

# Constant Approximating Parameterized $k$ -SETCOVER is $W[2]$ -hard

Bingkai Lin<sup>\*</sup>    Xuandi Ren<sup>†</sup>    Yican Sun<sup>‡</sup>    Xiuhan Wang<sup>§</sup>

October 22, 2022

## Abstract

In this paper, we prove that it is  $W[2]$ -hard to approximate  $k$ -SETCOVER within any constant ratio. Our proof is built upon the recently developed threshold graph composition technique. We propose a strong notion of threshold graphs and use a new composition method to prove this result. Our technique could also be applied to rule out polynomial time  $o\left(\frac{\log n}{\log \log n}\right)$  ratio approximation algorithms for the non-parameterized  $k$ -SETCOVER problem with  $k$  as small as  $O\left(\frac{\log n}{\log \log n}\right)^3$ , assuming  $W[1] \neq \text{FPT}$ . We highlight that our proof does not depend on the well-known PCP theorem, and only involves simple combinatorial objects.

## 1 Introduction

In the  $k$ -SETCOVER problem, we are given a bipartite graph  $G = (S \cup U, E)$  and an integer  $k$ , and the goal is to decide whether there exist  $k$  vertices in  $S$  such that each node in the *universe set*  $U$  is a neighbor of one of the  $k$  vertices. In the classic complexity regime, this problem is known to be NP-complete [Kar72]. Thus assuming  $P \neq \text{NP}$ , no algorithm can solve  $k$ -SETCOVER in polynomial time. To circumvent this intractability, many previous works have focused on efficient approximation algorithms for  $k$ -SETCOVER. It admits a simple greedy algorithm with approximation ratio  $(\ln n - \ln \ln n + \Theta(1))$  [Chv79, Joh74, Lov75, Sla97, Ste74]. On the opposite side, the hardness of approximation of  $k$ -SETCOVER has also been intensively studied [AMS06, DS14, Fei98, LY94, RS97]. The state-of-the-art result by Dinur and Stenur [DS14] shows that approximating  $k$ -SETCOVER within an  $(1 - \varepsilon) \cdot \ln n$  factor for any  $\varepsilon > 0$  is NP-hard.

To obtain a more fine-grained comprehension of NP-hard problems, it is natural to consider *parameterization*. In the parameterized complexity regime, people wonder whether there is any  $f(k) \cdot |G|^{O(1)}$  time algorithm (FPT algorithm) that can solve the  $k$ -SETCOVER problem, where  $f$

---

<sup>\*</sup>Nanjing University. Email: lin@nju.edu.cn

<sup>†</sup>University of California, Berkeley. Email: xuandi\_ren@berkeley.edu

<sup>‡</sup>School of Computer Science, Peking University. Email: sycpku@pku.edu.cn

<sup>§</sup>Tsinghua University. Email: wangzh19@mails.tsinghua.edu.cn

could be any computable function (e.g.,  $k!$  or  $k^k$ ). As was shown in [DF95],  $k$ -SETCOVER is the canonical  $W[2]$ -complete problem, which means that, unless  $W[2] = \text{FPT}$ ,  $k$ -SETCOVER does not admit any FPT algorithm. In parallel with the classical complexity regime, it is natural to ask:

*Assuming  $W[2] \neq \text{FPT}$ , does  $k$ -SETCOVER admit good FPT approximation algorithms?*

Note that since  $k$ -SETCOVER is a  $W[2]$ -complete problem, the question above is (almost) equivalent to asking whether there is an FPT self-reduction from exact  $k$ -SETCOVER to its gap version<sup>1</sup>.

In recent years, the parameterized inapproximability of  $k$ -SETCOVER has been established under various assumptions stronger than  $W[2] \neq \text{FPT}$ . Chen and Lin [CL19] proved that it is  $W[1]$ -hard to approximate  $k$ -SETCOVER within any constant ratio. Chalermsook et al. [CCK<sup>+</sup>17] showed that assuming Gap-ETH,  $k$ -SETCOVER cannot be approximated within a  $(\log n)^{O(1/k)}$  factor in  $n^{o(k)}$  time. Using the *Distributed PCP framework* [ARW17], Karthik, Laekhanukit and Manurangsi [KLM19] ruled out  $(\log n)^{1/\text{poly}(k)}$  ratio  $n^{k-\varepsilon}$ -time approximation algorithms for  $k$ -SETCOVER under SETH. They also proved similar inapproximability results under  $k$ -SUM Hypothesis, ETH and  $W[1] \neq \text{FPT}$ , respectively. The hardness of approximation factors of these results were improved to  $(\log n / \log \log n)^{1/k}$  in [Lin19] by combining  $k$ -SETCOVER instances with some gap-gadgets. Karthik and Navon [KN21] gave a simple construction of these gap-gadgets based on error correcting codes and named this technique *Threshold Graph Composition*.

In contrast to the success of proving inapproximability of  $k$ -SETCOVER under stronger assumptions, the FPT inapproximability of  $k$ -SETCOVER under  $W[2] \neq \text{FPT}$  remained completely open [KLM19, KN21, FKLM20]. Below we summarize the difficulties encountered in the previous approaches.

**Barriers of previous approaches.** First, neither of the two pioneering works [CCK<sup>+</sup>17, CL19] are applicable to this setting. In detail, the approach proposed in [CL19] requires the reduction’s starting point to admit a product structure (e.g.  $k$ -CLIQUE) in order to perform gap amplification. However, under the assumption  $W[2] \neq \text{FPT}$ , the starting point is the  $k$ -SETCOVER itself, whose product structure is notoriously hard to understand [KZ96]. On the other hand, the method of [CCK<sup>+</sup>17] highly depends on the inherent gap in the Gap-ETH assumption, thus could not be adapted to our setting.

Second, there are indeed some other advanced techniques, namely, the Distributed PCP Framework and the Threshold Graph Composition, which have been successfully used to prove the inapproximability of  $k$ -SETCOVER under  $W[1] \neq \text{FPT}$ , ETH, and SETH. However, under a weaker assumption  $W[2] \neq \text{FPT}$ , even constant FPT inapproximability of  $k$ -SETCOVER is difficult to reach using these two techniques. Generally speaking, in order to create a constant gap for  $k$ -SETCOVER under  $W[2] \neq \text{FPT}$ , both of these methods need to first establish the  $W[2]$ -hardness of  $k$ -SETCOVER with a small universe set. Nevertheless, establishing such a hardness result is highly non-trivial, and we believe it to be unrealizable [JP17].

Below, we illustrate their technique barriers in detail, respectively.

- The work [Lin19] proposes a gap-producing self-reduction for  $k$ -SETCOVER, i.e., it constructs a reduction from  $k$ -SETCOVER to  $c$ -gap  $k$ -SETCOVER. However, to create the constant gap  $c$ ,

---

<sup>1</sup>For any  $c > 1$ , on input a  $k$ -SETCOVER instance  $(G, k)$ , the goal of  $c$ -gap  $k$ -SETCOVER is to distinguish between the cases where  $G$  has  $k$ -size solution and the cases where  $G$  has no  $ck$ -size solution.

the running time of this reduction will be  $|U|^{(ck)^k}$ , where  $|U|$  represents the universe size for the input  $k$ -SETCOVER instance. Hence, to apply this method under our setting  $W[2] \neq \text{FPT}$ , we need to first prove that  $k$ -SETCOVER remains  $W[2]$ -hard even when  $|U| = n^{O(1/(ck)^k)}$ .

