Private Approximation of Search Problems*

Amos Beimel† Paz Carmi Kobbi Nissim Enav Weinreb

Department of Computer Science
Ben-Gurion University, Beer-Sheva, Israel
Email: {beimel,carmip,kobbi,weinrebe}@cs.bgu.ac.il

November 28, 2006

Abstract

Many approximation algorithms have been presented in the last decades for hard search problems. The focus of this paper is on cryptographic applications, where it is desired to design algorithms which do not leak unnecessary information. Specifically, we are interested in private approximation algorithms – efficient algorithms whose output does not leak information not implied by the optimal solutions to the search problems. Privacy requirements add constraints on the approximation algorithms; in particular, known approximation algorithms usually leak a lot of information.

For functions, [Feigenbaum et al., ICALP 2001] presented a natural requirement that a private algorithm should not leak information not implied by the original function. Generalizing this requirement to search problems is not straightforward as an input may have many different outputs. We present a new definition that captures a minimal privacy requirement from such algorithms – applied to an input instance, it should not leak any information that is not implied by its *collection of exact solutions*. Although our privacy requirement seems minimal, we show that for well studied problems, as vertex cover and max exact 3SAT, private approximation algorithms are unlikely to exist even for poor approximation ratios. Similar to [Halevi et al., STOC 2001], we define a relaxed notion of approximation algorithms that leak (little) information, and demonstrate the applicability of this notion by showing near optimal approximation algorithms for max exact 3SAT that leak little information.

Key words. Secure computation, Private approximation, Solution-list algorithms, Vertex cover.

*An extended abstract of this work appeared in Proc. of the 38th Annu. ACM Symp. on the Theory of Computing (STOC), 2006. This work was partially supported by the Frankel Center for Computer Science.

†Part of this work was done while the author was on sabbatical at the University of California, Davis, partially supported by the David and Lucile Packard Foundation.

1 Introduction

Approximation algorithms are currently one of the main research fields in computer science. The design of algorithms for approximating computationally hard problems has attracted substantial attention in the last few decades, as did the research on proving hardness of approximation. Frequently, approximation algorithms are applied to sensitive data, as in the distributed cryptographic setup of secure computation. In this paper we study privacy issues related to approximation algorithms of search problems.

We consider an abstract client-server setting. This scenario, besides being interesting on its own sake, is important since the multiparty distributed setting can be reduced to this model using secure function evaluation protocols [25, 10] (see discussion below). In the client-server setting, the server $\mathcal{S}$ is willing to let the client $\mathcal{C}$ learn a specific functionality f of its input. The standard requirement in this setting is that no other information would be leaked to $\mathcal{C}$. For concreteness, assume that $\mathcal{S}$ holds a graph G and consider the following examples:

1. If $f(G)$ is the diameter of G, then $\mathcal{S}$ can simply compute $f(G)$ and send it to $\mathcal{C}$. It is clear that $\mathcal{C}$ learns $f(G)$, but no other information about G.

2. If $f(G)$ is the number of perfect matchings in G, which is a well known $\#\mathcal{P}$-complete problem, then $\mathcal{C}$ and $\mathcal{S}$ have to settle for an approximation $\hat{f}$. This raises the question of which approximation to use, as $\mathcal{S}$ is only willing to reveal $f(G)$, but $\hat{f}$ may leak some other information. This was the setting in the work by Feigenbaum et al. [8]. They defined the notion of *private approximation* that combines two requirements: (i) *approximation:* $\hat{f}$ is an approximation to f, and (ii) *functional privacy:* $\hat{f}(G)$ can be simulated given $f(G)$; In particular, functional privacy implies that if $f(x) = f(y)$, then $\hat{f}(x)$ and $\hat{f}(y)$ are indistinguishable. It turns out that an efficient private approximation algorithm exists for the number of perfect matchings in a graph [8].

3. A more general and more typical case is when the server $\mathcal{C}$ needs to learn a "solution" for an optimization problem, rather than its optimization objective. For example, one is usually interested in finding a vertex cover of minimum size in G, rather than just learning the size of the optimal vertex cover. However, even if $\mathcal{C}$ and $\mathcal{S}$ are willing to settle for an approximation (i.e. finding a cover which is not much larger than the minimum), it is not clear which cover $\mathcal{C}$ should learn – the framework of private approximations of functions does not address this problem as there may be many solutions to the search problem. This more general case where one seeks not a computation of a function, but of a solution to a computational problem is the focus of this work.

