11] A. Meyerson, R. Williams, On the complexity of optimal K-anonymity, in: Proceedings of the Twenty-Third ACM Symposium on Principles of Database Systems (PODS), Paris, France, 2004.

[12] R. Peeters, The maximum edge biclique problem is **NP**-complete, Discrete Applied Mathematics 131 (2003) 651–654.

[13] J.H. van Lint, R.M. Wilson, A Course in Combinatorics, second ed., Cambridge University Press, 2001.

[14] V.V. Vazirani, Approximation Algorithms, Springer-Verlag, Berlin/Heidelberg, 2003.

[15] S.A. Vinterbo, Maximum k-intersection, edge labeled multigraph max capacity k-path, and max factor k-GCD are all **NP**-hard, Decision Systems Group, Harvard Medical School, Technical Report, 2002.

[16] S.A. Vinterbo, Privacy: A machine learning view, IEEE Transactions on Knowledge and Data Engineering 16 (8) (2004) 939–948.

[17] E.C. Xavier, A note on a maximum k-subset intersection problem, Information Processing Letters 112 (2012) 471–472.

[18] D. Zuckerman, Linear degree extractors and the inapproximability of max clique and chromatic number, Theory of Computing 3 (2007) 103–128.