- The approach of [KLM19] is more complicated. It starts with a  $(1 - 1/k)$ -gap  $k$ -MAXCOVER<sup>2</sup> instance which has  $q$  right variables and right alphabet  $\Sigma_U$ . It then uses the product method to amplify its gap to  $(1 - 1/k)^t$  with the price of increasing  $q$  to  $q^t$  and  $\Sigma_U$  to  $\Sigma_U^t$ . Finally it transfers a  $(1 - 1/k)^t$ -gap  $k$ -Max-Cover instance with alphabet  $\Sigma_U^t$  and variable  $q^t$  to a  $(1 - 1/k)^{-t/k}$ -gap  $k$ -SETCOVER instance with size at least  $q^t \cdot k^{|\Sigma_U|^t}$ . It is not hard to see that to obtain a constant gap for  $k$ -SETCOVER,  $t$  should be at least  $\Omega(k^2)$ , which means that  $q$  should be at most  $n^{O(1/k^2)}$ . However, as far as we known, there is no FPT-reduction from  $k$ -SETCOVER to  $(1 - 1/k)$ -gap  $k$ -MAXCOVER with  $q$  at most  $n^{O(1/k^2)}$ . We note that there is a simple reduction from  $k$ -SETCOVER to  $k$ -MAXCOVER with  $q = |U|$  and  $\Sigma_U = [k]$ . Such a reduction is far from satisfactory because the  $k$ -MAXCOVER instance it produces does not have a  $(1 - 1/k)$ -gap and it requires the  $W[2]$ -hardness of  $k$ -SETCOVER with  $|U| = n^{O(1/k^2)}$ .

Note that establishing such a hardness result is equivalent to making an FPT self-reduction that “compresses” the  $k$ -SETCOVER problem, i.e., the reduction starts with a  $k$ -SETCOVER instance with universe size  $n$  and results in another equivalent instance of  $k$ -SETCOVER with significantly smaller universe size  $n^{o(1)}$ . Unfortunately, there is no known FPT-reduction that achieves this, and we list two partial results below, which suggest that such a reduction might be unrealizable.

- First, we prove that (See Lemma 14 in the Appendix), unless  $W[2] = W[1]$ , there is no FPT self-reduction for  $k$ -SETCOVER that can compress the size of universe set to  $f(k) \cdot \log n$  for any computable function  $f$ .
- In addition, research from the lower bound of kernelization [JP17] shows that: unless  $\text{NP} \subseteq \text{coNP}/\text{poly}$  and the polynomial hierarchy PH collapses, for every  $\varepsilon > 0$ , there is no polynomial self-reduction for non-parameterized  $k$ -SETCOVER that compress the size of instance to  $n^{2-\varepsilon}$ . This also suggests that the corresponding parameterized version of  $k$ -SETCOVER may also be hard to compress.

**Our results.** We bypass these technique barriers by coming up with a strong version of the threshold graph and a new composition method that can create a constant gap for  $k$ -SETCOVER instance with a large universe set. Using this technique, we give the first inapproximability result for  $k$ -SETCOVER under  $W[2] \neq \text{FPT}$ . In summary, the main contribution of our paper is:

**Theorem 1.** *Assuming  $W[2] \neq \text{FPT}$ , there is no FPT algorithm which can approximate  $k$ -SETCOVER within any constant ratio.*

---

<sup>2</sup>In the  $k$ -MAXCOVER problem, we are given a bipartite graph  $I = (V, U, E)$  with  $|V| = k$  and  $|U| = q$ , two alphabet  $\Sigma_V$  and  $\Sigma_U$  and constraints  $C_e \subseteq \Sigma_V \times \Sigma_U$  for each  $e \in E$ . The goal is to find assignments  $\sigma_V : V \rightarrow \Sigma_V$  and  $\sigma_U : U \rightarrow \Sigma_U$  to maximize the number of  $u \in U$  whose edges’ constraints are all satisfied (Such vertex  $u$  are called *covered* by assignments  $\sigma_V$  and  $\sigma_U$ ). For  $\delta \in (0, 1)$ , the  $\delta$ -gap  $k$ -MAXCOVER is to distinguish between the cases where all vertices in  $U$  can be covered and the cases where at most  $\delta$ -fraction of  $U$  can be covered.

As a further application, our technique can also be used to rule out polynomial time  $o\left(\frac{\log n}{\log \log n}\right)$  ratio approximation algorithms for  $k$ -SETCOVER with  $k$  as small as  $O(\log^3 n)$  under  $W[1] \neq \text{FPT}$ .

**Theorem 2.** *Assuming  $W[1] \neq \text{FPT}$ , there is no polynomial time algorithm which can approximate  $k$ -SETCOVER within  $o\left(\frac{\log n}{\log \log n}\right)$  ratio, even if  $k$  is as small as  $O\left(\frac{\log n}{\log \log n}\right)^3$ .*

We emphasize that our Theorem 2 is non-trivial since it overcomes limitations of previous works in two aspects.

- *Limitations of results in parameterized complexity.* First, all the previous results [CL19, CCK<sup>+</sup>17, KLM19, Lin19] on the parameterized inapproximability of  $k$ -SETCOVER only ruled out  $o((\log n)^{1/k})$  approximation ratios, which was further pointed out as a barrier [KN21]. However, the folklore polynomial time greedy algorithm could approximate  $k$ -SETCOVER with approximation factor  $\Theta(\log n)$ . There is a huge gap between them. Our Theorem 2 is the first one to obtain inapproximability result for  $k$ -SETCOVER with a small solution size that can bypass the  $(\log n)^{1/k}$  barrier.
- *Limitations of the PCP theorem.* Second, the major technique used to rule out polynomial approximation algorithm for  $k$ -SETCOVER is the well-celebrated PCP theorem [Fei98, AMS06]. However, as pointed out in [CL19], results of this type are unlikely provable using the classic PCP machinery since the PCP theorem always produces  $k$ -SETCOVER instances with a large solution size  $\Omega(n)$ . However, our Theorem 2 could rule out polynomial time approximation with a small solution size  $O\left(\frac{\log n}{\log \log n}\right)^3$ . Furthermore, assuming the mild assumption  $\text{NP} \not\subseteq \text{TIME}(2^{\text{poly} \log n})$ ,  $k$ -SETCOVER with solution size  $\log^{O(1)} n$  is not even NP-hard, thus we could not derive hardness for  $k$ -SETCOVER with  $\text{poly} \log n$  solution size from gap SAT (or the PCP theorem).

Finally, as a suggestion to future work, we remark that: if one could improve the polynomial time lower bound in Theorem 2 to  $n^{k-o(1)}$  under SETH, then the ratio  $o\left(\frac{\log n}{\log \log n}\right)$  is tight. The reason is that: there is a simple algorithm with running time  $n^{k-\Omega(1)}$  that approximates  $k$ -SETCOVER with approximation factor  $O\left(\frac{\log |U|}{\log k}\right)$ . Under our setting, we have that  $k = \log^{\Theta(1)} n$ , hence, the algorithm could approximate  $k$ -SETCOVER with factor  $O\left(\frac{\log |U|}{\log k}\right) = O\left(\frac{\log n}{\log \log n}\right)$  in  $n^{k-\Omega(1)}$  time. The algorithm simply modifies the well-known greedy algorithm for  $k$ -SETCOVER, and is illustrated detailedly as follows.

*Details of the algorithm.* Fix some integer  $2 \leq T < k$ , our algorithm repeatedly picks  $k - T$  vertices into the final solution in  $O(n^{k-T+1})$  time, such that these vertices cover the largest fraction of remaining elements in the universe. Repeat the procedure above for  $r$  times with  $r$  to be determined, we could obtain a solution of size  $r \cdot (k - T)$ . We analyze the correctness and the approximation ratio of our algorithm below.