The generalization of the definition of (functional) private approximation to search problems is not straightforward. As one input may have many different outputs, it is not clear in which cases we need $\hat{f}(x)$ and $\hat{f}(y)$ to be indistinguishable. For our example of vertex cover, it seems that a minimal requirement is that a private approximation algorithm should not distinguish between graphs G_1, G_2 that are equivalent in the sense that they have exactly the same set of solutions, i.e., every minimum cover for G_1 is also a minimum cover for G_2 and vice versa. Note that this is a rather weak requirement, as it does not restrict the approximation algorithm with respect to non-equivalent graphs. Furthermore, it might leak more information than one solution to the problem. Notice that the number of equivalence classes of graphs is exponential in $|V|$, thus, there may be many equivalence classes that contain only a few graphs. This is in sharp contrast with functional privacy that divides the graphs to only $|V|$ equivalence classes – as there are $|V|$ possible answers to the functional problem.

We emphasize that impossibility results for private approximation of the optimization objective value (e.g., approximating the size of a minimal vertex cover) do not imply impossibility results for the task of finding a solution with near optimal optimization value (e.g., finding a small vertex cover). For non private computation, if there is an efficient algorithm whose output is a near optimal solution, then one can compute its value. However, this is not a private reduction as it leaks more information than implied by the optimization value. Furthermore, for private approximation the above two tasks are incomparable as the privacy requirement of "not learning information" is applied to a different output. For example, consider two graphs G_1 and G_2 such that the size of a minimum vertex covers of G_1 and G_2 are the same, but the sets of minimum covers are different. A private approximation algorithm for the search problem of vertex cover can return covers of different size, while a private approximation algorithm for the size of the vertex cover must return the same answer for G_1 and G_2.

To understand the nature of private approximation of search problems, we concentrate in this work on two optimization problems with different characteristics – vertex cover and max exact 3SAT. The first problem is a minimization problem while the second is a maximization problem. More importantly, for vertex cover we want a solution satisfying *all* constraints (covering all edges) and we compromise by allowing a solution whose size is not optimal. In contrast, for max exact 3SAT we seek a solution satisfying *most* constraints. The same methods we present in this paper for these two problems are applicable for other optimization problems (e.g., max-cut).

We show that vertex cover cannot be approximated privately even to an approximation ratio as poor as $n^{1-\epsilon}$. A similar result is shown for max exact 3SAT. This means that although our notion of privacy is minimal and it seems that any reasonable notion of privacy for search problems should imply it, there are natural problems for which it is too strong.

Our proof techniques for the impossibility results are different from those used for obtaining the inapproximability results for the functional version of the problem [14]. We show how to use a private approximation algorithm for a problem in order to solve the problem *exactly* in polynomial time. All our lower bounds have the same structure, where a solution is constructed in an iterative manner. For example, for vertex cover, in each iteration a node is examined. If that node appears in some optimal solution, we add it to the solution and remove it and its neighbors from the graph. If there exists some optimal solution that does not include this node, we remove it from the graph. When both conditions are met, we choose one of them arbitrarily. The crux of the algorithm is that the private approximation algorithm is used for deciding which of the conditions hold.

In view of the impossibility results, it is natural to look for a relaxation of the definition. In the setting of functional privacy, Halevi et al. [14] defined the notion of *almost private algorithms* that are allowed to leak (little) information beyond what is leaked by the exact functionality. This notion falls elegantly within our definitional framework, where it is generalized to search problems. We say that an algorithm leaks at most k bits if it refines each equivalence class by dividing it to at most 2^k sub-classes. For max exact 3SAT, this relaxation results in a tremendous improvement, and there exists an efficient approximation algorithm for max exact 3SAT with a near optimal approximation ratio of $7/8 - \epsilon$ that leaks only $\log \log n$ bits of information.

Interestingly, the algorithm for max exact 3SAT and an algorithm we describe for vertex cover fall into a class of approximation algorithms that we call *solution-list algorithms*. This class of algorithms provides a much stronger privacy guarantee. Intuitively, for every input size n, the solution-list algorithm deterministically computes a list of possible outcomes. Upon seeing the actual input, the algorithm is restricted to output a solution from the list. Thus, the number of bits the algorithm leaks is at most the logarithm of the size of the list. For max exact 3SAT, our algorithm efficiently computes a list of length $O(\log n)$ assignments such that for every exact 3 CNF formula ϕ there exists an assignment in the list that is a good approximation for ϕ, hence the algorithm leaks $O(\log \log n)$ bits. For vertex cover, we obtain a solution-list algorithm that is also significant, but not as dramatic as for max exact 3SAT, and we obtain a tradeoff between the amount of leakage and approximation quality.

We also show an impossibility result for approximating vertex cover while leaking at most $O(\log n)$ bits. This suggests that the solution-list algorithm for the problem may be optimal. However, there is still an exponential gap between the lower and upper bounds on the number of bits leaked by a polynomial time algorithm. Finding algorithmic techniques, possibly stronger than solution-list, for private approximation algorithms is an open problem.

The notion of private search is applicable also to search problems in $\mathcal{P}$. For example, suppose a server holds a graph G and a client wants to learn a maximal matching of this graph. What kind of information can the client derive from the server's output? Can the client rule-out many input graphs given the answer? It is somewhat surprising that the problem of private computation of search problems in $\mathcal{P}$ was not considered in previous papers. For many problems in $\mathcal{P}$, there is an efficient private algorithm solving the search problem. One option is choosing the lexicographically first exact solution (e.g., for maximum matching and shortest path). Another option is constructing a randomized algorithm that chooses a random exact solution according to some distribution. However, for some problems finding a solution privately imposes additional constraints on the algorithm. We show that these constraints can make the problem much harder. We present a search problem which can be easily solved without privacy constraints, but cannot be efficiently and privately solved unless $\mathcal{NP} \subseteq \mathcal{P}/\text{poly}$. We believe that more research should be conducted on private solutions for search problems in $\mathcal{P}$.

Multiparty Approximations of Search Problems. As mentioned before, the abstract client-server setting we consider in this paper is important since the multiparty distributed setting can be reduced to this model. Consider the following scenario in the multiparty model: There is some secret input shared by the parties (that is, all parties together can compute the input, while smaller sets of parties have no information on the input). The parties want to solve some search problem on the input and make the output public without disclosing extra information. Clearly, our impossibility results hold for this model as well. Furthermore, using general secure function evaluation protocols [25, 10], efficient private algorithms in the server/client model can be transformed to protocols in the multiparty model. That is, the efficient algorithm and the sharing of the secret input imply that there is a small circuit, whose input are the inputs of the parties, computing the private functionality. Using the constructions of [25, 10], there is a private protocol computing this private functionality.

1.1 Related Work

Feigenbaum et al. [8] initiated the discussion of private approximation of functions. They observed that combining approximation algorithms and secure function evaluation protocols might result in protocols that are not private as the output of the approximation algorithm might leak information. The definition of functional privacy put forward by [8] is a simulation based definition, where the simulator's input is the exact value $f(x)$ and its output distribution is computationally indistinguishable from $\hat{f}(x)$. Under this definition, they provided a protocol for approximating the Hamming distance of two n-bit strings with communication complexity $\tilde{O}(\sqrt{n})$, and polynomial solutions for approximating the permanent and other natural $\#\mathcal{P}$ problems. Other private approximation protocols published since include [19, 9, 17]. In particular, Indyk and Woodruff [17] provide a polylogarithmic communication approximation for the Hamming distance and a secure approximation of the near neighbor search problem.

Inapproximability results for computing the *size* of a minimum vertex cover within approximation $n^{1-\epsilon}$ were proved by Halevi et al. [14]. Their proof uses a special *sliding-window* reduction that translates a SAT instance ϕ to an instance G of vertex cover such that if ϕ is satisfiable, then G has a vertex cover of size z, and otherwise any vertex cover for G is of size at least $z + 1$. These techniques do not apply in our setting, as the large number of equivalence classes does not allow a simple averaging argument, as used in [14].

The notion of almost private approximation was introduced in [14]. Their definition modifies that of [8] by allowing the simulator to consult a deterministic predicate of the input. They showed that by this slight compromise in privacy, one can get fairly good approximations for any problem that admits a good deterministic approximation. For the functional version of vertex cover this yields an approximation ratio 4 (more generally, there is a tradeoff between the leakage and approximation ratio). A similar relaxation of privacy is the notion of additional information in secure two-party protocols by Bar-Yehuda et al. [3]. Related ideas can be found in the study of knowledge complexity [13, 12, 5, 11, 24].