*Analysis.* In each repetition, we will find  $k - T$  vertices that cover at least  $1 - T/k$  fraction of remaining elements in the universe. Thus after  $r$  repetitions, the number of the uncovered

elements will be no more than  $(T/k)^r \cdot |U|$ . To ensure that we have covered all elements in the universe, we need to choose the number of repetitions  $r = \frac{\log |U|}{\log k - \log T}$  by solving the inequality for uncovered elements  $(T/k)^r \cdot |U| < 1$ .

Note that proving tight lower bounds and upper bounds for approximation algorithms of  $k$ -SETCOVER for different choices of  $k$  and approximation ratios is still a big challenge. We would like to mention that our proof might shed a new light on the study of this problem since it is an elementary proof, which circumvents the complex PCP machinery and only involves simple combinatorial objects such as error correcting codes.

## 1.1 Our Techniques

Our main technique is the *threshold graph composition*, which was first used in [Lin18], and has been applied to create gaps for many parameterized problems [CL19, Lin19, BBE<sup>+</sup>21, KN21]. At a very high level, this method first constructs a graph with some threshold properties and then combines this graph with the input instance to produce a gap instance of the desired problem.

In this paper, we introduce a strong variant of the threshold graph and propose a new way to compose this threshold graph with the original  $k$ -SETCOVER instance. Below we firstly overview the proof in the work [Lin19], and then illustrate our proof for Theorem 1 briefly. Theorem 2 could be analogously proved by simply choosing another combination of parameters.

**The Original Proof in [Lin19].** The threshold graph appeared in [Lin19] are bipartite graphs  $T = (A \dot{\cup} B, E_T)$  with the following properties:

- (i)  $A = A_1 \dot{\cup} A_2 \dot{\cup} \dots \dot{\cup} A_k$ .
- (ii)  $B = B_1 \dot{\cup} B_2 \dot{\cup} \dots \dot{\cup} B_m$ .
- (iii) For any  $a_1 \in A_1, \dots, a_k \in A_k$  and  $i \in [m]$ ,  $a_1, \dots, a_k$  have a common neighbor in  $B_i$ .
- (iv) For any  $X \subseteq A$  and  $b_1 \in B_1, \dots, b_m \in B_m$ , if every  $b_i$  has  $k + 1$  neighbors in  $X$ , then  $|X| > h$ .

Given a set cover instance  $\Gamma = (S \dot{\cup} U, E)$  and a threshold graph  $T = (A \dot{\cup} B, E_T)$  with  $A = A_1 \dot{\cup} A_2, \dots, A_k$ ,  $B = B_1 \dot{\cup} \dots \dot{\cup} B_m$  and  $|A_i| = |S|$ , the reduction in [Lin19] treats each  $A_i$  as a copy of  $S$  and creates a set cover instance  $\Gamma_i = (A, U^{B_i}, E_i)$  for every  $i \in [m]$  so that in order to cover  $U^{B_i}$ , one has to pick  $\ell$  notes  $\{a_1, \dots, a_\ell\}$  from  $A$  satisfying the following two conditions:

- (a)  $a_1, \dots, a_\ell$  cover  $U$  in the instance  $\Gamma$ ,
- (b)  $a_1, \dots, a_\ell$  have a common neighbor in  $B_i$  in the threshold graph  $T$ .

After that, it takes the union of all  $\Gamma_i = (A, U^{B_i}, E_i)$  for  $i \in [m]$ . It is not hard to see that if the input instance  $\Gamma$  has a solution of size  $k$ , then there exist  $a_1 \in A_1, \dots, a_k \in A_k$  which cover all vertex in  $U$ . By (a), (b) and the property (iii), these vertices  $a_1, \dots, a_k$  also cover  $U^{B_i}$  for all  $i \in [m]$ .

On the other hand, if the input instance  $\Gamma$  has no solution of size  $k$ , below we show that every solution  $X$  of the output instance must have a size larger than  $h$ . To cover the part  $\Gamma_i$  in the output

instance, any solution  $X$  has to incorporate  $k + 1$  vertices in to cover  $U$  to satisfy condition (a). Furthermore, these vertices share a common neighbor  $b_i \in B_i$  to satisfy condition (b). This means that we could find a vertex  $b_i \in B_i$  with  $k + 1$  neighbors in our solution  $X$ . Since this works for every  $i \in [m]$ , we conclude by the threshold property (iv) that  $|X| > h$ .

The reduction above fails to prove constant inapproximability of  $k$ -SETCOVER under  $W[2] \neq$  FPT because to get a constant hardness factor, one needs to set  $h = c \cdot k$  for some constant  $c$ . Unfortunately, the current construction of threshold graph has  $|B_i| = h^k$ . Thus the reduction has running time at least  $|U|^{h^k}$ , which is not FPT when  $|U| = \Omega(n)$ . Previous reductions [Lin19, KLM19] use the fact that under stronger hypotheses,  $k$ -SETCOVER remains  $W[1]$ -hard even when  $|U| = k^{O(1)} \log n$ . However, unless  $W[2] = W[1]$ , one cannot prove  $k$ -SETCOVER remains  $W[2]$ -hard when  $|U| = k^{O(1)} \log n$  (See Lemma 14 in the Appendix). Jansen and Pieterse [JP17] proved that unless  $NP \subseteq \text{coNP}/\text{poly}$ , there is no  $O(n^{2-\epsilon})$ -size compression for general non-parameterized  $k$ -SETCOVER, which suggests it might not be possible to prove  $k$ -SETCOVER with small  $|U|$  is  $W[2]$ -hard.

**Our Reduction.** Below we illustrate our reduction at a high level. We firstly define a stronger version of the threshold graph by replacing (iv) with a property reminiscent of the soundness condition of *multi-assignment* PCP [AMS06].

(iv') For any  $X \subseteq A$  and  $I \subseteq [m]$  with  $|I| \geq \epsilon m$ , if for every  $i \in I$ ,  $b_i$  has  $k + 1$  neighbors in  $X$ , then  $|X| > h$ .

It turns out that such strong threshold graphs can be constructed using Error Correcting Codes, which was proposed by Karthik and Navon [KN21].

Then we define a new threshold graph composition which replaces  $\Gamma_i$  by  $\Gamma'_i = (A \cup B_i, U^c \times B_i^c, E'_i)$ . By choosing the edge set  $E'_i$  carefully, we can guarantee that in order to cover  $U^c \times B_i^c$ , one has to

- (a) either pick  $c + 1$  vertices from  $B_i$ ,
- (b) or pick  $a_1, \dots, a_\ell$  from  $A$  and  $b_i \in B_i$  such that
  - (b.1)  $a_1, \dots, a_\ell$  cover  $U$  in the instance  $\Gamma$ ,
  - (b.2)  $b_i$  is a common neighbor of  $a_1, \dots, a_\ell$  in the threshold graph.

Suppose the input instance  $\Gamma$  has a solution of size  $k$ . Since each  $A_i$  is a copy of the set  $S$  in  $\Gamma$ , let  $a_1 \in A_1, \dots, a_k \in A_k$  be the vertices that can cover  $U$ . By the property of threshold graph,  $a_1, \dots, a_k$  have a common neighbor  $b_i \in B_i$  for every  $i \in [m]$ . The vertices  $a_1, \dots, a_k$  and  $b_1, \dots, b_m$  together can cover  $U^c \times B_i^c$  for all  $i \in [m]$ .