We next compare the possibility and impossibility results for vertex cover and max exact 3SAT to non-private computation and privately approximating the size of the solutions. The vertex cover problem can be (non-privately) $(2 - o(1))$-approximated in deterministic polynomial time [21, 4, 15], and if $\mathcal{P} \neq \mathcal{NP}$, then there is no polynomial-time 1.3606-approximation algorithm for vertex cover [7]. In contrast, we show that if $\mathcal{RP} \neq \mathcal{NP}$, then this problem cannot be privately $n^{1-\epsilon}$-approximated (even if a leakage of $\log n$ bits is allowed). The max exact 3SAT problem can be $7/8$-approximated [18] and if $\mathcal{P} \neq \mathcal{NP}$ there is no polynomial-time $(7/8 + \epsilon)$-approximation algorithm for it [16]. In contrast, we show that if $\mathcal{RP} \neq \mathcal{NP}$, then this problem cannot be privately $n^{1-\epsilon}$-approximated, however, with $O(\log \log n)$ leakage there is a $(7/8 - \epsilon)$-approximation algorithm for max exact 3SAT. By [14], the (non-private) approximation algorithms imply that the size of the optimal solutions of these problems can be privately approximated with a constant factor leaking one bit. Finally, if $\mathcal{NP} \not\subseteq \mathcal{BPP}$, then for every constant $\epsilon > 0$ there is no private $n^{1-\epsilon}$-approximation algorithm for the size of the minimum vertex cover [14].

Organization. In Section 2 we define private algorithms for search problems. In Section 3 we provide impossibility results for the minimum vertex cover and the maximum exact 3SAT search problems. In Section 4 we discuss algorithms that leak (little) information and describe such algorithms for both search problems. Later, in Section 5, we prove our strongest impossibility result, showing vertex cover cannot be privately approximated even if a leakage of $O(\log n)$ bits is allowed. In Section 6, we present an impossibility result for a search problem in $\mathcal{P}$. In Appendix A, we discuss an equivalent definition for private algorithms for search problems. In Appendices B and C, we complete the proofs of the impossibility results.

2 Definition of Private Algorithms with respect to a Privacy Structure

There are two different aspects of private algorithms – the utility of the algorithm (what should be computed) and the privacy requirement (what should be protected, that is, what information should not be revealed by the computation). In computing functions this is quite straightforward, we want to compute (or approximate) a function and we want to protect inputs with the same output. For search algorithms, we know what should be computed. However, since these algorithms may output different outputs on the same input, it is less clear what should be protected. We, thus, separate the specification of what we want to protect from what we compute. In general, we require the output of the algorithm on certain pairs of inputs to be indistinguishable. In this section we define the pairs of inputs that should be protected by a private algorithm.

Definition 2.1 (Privacy Structure) *A privacy structure $\mathcal{R} \subseteq \{0,1\}^* \times \{0,1\}^*$ is an equivalence relation on instances. For $\langle x, y \rangle \in \mathcal{R}$, we use the notation $x \equiv_{\mathcal{R}} y$.*

We will only discuss privacy structures of the form $\mathcal{R} = \cup_{n \in \mathcal{N}} \mathcal{R}_n$, where $\mathcal{R}_n$ is an equivalence relation between instances of size n, such as graphs on n vertices or Boolean formulae over n variables. We say that an algorithm $\mathcal{A}$ is *private* with respect to a privacy structure $\mathcal{R}$ if the results of executing $\mathcal{A}$ on two $\mathcal{R}$-equivalent inputs are computationally indistinguishable.

Definition 2.2 (Private Algorithm) *Let $\mathcal{R}$ be a privacy structure. A probabilistic polynomial time algorithm $\mathcal{A}$ is* private *with respect to $\mathcal{R}$ if for every polynomial-time algorithm $\mathcal{D}$ and for every positive polynomial $p(\cdot)$, there exists some $n_0 \in \mathbb{N}$ such that for every $x, y \in \{0,1\}^*$ such that $x \equiv_{\mathcal{R}} y$ and $|x| = |y| \geq n_0$*

$$|\Pr[\mathcal{D}(\mathcal{A}(x), x, y) = 1] - \Pr[\mathcal{D}(\mathcal{A}(y), x, y) = 1]| \leq \frac{1}{p(|x|)},$$

where the probabilities are taken over the random choices of $\mathcal{A}$ and $\mathcal{D}$. That is, when $x \equiv_{\mathcal{R}} y$, an algorithm $\mathcal{D}$ given an output of $\mathcal{A}$ cannot distinguish if the input of $\mathcal{A}$ is x or y.

Example 2.3 Let $f : \{0,1\}^* \to \mathbb{N}$ be a function. Define $\mathcal{R}_f = \{\langle x, y \rangle : |x| = |y|, f(x) = f(y)\}$. The relation $\mathcal{R}_f$ is the relation implicitly considered when discussing private computation of functions.

In Appendix A, we present an equivalent definition of privacy which requires "semantic security," and prove that this definition is equivalent to Definition 2.2.

We next recall the definition of search problem and define the privacy structure associated with it.

Definition 2.4 *A bivariate relation Q is polynomially-bounded if there exists a constant c such that $|w| \leq |x|^c$ for every $\langle x, w \rangle \in Q$. The* decision problem *for Q is, given an input x, decide if there exists a w such that $\langle x, w \rangle \in Q$ or not. The* search problem *for Q is, given an input x,* find *a w such that $\langle x, w \rangle \in Q$ if such w exists.*

Search problems are the more common algorithmic task of finding a solution to a problem (rather than deciding whether the problem has a solution or not). To define private solution or private approximation of a search problem, one must determine the privacy structure the algorithm should respect. It is a *minimal* requirement to demand that if two inputs have the same set of answers to the search problem, the approximation algorithm should not enable to distinguish between them.

Definition 2.5 (Privacy Structure of a Search Problem) *The privacy structure $\mathcal{R}_Q$ related to a relation Q is defined as follows: $x \equiv_{\mathcal{R}_Q} y$ iff*

- *$|x| = |y|$, and*
- *$\langle x, w \rangle \in Q$ iff $\langle y, w \rangle \in Q$ for every w.*

That is, $x \equiv_{\mathcal{R}_Q} y$ if they have the same set of solutions.

We first give an example of a search problem in $\mathcal{P}$.

Example 2.6 Let maxMatch be the maximum matching relation, that is, $\langle G, M \rangle \in$ maxMatch if M is a maximum matching in G. In this case, $\mathcal{R}_{\text{maxMatch}}$ contains all pairs of graphs $G_1 = \langle V, E_1 \rangle, G_2 = \langle V, E_2 \rangle$ for which M is a maximum matching for G_1 iff it is a maximum matching for G_2.

We give two examples of privacy structures, for specific relations, that would be the focus of this paper.

Example 2.7 Let minVC be the minimum vertex cover relation, that is, $\langle G, C\rangle \in$ minVC if C is a minimum vertex cover in G. In this case, $\mathcal{R}_{\text{minVC}}$ contains all pairs of graphs $G_1 = \langle V, E_1\rangle, G_2 = \langle V, E_2\rangle$ for which $C \subseteq V$ is a minimum vertex cover for G_1 iff it is a minimum vertex cover for G_2.

Example 2.8 An exact 3CNF formula is a CNF formula that contains exactly three different literals in each clause. Let maxE3SAT be the maximum exact three SAT relation, that is, $\langle \phi, a\rangle \in$ maxE3SAT if ϕ is an exact 3CNF formula over n variables, and a is an assignment to the n variables that satisfies the maximum possible number of clauses in ϕ. In this case, the privacy structure $\mathcal{R}_{\text{maxE3SAT}}$ contains all pairs of exact 3CNF formulae ϕ_1, ϕ_2 over n variables for which an assignment a satisfies the maximum number of clauses in ϕ_1 iff it satisfies the maximum number of clauses in ϕ_2.

We note that a related definition of private approximation was recently presented in the context of the Nearest Neighbor problem [17]. Their privacy requirement is that instances with identical sets of *approximate* solutions should not be told apart by the private approximation algorithm. This definition may be cast in our framework by constructing a privacy structure where instances that have the same collection of approximate solutions are considered equivalent.

3 Impossibility Results for Private Approximation

3.1 Impossibility Results for Private Approximation of Vertex Cover

In this section we show that private approximation of the vertex cover search problem with respect to $\mathcal{R}_{\text{minVC}}$ (defined in Example 2.7) is a hard task. We start with defining private approximation of vertex cover, and then prove impossibility results for both the deterministic and the randomized settings.

Definition 3.1 (Private Approximation of Vertex Cover) *An algorithm $\mathcal{A}$ is a* private $c(n)$-approximation *algorithm for* minVC *if: (i) $\mathcal{A}$ runs in polynomial time. (ii) $\mathcal{A}$ is a $c(n)$-approximation algorithm for* minVC*, that is, for every graph G with n vertices, it returns a vertex cover whose size is at most $c(n)$ times the size of the smallest vertex cover of G. (iii) $\mathcal{A}$ is private with respect to $\mathcal{R}_{\text{minVC}}$.*

To illustrate our definitions, we present a private $(n/\log n)$-approximation algorithm for the vertex cover problem. This algorithm is based on the polynomial algorithm of [23] that returns a minimum vertex cover if the size of the vertex cover is at most $\log n$. Actually, in this case there are at most n^2 such covers,[1] and the algorithm can efficiently compute all of them. Thus, we can define any rule to choose one of them (e.g., the lexicographically first or uniform distribution). To approximate minVC we do the following:

If there is a cover of size at most $\log n$, return the lexicographically first minimum vertex cover. Otherwise, return the entire set of vertices.