On the other hand, if  $\Gamma$  has no  $k$ -size solution, then in order to cover every  $U^c \times B_i^c$ , one should pick either  $(1 - \epsilon)(c + 1)m$  vertices in  $B$  or pick  $X \subseteq A$  and  $\epsilon m$  vertices from  $B$  such that each vertex has  $k + 1$  neighbors in  $X$ . By the property (iv'), the later implies that  $|X| > h$ . Thus, either  $(1 - \epsilon)(c + 1)m$  vertices in  $B$  or  $h$  vertices in  $A$  must be picked in this case. To obtain a constant gap, we assign  $m/k$  weight to every vertex in  $A$  and let  $h = k^2$ ,  $m = k^5$ . It is routine to check that the reduction is FPT.



We remark that  $|B_i|$  can be sufficiently small in the threshold graph construction. Assuming  $W[1] \neq \text{FPT}$ ,  $|U|$  is  $k^{O(1)} \log n$  and thus we can set  $c = \frac{\log n}{k \log \log n}$ . Since  $k$  is arbitrarily small with respect to  $n$ , by choosing appropriate parameters, we can rule out  $o\left(\frac{\log n}{\log \log n}\right)$  factor polynomial time algorithms for the non-parameterized  $k$ -SETCOVER problem.

## 2 Preliminaries

In this section, we formally define the  $k$ -SETCOVER problem and related hypothesis.

For every graph  $G$ , we use  $V(G)$  and  $E(G)$  to denote its vertex set and edge set. For every vertex  $v \in V(G)$ , let  $N(v) \subseteq V(G)$  be the set of neighbors of  $v$  in  $G$ . For every vertex set  $C \subseteq V(G)$ , define  $N(C) = \bigcup_{v \in C} N(v)$ .

The (weighted variant of) SETCOVER is defined as follows. An instance  $\Gamma$  consists of a bipartite graph  $G = (S, U, E)$  and a weight function  $w : S \rightarrow \mathbb{N}^+$ . The goal is to find a set  $C \subseteq S$  such that  $N(C) = U$  and the total weight  $\sum_{s \in C} w(s)$  is minimal. We use  $\text{OPT}(\Gamma)$  to represent the minimum total weight. For every  $\chi \in \mathbb{N}$ , we say an instance  $\Gamma$  is  $\chi$ -weighted if the number of different weights is upper bounded by  $\chi$ , i.e.  $|\{w(v) : v \in S\}| \leq \chi$ . The unweighted SETCOVER is equivalent to 1-weighted SETCOVER.

Given a constant  $\chi$ , by duplicating every vertex according to its weight, any  $\chi$ -weighted SETCOVER instance  $\Gamma$  can be reduced to an unweighted (1-weighted) SETCOVER instance  $\Gamma'$  in time  $(|\Gamma| \cdot \max\{w(s) : s \in S\})^{O(1)}$ , while preserving the optimum.

**Lemma 3** (Lemma 16 in [CL19]). *For any constant  $\chi$ , there is a reduction which, given any  $\chi$ -weighted SETCOVER instance  $\Gamma = (S, U, E, w)$ , outputs an unweighted SETCOVER instance  $\Gamma' = (S', U', E')$  in  $O((|\Gamma| \max\{w(s) : s \in S\})^{O(1)})$  time, such that for every  $k < |\Gamma|$ ,  $\text{OPT}(\Gamma) \leq k$  if and only if  $\text{OPT}(\Gamma') \leq k$ .*

In parameterized complexity theory, we consider problems  $L \subseteq \{0, 1\}^*$  with a computable function  $\kappa : \{0, 1\}^* \rightarrow \mathbb{N}$  which returns a parameter  $\kappa(x) \in \mathbb{N}$  for every input instance  $x \in \{0, 1\}^*$ . Since we mainly focus on graph related problems, it is convenient to treat each input as a pair  $x = (G, k)$  with  $G$  a graph and  $k$  an integer and let  $\kappa(x) = \kappa(G, k) = k$ . A parameterized problem  $(L, \kappa)$  is *fixed parameter tractable* (FPT) if it has an algorithm which for every input  $x \in \{0, 1\}^*$  decides if  $x$  is a yes-instance of  $L$  in  $f(\kappa(x)) \cdot |x|^{O(1)}$ -time for some computable function  $f$ . An FPT-reduction from problem  $(L, \kappa)$  to  $(L', \kappa')$  is an algorithm  $A$  which on every input  $x$ , outputs an instance  $x'$  in  $f(\kappa(x)) \cdot |x|^{O(1)}$ -time such that  $x$  is a yes-instance if and only if  $x'$  is a yes-instance and  $\kappa'(x') = g(\kappa(x))$  for some computable function  $f$  and  $g$ . In the parameterized version of SETCOVER, i.e.,  $k$ -SETCOVER, we are given an instance  $\Gamma$  of unweighted SETCOVER and an integer  $k$  as its parameter (i.e.  $\kappa(\Gamma, k) = k$ ). The goal is to decide if  $\text{OPT}(\Gamma) \leq k$ . The  $c$ -gap  $k$ -SETCOVER problem is to distinguish between the cases  $\text{OPT}(\Gamma) \leq k$  and  $\text{OPT}(\Gamma) > ck$ . Another fundamental parameterized problem is the  $k$ -CLIQUE problem, whose goal is to decide whether an input graph  $G$  contains a clique of size  $k$ .

At the end, we present the hypothesis  $W[2] \neq \text{FPT}$  and  $W[1] \neq \text{FPT}$  on which our results based. For simplicity, we omit the definition of W-Hierarchy and only use an equivalent form.

**Hypothesis 4** ( $W[2] \neq \text{FPT}$ ).  $k\text{-SETCOVER}$  cannot be solved in  $f(k) \cdot n^{O(1)}$  time for any computable function  $f$ .

**Hypothesis 5** ( $W[1] \neq \text{FPT}$ ).  $k\text{-CLIQUE}$  cannot be solved in  $f(k) \cdot n^{O(1)}$  time for any computable function  $f$ .

Analogously to the definition of NP-hardness, we say a problem is  $W[1]$ -hard or  $W[2]$ -hard if there is an FPT-reduction from  $k\text{-CLIQUE}$  or  $k\text{-SETCOVER}$  to it, respectively. Similarly, a problem is in  $W[1]$  or  $W[2]$  if there is an FPT-reduction from it to  $k\text{-CLIQUE}$  or  $k\text{-SETCOVER}$ , respectively. Note that  $W[1] \neq \text{FPT}$  implies  $W[2] \neq \text{FPT}$ , because there is an FPT reduction which transforms a  $k\text{-CLIQUE}$  instance to an unweighted  $k\text{-SETCOVER}$  instance  $\Gamma$  with small universe size  $|U| = k^3 \cdot \log |\Gamma|$  [KLM19, Lin19]. Thus, in our paper, the starting point of our reduction for Theorem 2 is an unweighted  $k\text{-SETCOVER}$  with small universe size rather than a  $k\text{-CLIQUE}$  instance. Formally, the reduction [KLM19, Lin19] is stated below.

**Lemma 6.** *There is a polynomial time reduction which, given an  $n$ -vertex graph  $G = (V, E)$  and a parameter  $k$ , outputs a  $k\text{-SETCOVER}$  instance  $\Gamma = (S, U, E)$  where  $|U| = O(k^3 \log n)$  and  $|S| = |E| = O(n^2)$ , such that:*

- if  $G$  contains a  $k$ -clique, then  $\text{OPT}(\Gamma) \leq \binom{k}{2}$ ;
- if  $G$  contains no  $k$ -clique, then  $\text{OPT}(\Gamma) > \binom{k}{2}$ .