We show impossibility results for privately approximating vertex cover in the deterministic and in the randomized setting.

Theorem 3.2 *Let $\epsilon > 0$ be a constant.*

[1] The algorithm of [23] first finds a maximal matching of size at most $\log n$ (the size of any matching is a lower bound on the size of a vertex cover). Let P be the end-points of the edges in this matching (thus, $|P| \leq 2\log n$). In [23] it is proven that any minimum cover is composed of a subset $P' \subseteq P$ and all the neighbors of $P \setminus P'$. As there are at most n^2 subsets of P, there are at most n^2 minimum vertex covers for the graph.

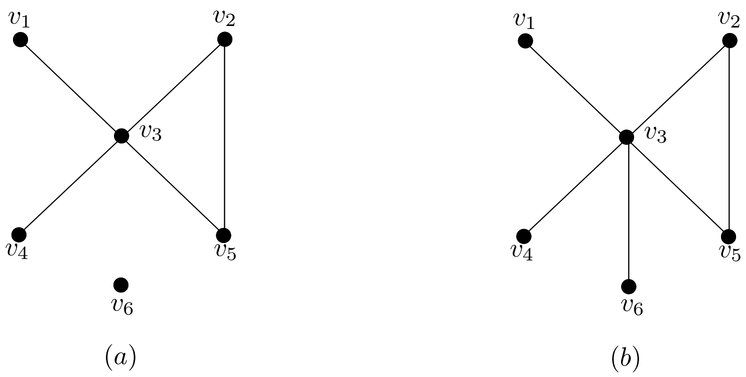

Figure 1: A pair of graphs equivalent under $\mathcal{R}_{\text{minVC}}$.

1. *If $\mathcal{P} \neq \mathcal{NP}$, then there is no* deterministic *private $n^{1-\epsilon}$-approximation algorithm for the search problem of* minVC.
2. *If $\mathcal{RP} \neq \mathcal{NP}$, then there is no* randomized *private $n^{1-\epsilon}$-approximation algorithm for the search problem of* minVC.

Part 1 of Theorem 3.2 is proven in the rest of this section. Part 2 of Theorem 3.2 is a special case of Theorem 5.1 proven in Appendix B.

3.1.1 Relevant and Critical Vertices

The framework for proving Theorem 3.2 is the following: We assume the existence of the appropriate private approximation algorithm and derive a greedy algorithm that solves vertex cover *exactly*. The following definitions are central for both the deterministic and the randomized case.

Definition 3.3 (Critical Vertices and Relevant Vertices) *Let $G = \langle V, E \rangle$ be a graph and $v \in V$ be a vertex of G. We say that v is* critical *for G if* every *minimal vertex cover of G contains v. We say that v is* relevant *for G if there* exists *a minimal vertex cover of G that contains v.*

Observation 3.4 *Every vertex is relevant or non-critical (or both).*

Example 3.5 To illustrate the relation $\mathcal{R}_{\text{minVC}}$ we show a pair of graphs that are equivalent under the relation. One way to create such a pair is to pick a graph and identify a vertex that is critical for this graph. For example, vertex v_3 in Figure 1(a) is a critical vertex. To get the second graph we connect the critical vertex to some other vertex. In Figure 1(b), vertex v_3 is connected to v_6. It is easy to verify that the set of minimum covers in both graphs is $\{\{v_3, v_2\}, \{v_3, v_5\}\}$. The equivalence can also be derived from Claim 3.8 below.

We reduce the design of a greedy algorithm for vertex cover, to solving the following problem.

Definition 3.6 (The Relevant / Non-Critical Problem)
INPUT: *A graph $G = \langle V, E \rangle$ and a vertex $v \in V$.*
OUTPUT: *One of the following: (i) v is relevant for G. (ii) v is non-critical for G.*

```
Algorithm Greedy Vertex Cover
```

INPUT: A graph $G = \langle V, E \rangle$.
OUTPUT: A minimal vertex cover of G.