### 3 Strong Threshold Graphs

In this part, we introduce a combinatorial object called strong threshold graph, which plays an important role in our proof. Intuitively, we compose strong threshold graphs with the original  $k\text{-SETCOVER}$  instance to produce gaps. Our threshold graph construction comes from [KN21]. However, we take a new analysis to this construction and establish the stronger threshold property on this type of graphs. Since the construction involves error correcting codes. For the sake of self-containedness, we first put the definition of error correcting codes here. Then, we formally define the strong threshold property used in this work, and finally, we use a new analysis to the construction in [KN21] and prove (See Theorem 10) that it is a strong threshold graph.

**Definition 7** (Error Correcting Codes). *Let  $\Sigma$  be a finite set, a subset  $\mathcal{C} : \Sigma^r \rightarrow \Sigma^m$  is an error correcting code with message length  $r$ , block length  $m$  and relative distance  $\delta$  if for every  $x, y \in \Sigma^r$ ,  $\Delta(\mathcal{C}(x), \mathcal{C}(y)) \geq \delta$ . We denote then  $\Delta(\mathcal{C}) = \delta$ . Here  $\Delta(x, y) = \frac{1}{m} |\{i \in [m] : x_i \neq y_i\}|$ .*

We sometimes abuse notations a little and treat an error correcting code as its image, i.e.,  $\mathcal{C} \subseteq \Sigma^m$  and  $|\mathcal{C}| = \Sigma^r$ .

Throughout our paper, we use Reed-Solomon Code (RS code). Let  $\Sigma$  be a field and  $r \leq m \leq |\Sigma|$ . Fix  $m$  elements  $f_1, \dots, f_m \in \Sigma$ . The RS code  $\mathcal{C}^{\text{RS}} : \Sigma^r \rightarrow \Sigma^m$  is defined as

$$\forall (a_1, \dots, a_r) \in \Sigma^r, \mathcal{C}^{\text{RS}}(a_1, a_2, \dots, a_r) = \left( \sum_{i \in [r]} a_i f_1^{i-1}, \sum_{i \in [r]} a_i f_2^{i-1}, \dots, \sum_{i \in [r]} a_i f_m^{i-1} \right).$$



**Lemma 8** (RS code [RS60]). *Given input and output length  $r, m$  and alphabet  $\Sigma$  such that  $|\Sigma| \geq m$ , the RS code  $\mathcal{C}^{\text{RS}} : \Sigma^r \rightarrow \Sigma^m$  satisfies  $\Delta(\mathcal{C}^{\text{RS}}) \geq 1 - \frac{r}{m}$ .*

Below we describe the properties of threshold graphs we need. Note that in Section 1.1 we have compared our threshold property with those in previous work [Lin19].

**Definition 9** ( $(n, k, t, m, h, \varepsilon)$ -threshold graph). *Given  $n, k, t, m, h \in \mathbb{N}^+$  and  $\varepsilon \in (0, 1)$ , a bipartite graph  $T = (A \dot{\cup} B, E)$  is an  $(n, k, t, m, h, \varepsilon)$ -threshold graph if it has the following properties:*

- *A consists of  $k$  disjoint parts  $A = A_1 \dot{\cup} A_2 \dot{\cup} \dots \dot{\cup} A_k$  with  $|A_i| = n$  for all  $i \in [k]$ .*
- *B consists of  $m$  disjoint parts  $B = B_1 \dot{\cup} B_2 \dot{\cup} \dots \dot{\cup} B_m$  with  $|B_i| = t$  for all  $i \in [m]$ .*
- *For any  $a_1 \in A_1, a_2 \in A_2, \dots, a_k \in A_k$  and every  $j \in [m]$ , there is a vertex  $b \in B_j$  which is a common neighbor of  $\{a_1, \dots, a_k\}$ .*
- *For any  $X \subseteq A$  and  $b_1 \in B_1, \dots, b_m \in B_m$ , if there are  $\varepsilon m$  indices  $j$  such that  $|N(b_j) \cap X| \geq k + 1$ , then  $|X| > h$ .*

At the end, we analyze the construction in [KN21] and prove that it is a strong threshold graph. Formally, we have that:

**Theorem 10.** *Given an error correcting code  $\mathcal{C} : \Sigma^r \rightarrow \Sigma^m$  with distance  $\delta$ , then for any  $k \in \mathbb{N}$  and  $\varepsilon \in (0, 1)$ , a  $(|\Sigma|^r, k, |\Sigma|^k, m, \sqrt{\frac{2\varepsilon}{1-\delta}}, \varepsilon)$ -threshold graph  $G$  can be constructed in time  $O((k + m) \cdot |\Sigma|^{O(r+k)})$ .*

*Proof.* The construction is as follows.

- For every  $i \in [k]$ ,  $A_i = \{\mathcal{C}(x) : x \in \Sigma^r\}$ .
- For every  $j \in [m]$ ,  $B_j = \Sigma^k$ .
- A vertex  $a \in A_i$  and a vertex  $b \in B_j$  are linked if and only if  $(a)_j = (b)_i$ , where  $(a)_j$  means the  $j$ -th element of vector  $a$ .

Fix any  $a_1 \in A_1, a_2 \in A_2, \dots, a_k \in A_k$  and an index  $j \in [m]$ . Let  $b = ((a_1)_j, (a_2)_j, \dots, (a_k)_j)$ . It's easy to see  $b \in B_j$  is a common neighbor of  $a_1, a_2, \dots, a_k$ .

Fix any  $b_1 \in B_1, b_2 \in B_2, \dots, b_m \in B_m$  and  $X \subseteq A$ . Suppose there are  $\varepsilon m$  indices  $j$  such that  $|N(b_j) \cap X| \geq k + 1$ . Since  $A$  is divided into  $k$  parts, for each such index  $j$ , there must be an index  $i \in [k]$  such that  $N(b_j)$  contains at least two vertices in  $X \cap A_i$ . Let  $x, x' \in X \cap A_i$  be two different vertices, then  $(x)_j = (x')_j = (b_j)_i$ . Define

$$L_{x,x'} = \{i \in [m] : (x)_i = (x')_i\},$$

we have

$$\sum_{x \neq x' \in X} |L_{x,x'}| \geq \varepsilon m.$$

However, according to the distance of codewords, for every  $x, x' \in X, x \neq x'$ ,

$$|L_{x,x'}| \leq (1 - \delta)m.$$

This leads to  $\binom{|X|}{2}(1 - \delta) \geq \varepsilon$ , i.e.,  $|X| > \sqrt{\frac{2\varepsilon}{1-\delta}}$ . □

## 4 Proof of the Main Theorem

**Theorem 11.** *There is a reduction which, given an unweighted  $k$ -SETCOVER instance  $\Gamma = (S, U, E)$ , an  $(n, k, t, m, h, \varepsilon)$ -threshold graph  $T = (A \dot{\cup} B, E_T)$  where  $n = |S|$  and  $m \leq n^{O(1)}$ , and an integer  $c \in \mathbb{N}^+$ , outputs a new 2-weighted  $k$ -SETCOVER instance  $\Gamma' = (S', U', E', w)$  with the following properties:*

- (Completeness) *If  $\text{OPT}(\Gamma) \leq k$ , then  $\text{OPT}(\Gamma') \leq 2m$ .*
- (Soundness) *If  $\text{OPT}(\Gamma) > k$ , then  $\text{OPT}(\Gamma') > \min\{mh/k, (1 - \varepsilon)mc\}$ .*
- *The reduction runs in  $|\Gamma|^{O(1)} \cdot (|U|t)^{O(c)}$  time.*

*Proof.* Let  $A = A_1 \dot{\cup} A_2 \dot{\cup} \dots \dot{\cup} A_k$  and  $B = B_1 \dot{\cup} B_2 \dot{\cup} \dots \dot{\cup} B_m$ . For every  $i \in [k]$ , we treat each  $A_i$  as a copy of  $[n]$ . Let  $s : [n] \rightarrow S$  be a bijection. The new instance  $\Gamma' = (S', U', E')$  and  $w$  is defined as follows.