1. If $V = \emptyset$, return $\emptyset$.
2. Pick a vertex $v \in V$ and execute Algorithm `Relevant-Non-Critical` on G and v.
3. If the answer is "RELEVANT":
   (a) Run `Greedy Vertex Cover` on the graph $G \setminus \{v\}$. Denote the answer by C_v.
   (b) Return $C_v \cup \{v\}$.
4. If the answer is "NON-CRITICAL":
   (a) Let $N(v)$ be the neighbors of v in G.
   (b) Run `Greedy Vertex Cover` on the graph $G \setminus (\{v\} \cup N(v))$. Denote the answer by $C_{N(v)}$.
   (c) Return $C_{N(v)} \cup N(v)$.

Figure 2: A greedy algorithm using Algorithm `Relevant-Non-Critical` to find a minimum vertex cover.

In Figure 2, we describe a greedy algorithm for vertex cover given an access to an algorithm that solves the Relevant / Non-Critical problem. In subsequent sections, we solve the latter using oracle access to private approximation algorithms for vertex cover. The following claim asserts the correctness of the greedy algorithm.

Claim 3.7 *If Algorithm* `Relevant-Non-Critical` *is polynomial and correct, then Algorithm* `Greedy Vertex Cover` *is polynomial and correct.*

Proof: The proof is by induction on $|V|$. The algorithm is trivially correct for $V = \emptyset$. Now suppose $V \neq \emptyset$, and we start from the case where v is relevant. In this case, there is a minimal cover C for G that contains v. Denote $d = |C|$, and note that the set $C \setminus \{v\}$ is a cover of size $d - 1$ for the graph $G \setminus \{v\}$. By the induction hypothesis, the set C_v is a minimal cover for $G \setminus \{v\}$. We claim that $C_v \cup \{v\}$ is a minimal cover for G. First note that it is indeed a cover of G as any edge adjacent to v is covered by v, and all the rest are covered by C_v. Since C_v is a minimal cover of $G \setminus \{v\}$, it is of size at most $d - 1$. Therefore, the size of $C_v \cup \{v\}$ is d, and it is a minimal cover of G.

In the other case, v is not critical for G. Thus, there is a minimal cover C' for G that does not contain v, and, therefore, contains all vertices in $N(v)$. Denote $d' = |C'|$ and $h = |N(v)|$. In this case the set $C' \setminus N(v)$ is a cover of size $d' - h$ of the graph $G \setminus (N(v) \cup \{v\})$. By the induction hypothesis, the set $C_{N(v)}$ is a minimal cover for $G \setminus (N(v) \cup \{v\})$. We claim that $C_{N(v)} \cup N(v)$ is a minimal cover for G. First note that it is indeed a cover of G as any edge adjacent to $N(v) \cup \{v\}$ is covered by $N(v)$, and all other edges are covered by $C_{N(v)}$. Since $C_{N(v)}$ is a minimal cover of $G \setminus (N(v) \cup \{v\})$, it is of size at most $d' - h$. Therefore, the size of $C_{N(v)} \cup N(v)$ is d', and it is a minimal cover of G.

It is straightforward to verify that if `Relevant-Non-Critical` is polynomial, then the greedy algorithm is polynomial. $\square$

3.1.2 Combinatorial Claims

The following combinatorial claims are helpful in designing algorithms for the Relevant / Non-Critical problem in both the deterministic and the randomized settings. Intuitively, a private approximation algorithm must be "sensitive" to small changes in the set of minimum vertex covers of its input graph. We study the connection between the $\mathcal{R}_{\text{minVC}}$ relation and the set of critical and relevant vertices in a graph.

Claim 3.8 *Let $G = \langle V, E\rangle$ be a graph, $u, v \in V$ such that $(u, v) \notin E$, and $G^* = \langle V, E^*\rangle$, where $E^* = E \cup (u, v)$. If u is critical for G, then $G \equiv_{\mathcal{R}_{\text{minVC}}} G^*$.*

Proof: We first show that every minimum vertex cover of G is a minimum vertex cover of G^*. Let C be a minimal cover of G. As u is critical for G, we get that $u \in C$. Therefore, C covers the edge (u, v) and thus it is a cover of G^*. Note that every cover of G^* is also a cover of G, and thus C is a minimal cover of G^*.

For the other direction, let C^* be a minimal cover of G^*. Let c be the size of a minimal cover of G. As u appears in at least one minimal cover of G, which is also a cover of G^*, the size of C^* is at most c. On the other hand, as $E \subseteq E^*$, the set C^* is also a cover of G and thus the size of C^* is exactly c. Therefore, C^* is a minimal cover of G. $\square$

We will later see that if a vertex u is not in the result of the private approximation algorithm $\mathcal{A}$, then it is non-critical for the input graph G. However, if the vertex cover size of the input graph is large, the approximation algorithm may return the entire set V as its result. To avoid this, we add a large set of isolated vertices to G. (The size of this set is a function of the approximation ratio.) The mere fact that an isolated vertex is non-critical for G is of-course not helpful. Nevertheless, we gain information by connecting this isolated vertex to the vertex v and running $\mathcal{A}$ on the new graph. It will also be helpful to consider duplicating the graph G and connecting the isolated vertex to both copies of the original vertex v.