- $S' = A \dot{\cup} B$ .
- $U' = \{(u_1, \dots, u_c, b_1, \dots, b_c, i) : (u_1, \dots, u_c) \in U^c, (b_1, \dots, b_c) \in B_i^c, i \in [m]\}$ .
- For every  $a \in A$ ,  $w(a) = m/k$ . For every  $b \in B$ ,  $w(b) = 1$ .
- For every  $a \in A$  and  $\vec{u} = (u_1, \dots, u_c, b_1, \dots, b_c, i) \in U'$ , we link  $a$  and  $\vec{u}$  if there exists  $j \in [c]$  such that  $(a, b_j) \in E_T$  and  $(s(a), u_j) \in E$ , where  $s(a)$  is the matching vertex of  $a$  in  $S$ .
- For every  $b \in B$  and  $\vec{u} = (u_1, \dots, u_c, b_1, \dots, b_c, i) \in U'$ , we link  $b$  and  $\vec{u}$  if  $b \in B_i$  and  $b \neq b_j$  for all  $j \in [c]$ .

It is easy to see the reduction can be done in  $((nk + tm) \cdot (|U|t)^c \cdot m)^{O(1)} = |\Gamma|^{O(1)} \cdot (|U|t)^{O(c)}$  time.

For the completeness case, let the solution in  $\Gamma$  be  $a_1, \dots, a_k \in S$ . By the property of threshold graph,  $a_1 \in A_1, \dots, a_k \in A_k$  have common neighbors  $b_1 \in B_1, \dots, b_m \in B_m$  in  $T$ .

We claim that  $\{a_1, \dots, a_k, b_1, \dots, b_m\} \subseteq S'$ , which is of weight  $2m$ , is a valid solution of  $\Gamma'$ .

For every  $\vec{u} = (u_1, \dots, u_c, \hat{b}_1, \dots, \hat{b}_c, i) \in U'$  with  $(\hat{b}_1, \dots, \hat{b}_c) \in B_i^c$  for some  $i \in [m]$ ,

- If  $b_i$  does not appear in  $\{\hat{b}_1, \dots, \hat{b}_c\}$ , then according to our construction,  $(b_i, \vec{u}) \in E'$ .
- Otherwise suppose  $b_i = \hat{b}_j$  for some  $j \in [c]$ . Since  $\{a_1, \dots, a_k\}$  is a valid covering of  $U$  in  $\Gamma$ , there exists  $j^* \in [k]$  such that  $(s(a_{j^*}), u_j) \in E$ . Note that  $(a_{j^*}, \hat{b}_j) \in E_T$ , we have  $(a_{j^*}, \vec{u}) \in E'$  by definition.

For the soundness case, consider a solution  $\{a_1, \dots, a_q, b_1, \dots, b_r\}$  with  $a_i \in A$  and  $b_j \in B$ , we first prove that, for every  $i \in [m]$ ,

- Either  $\{b_1, \dots, b_r\}$  contains at least  $c + 1$  vertices in  $B_i$ ,

- Or there exists  $j \in [r]$ , such that  $b_j \in B_i$  is a common neighbor of at least  $k + 1$  vertices in  $\{a_1, \dots, a_q\}$  in the threshold graph  $T$ .

Suppose it is not the first case, i.e., the set  $\{b_1, \dots, b_r\}$  contains no more than  $c$  vertices in  $B_i$ , for simplicity of notation let those vertices be  $\hat{b}_1, \dots, \hat{b}_c$  (we allow duplication so that the number of vertices can always be  $c$ ).

Consider the set  $U_{\hat{b}_1, \dots, \hat{b}_c} = \{(u_1, \dots, u_c, \hat{b}_1, \dots, \hat{b}_c, i) \in U' : (u_1, \dots, u_c) \in U^c\}$ . According to our assumption,  $\{b_1, \dots, b_r\}$  does not cover  $U_{\hat{b}_1, \dots, \hat{b}_c}$ . So  $U_{\hat{b}_1, \dots, \hat{b}_c}$  can only be covered by  $\{a_1, \dots, a_q\}$ . Suppose by contradiction that for every  $j \in [c]$ ,  $\hat{b}_j$  has at most  $k$  neighbors in  $\{a_1, \dots, a_q\}$ , i.e. the set  $N_{\hat{b}_j} = \{a_\ell : \ell \in [q], a_\ell \in N(\hat{b}_j)\}$  has size at most  $k$ . Since  $\text{OPT}(\Gamma) > k$ , there must be some  $\hat{u}_j \in U$  not covered by  $N_{\hat{b}_j}$ . Hence for every  $j \in [c]$  and  $\ell \in [q]$ , either  $(a_\ell, \hat{b}_j) \notin E_T$  or  $(s(a_\ell), \hat{u}_j) \notin E$ , i.e. the vertex  $\vec{u} = (\hat{u}_1, \dots, \hat{u}_c, \hat{b}_1, \dots, \hat{b}_c, i)$  is not covered by  $\{a_1, \dots, a_q\}$ .

If there are  $\varepsilon m$  indices  $i \in [m]$  such that in the threshold graph  $T$ , there exists  $b_j \in B_i$  that has  $k + 1$  neighbors in  $\{a_1, \dots, a_q\}$ , then by the property of threshold graph,  $q > h$ . It follows that  $w(\{a_1, \dots, a_q\}) > h \cdot m/k$ . Otherwise, there are  $(1 - \varepsilon)m$  indices  $i \in [m]$  such that  $\{b_1, \dots, b_r\}$  contains  $c + 1$  vertices in  $B_i$ , we have  $w(\{b_1, \dots, b_r\}) > (1 - \varepsilon)cm$ .  $\square$

Now we are ready to prove the  $W[2]$ -hardness of constant gap  $k$ -SETCOVER.

**Theorem 12** (Restated version of Theorem 1). *Assuming  $W[2] \neq \text{FPT}$ , there is no deterministic FPT algorithm which can approximate  $k$ -SETCOVER within any constant ratio.*

*Proof.* For any constant  $c_0 > 0$ , we give an FPT reduction from a  $k$ -SETCOVER instance  $\Gamma = (S, U, E)$  to a  $c_0$ -gap  $k'$ -SETCOVER instance as follows. Note that without loss of generality, we could assume that our reduction only holds for large enough  $\kappa(x)$  and  $|x|$ , since we could apply enumerative search to solve small instances. We assume that  $k \geq 2c_0$ ,  $|S| \geq k^{5k}$  and  $|U| \leq |S|$ . Let

- $n = |S|$ ,
- $r = k$ ,
- $m = k^5$ ,
- $c = 4c_0$ ,
- $|\Sigma| = n^{1/k} \geq m$ ,
- $\varepsilon = 1/2$ .

By Lemma 8, the Reed-Solomon code  $\mathcal{C}^{\text{RS}} : \Sigma^r \rightarrow \Sigma^m$  has distance at least  $\delta = 1 - \frac{r}{m} = 1 - \frac{1}{k^4}$ . Then by Theorem 10, we can construct a  $(|\Sigma|^r = n, k, |\Sigma|^k = n, m = k^5, \sqrt{\frac{2 \cdot \varepsilon}{1 - \delta}} = k^2, \varepsilon = \frac{1}{2})$ -threshold graph in  $O((k + m) \cdot |\Sigma|^{O(r+k)}) = n^{O(1)}$  time. After that, we apply Theorem 11 to get a 2-weighted gap SETCOVER instance in time  $(n^{O(1)}) \cdot (n^2)^{O(c)} = n^{O(c)}$ , where in the yes case, the optimal solution has size at most  $2m$ , and in the no case, the optimal solution has size greater than  $\min\{m \cdot k^2/k, (1 - \varepsilon) \cdot mc\} = 2c_0m$ . Finally, we apply Lemma 3 to remove the weights in

$\left(n^{O(c)} \cdot \frac{m}{k}\right)^{O(1)} = n^{O(c)}$  time. The whole reduction runs in time polynomial in  $n$ , and the new parameter  $2m = 2k^5$  is a function of  $k$ . Thus our reduction is an FPT reduction.  $\square$

Next we prove the hardness of approximating non-parameterized  $k$ -SETCOVER without using the PCP theorem.

**Theorem 13** (Restated version of Theorem 2). *Assuming  $W[1] \neq \text{FPT}$ , there is no polynomial time algorithm which can approximate non-parameterized  $k$ -SETCOVER within  $o\left(\frac{\log n}{\log \log n}\right)$  ratio, even if  $k$  is as small as  $O\left(\frac{\log n}{\log \log n}\right)^3$ .*

*Proof.* Suppose by contradiction there is an algorithm which can approximate non-parameterized  $k$ -SETCOVER within a ratio less than  $\frac{\log n}{g(n)\log \log n}$  for some  $g(n) = \omega(1)$  in  $n^{O(1)}$  time, then we give an algorithm which can solve  $k$ -CLIQUE in FPT time. W.l.o.g. assume  $g$  is a non-decreasing function.

Given a  $k$ -CLIQUE instance  $G = (V, E)$ , we first reduce it to a  $k'$ -SETCOVER instance  $\Gamma = (S, U, E)$  where  $|U| = O((k')^2 \log |S|)$  by Lemma 6. We will use  $k$  for  $k'$  in the following for clarity. W.l.o.g. assume  $n = |S|$  is large enough that

$$\min \left\{ g(n), \frac{\log n}{\log \log n} \right\} \geq \max\{2k, 100\},$$

otherwise we use brute-force to solve  $\Gamma$ , and thus solve  $G$  in FPT time. Let

- $r = \frac{\log n}{\log \log n}$ ,
- $m = \left(\frac{\log n}{\log \log n}\right)^3$ ,
- $c = \frac{\log n}{k \log \log n}$ ,
- $|\Sigma| = \left(\frac{\log n}{\log \log n}\right)^3$ ,
- $\varepsilon = 1/2$ .

By Lemma 8, the Reed-Solomon code  $\mathcal{C}^{RS} : \Sigma^r \rightarrow \Sigma^m$  has distance at least  $\delta = 1 - \frac{r}{m} = 1 - \left(\frac{\log \log n}{\log n}\right)^2$ . Then by Theorem 10, we can construct a  $(|\Sigma|^r = n^{O(1)}, k, |\Sigma|^k = \left(\frac{\log n}{\log \log n}\right)^{3k}, m = k^5, \sqrt{\frac{2 \cdot \varepsilon}{1 - \delta}} = \frac{\log n}{\log \log n}, \varepsilon = \frac{1}{2})$ -threshold graph in  $O((k + m) \cdot |\Sigma|^{O(r+k)}) = n^{O(1)}$  time. After that, we apply Theorem 11 to get a 2-weighted gap SETCOVER instance in time  $n^{O(1)} \cdot \left(k^2 \log n \cdot \left(\frac{\log n}{\log \log n}\right)^{3k}\right)^c = n^{O(1)}$ , where in the yes case, the optimal solution has size at most  $2m = 2 \left(\frac{\log n}{\log \log n}\right)^3$  and in the no case, the optimal solution has size greater than  $\min\{m \cdot \frac{\log n}{\log \log n} / k, (1 - \varepsilon) \cdot mc\} = \frac{\log n}{2k \log \log n}$ . We apply Lemma 3 to remove the weights in  $n^{O(1)}$  time. Finally, as  $\frac{\log n}{2k \log \log n} \geq \frac{\log n}{g(n) \log \log n}$ , the presumed approximation algorithm for non-parameterized  $k$ -SETCOVER can distinguish between the two cases in  $n^{O(1)}$  time, and thus solve  $k$ -CLIQUE in FPT time, contradicting  $W[1] \neq \text{FPT}$ .  $\square$

## 5 Conclusion

In this paper, we settle the  $W[2]$ -hardness of approximating  $k$ -SETCOVER with constant approximation ratio. Our result could also be applied to rule out polynomial algorithm approximating non-parameterized  $k$ -SETCOVER within ratio  $o\left(\frac{\log n}{\log \log n}\right)$ , with  $k$  as small as  $O\left(\frac{\log n}{\log \log n}\right)^3$ , assuming  $W[1] \neq \text{FPT}$ .

As further research questions, it is interesting to consider  $W[2]$ -hardness of approximating  $k$ -SETCOVER beyond constant ratio.

**Question 1.** *Is it  $W[2]$ -hard to approximate  $k$ -SETCOVER with super-constant ratio?*

Our technique fails to answer this question, because given approximation ratio  $c$ , our reduction runs in time  $\Omega(|U|^c)$ , which will result in a non-FPT reduction if  $c$  is not constant.

It is well-known that the textbook greedy algorithm approximates  $k$ -SETCOVER with ratio  $O(\log n)$ , and this is also the best approximation algorithm. However, assuming  $W[1] \neq \text{FPT}$ , the state-of-the-art result [Lin19] rules out FPT algorithms approximating  $k$ -SETCOVER within approximation ratio  $\left(\frac{\log n}{\log \log n}\right)^{1/k}$ , and there is still a gap. Could we further improve this inapproximability, or does there exist FPT algorithms approximating  $k$ -SETCOVER with better approximation ratio? This leads to the following question:

**Question 2.** *Is there any FPT algorithm approximating  $k$ -SETCOVER within approximation ratio  $o(\log n)$ ?*

Finally, we ask whether the lower bound of our Theorem 2 could be improved under stronger assumptions (e.g., SETH), which leads to the following question:

**Question 3.** *For  $k = (\log n)^{O(1)}$ , assuming SETH, is there any algorithm with running time  $n^{k-o(1)}$  approximating  $k$ -SETCOVER with approximation ratio  $o\left(\frac{\log n}{\log \log n}\right)$ ?*

## References

- [AMS06] Noga Alon, Dana Moshkovitz, and Shmuel Safra. Algorithmic construction of sets for  $k$ -restrictions. *ACM Transactions on Algorithms (TALG)*, 2(2):153–177, 2006.
- [ARW17] Amir Abboud, Aviad Rubinfeld, and Ryan Williams. Distributed PCP theorems for hardness of approximation in P. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 25–36. IEEE, 2017.
- [BBE<sup>+</sup>21] Arnab Bhattacharyya, Édouard Bonnet, László Egri, Suprovat Ghoshal, Karthik C. S., Bingkai Lin, Pasin Manurangsi, and Dániel Marx. Parameterized intractability of even set and shortest vector problem. *J. ACM*, 68(3):16:1–16:40, 2021.
- [CCK<sup>+</sup>17] Parinya Chalermsook, Marek Cygan, Guy Kortsarz, Bundit Laekhanukit, Pasin Manurangsi, Danupon Nanongkai, and Luca Trevisan. From gap-eth to fpt-inapproximability: Clique, dominating set, and more. In Chris Umans, editor, *58th*

- IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 743–754. IEEE Computer Society, 2017.
- [Chv79] V. Chvátal. A greedy heuristic for the set-covering problem. *Mathematics of Operations*, 4(3):233 – 235, 1979.
- [CL19] Yijia Chen and Bingkai Lin. The constant inapproximability of the parameterized dominating set problem. *SIAM J. Comput.*, 48(2):513–533, 2019.
- [DF95] Rodney G. Downey and Michael R. Fellows. Fixed-parameter tractability and completeness I: basic results. *SIAM J. Comput.*, 24(4):873–921, 1995.
- [DS14] I. Dinur and D. Steurer. Analytical approach to parallel repetition. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 624–633, 2014.
- [Fei98] Uriel Feige. A threshold of  $\ln n$  for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.
- [FKLM20] Andreas Emil Feldmann, Karthik C. S., Euiwoong Lee, and Pasin Manurangsi. A survey on approximation in parameterized complexity: Hardness and algorithms. *Algorithms*, 13(6):146, 2020.
- [Joh74] D. S. Johnson. Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, 9(3):256–278, 1974.
- [JP17] Bart M. Jansen and Astrid Pieterse. Sparsification upper and lower bounds for graph problems and not-all-equal sat. *Algorithmica*, 79(1):3–28, sep 2017.
- [Kar72] R. M. Karp. Reducibility among combinatorial problems. In *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.*, pages 85–103, 1972.
- [KLM19] Karthik C. S., Bundit Laekhanukit, and Pasin Manurangsi. On the parameterized complexity of approximating dominating set. *J. ACM*, 66(5):33:1–33:38, 2019.
- [KN21] Karthik C. S. and Inbal Livni Navon. On hardness of approximation of parameterized set cover and label cover: Threshold graphs from error correcting codes. In Hung Viet Le and Valerie King, editors, *4th Symposium on Simplicity in Algorithms, SOSA 2021, Virtual Conference, January 11-12, 2021*, pages 210–223. SIAM, 2021.
- [KZ96] Sandi Klavžar and Blaž Zmazek. On a vizing-like conjecture for direct product graphs. *Discrete Math.*, 156(1–3):243–246, sep 1996.
- [Lin18] Bingkai Lin. The parameterized complexity of the  $k$ -biclique problem. *J. ACM*, 65(5):34:1–34:23, 2018.
- [Lin19] Bingkai Lin. A simple gap-producing reduction for the parameterized set cover problem. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming, ICALP*



- 2019, July 9-12, 2019, Patras, Greece, volume 132 of *LIPICs*, pages 81:1–81:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.
- [Lov75] L. Lovász. On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4):383 – 390, 1975.
- [LY94] C. Lund and M. Yannakakis. On the hardness of approximating minimization problems. *Journal of the ACM*, 41(5):960–981, 1994.
- [RS60] I. S. Reed and G. Solomon. Polynomial codes over certain finite fields. *Journal of the Society for Industrial and Applied Mathematics*, 8(2):300–304, 1960.
- [RS97] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computing, STOC 1997, El Paso, Texas, USA, May 4-6, 1997*, pages 475–484, 1997.
- [Sla97] P. Slavík. A tight analysis of the greedy algorithm for set cover. *Journal of Algorithms*, 25(2):237 – 254, 1997.
- [Ste74] S. K. Stein. Two combinatorial covering theorems. *Journal of Combinatorial Theory, Series A*, 16(3):391–397, 1974.

## Appendix

We prove that, for any computable function  $f$ ,  $k$ -SETCOVER with universe set size at most  $f(k) \log n$  is in  $W[1]$ . We show a reduction from such a  $k$ -SETCOVER instance to a  $k'$ -CLIQUE instance. The idea is as follows. For each element in  $U$ , it should be covered by some set from one of the  $k$  groups. We divide  $U$  into  $f(k) \cdot \log k$  parts of size  $\log n / \log k$ , and use a variable in  $[k]^{\log n / \log k}$  for each part to encode which groups the sets covering those  $\log n / \log k$  elements are from. For each group of sets, we also use a variable to indicate which set is picked from this group. We add constraints between the two types of variables to check whether the alleged set is eligible for covering corresponding elements, and transform this 2CSP instance to a CLIQUE instance in the canonical way.

**Lemma 14.** *There is an FPT-reduction which, given an instance  $\Gamma = (S, U, E)$  of  $k$ -SETCOVER with  $|U| = f(k) \cdot \log n$ , outputs a graph  $G$  and an integer  $k' = k + f(k) \cdot \log k$ , such that*

- if  $\text{OPT}(\Gamma) = k$ , then  $G$  contains a  $k'$ -clique,
- if  $\text{OPT}(\Gamma) < k$ , then  $G$  contains no  $k'$ -clique.

*Proof.* Let  $h = f(k) \cdot \log k$ . We divide  $U$  into  $h$  groups of size  $\log n / \log k$  and index every element in  $U$  by a pair  $(i, j)$  where  $i \in [h], j \in [\log n / \log k]$ . We define a graph  $G$  as follows.

- $V(G) = V_1 \cup V_2, \dots, V_k \cup W_1 \cup W_2, \dots, W_h$ .

- For every  $i \in [k]$ ,  $V_i$  is a copy of  $S$ .
- For every  $i \in [h]$ ,  $W_i = [k]^{\log n / \log k}$ .
- Make each  $V_i$  and  $W_j$  an independent set. Add edges between different  $V_i$  and  $V_j$ , and between different  $W_i$  and  $W_j$ .
- For every  $v \in V_i$  and  $w \in W_j$ , add an edge between them if and only if for all  $\ell \in [\log n / \log k]$ ,  $w[\ell] = i$  implies  $v$  can cover the  $(j, \ell)$ -th element in  $U$ .

The running time of this reduction is at most  $O(kh \cdot k^{\log n / \log k} \cdot |G| \cdot |U|) = k \log k \cdot f(k) \cdot n^{O(1)}$ .

Suppose  $\Gamma$  has a size- $k$  solution  $v_1, \dots, v_k \in S$ . We define  $w_1 \in W_1, \dots, w_h \in W_h$  as follows. For every  $j \in [h]$  and  $\ell \in [\log n / \log k]$ , let  $w_j[\ell] = i$  such that  $v_i$  can cover the  $(j, \ell)$ -th element in  $U$ . By our construction, each  $w_j$  is adjacent to all  $v_1, \dots, v_k$ . Thus, we obtain a  $k'$ -clique in graph  $G$ .

Suppose  $G$  contains a clique  $X$  of size  $k'$ . By our construction,  $|X \cap V_i| = |X \cap W_j| = 1$  for every  $i \in [k], j \in [h]$ . Let  $v_i \in |X \cap V_i|, w_j \in |X \cap W_j|$  be the vertices in the clique. For every  $j \in [h]$  and  $\ell \in [\log n / \log k]$ , let  $i = w_j[\ell]$ , then by the existence of an edge between  $v_i$  and  $w_j$ , the  $(j, \ell)$ -th element in  $U$  can be covered by  $v_i$ . Thus,  $U$  can be covered by  $k$  sets  $v_1, \dots, v_k$ .  $\square$