Definition 3.9 (The Graphs G_2 and $G({}^{i}_{\wedge})$) *Let $G = \langle V, E\rangle$ be a graph, $v \in V$ be a vertex, I be a set of vertices, and $i \in I$. The graph G_2 is defined as $G_2 = \langle V_2, E_2\rangle$ where $V_2 \stackrel{\text{def}}{=} (V \times \{1, 2\}) \cup I$ and $E_2 \stackrel{\text{def}}{=} \{(\langle u, j\rangle, \langle w, j\rangle) : (u, w) \in E, j \in \{1, 2\}\}$. The graph $G({}^{i}_{\wedge})$ is defined as $G({}^{i}_{\wedge}) = \langle V_2, E({}^{i}_{\wedge})\rangle$ where $E({}^{i}_{\wedge}) \stackrel{\text{def}}{=} E_2 \cup \{(\langle v, j\rangle, i) : j \in \{1, 2\}\}$.*

The graphs G_2 and $G({}^{i}_{\wedge})$ are illustrated in Figure 3. The following claims summarize the properties of G_2 and $G({}^{i}_{\wedge})$.

Claim 3.10 *If v is critical for G, then $G_2 \equiv_{\mathcal{R}_{\text{minVC}}} G({}^{i}_{\wedge})$.*

Proof: Since G_2 contains two separate copies of G and v is critical for G, the vertices $\langle v, 1\rangle$ and $\langle v, 2\rangle$ are critical for G_2. Hence, by Claim 3.8, adding the edges $(\langle v, 1\rangle, i)$ and $(\langle v, 2\rangle, i)$ does not change the set of minimal vertex covers of the graph. $\square$

Claim 3.11 *If v is not relevant for G, then i is critical for $G({}^{i}_{\wedge})$.*

Proof: Assume towards contradiction that the vertex i is non-critical for $G({}^{i}_{\wedge})$. Hence, there must be a minimal cover $C({}^{i}_{\wedge})$ of $G({}^{i}_{\wedge})$ that contains $\langle v, 1\rangle$ and $\langle v, 2\rangle$. The intersection of $C({}^{i}_{\wedge})$ with each copy of G contains a cover of G that contains the appropriate copy of v. As v is not relevant for G, these covers are not optimal. Let c be the size of a minimum vertex cover of G. Then $\left|C({}^{i}_{\wedge})\right| \geq 2(c + 1) = 2c + 2$. On the other hand, let C be a minimum cover of G of size c. Then, the set $(C \times \{1, 2\}) \cup \{i\}$ is a cover of $G({}^{i}_{\wedge})$ of size $2c + 1$, in contradiction to the minimality of $C({}^{i}_{\wedge})$. Hence, i is critical for $G({}^{i}_{\wedge})$. $\square$

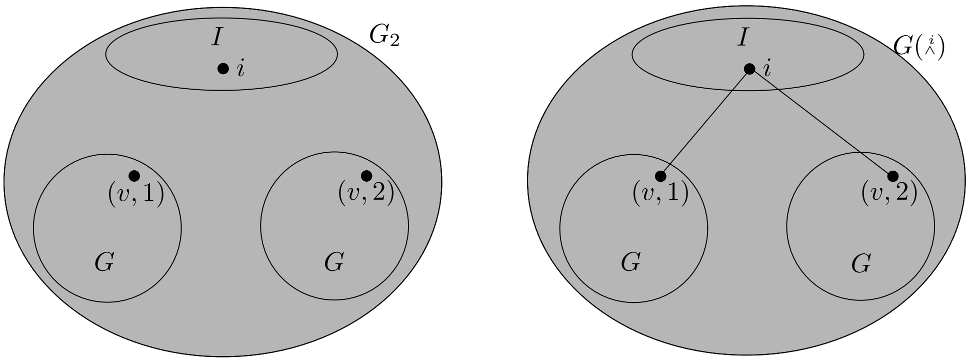

Figure 3: The graphs G_2 and $G\binom{i}{\wedge}$.

3.1.3 Impossibility Result for Deterministic Private Approximation

In this section we show an algorithm that solves the Relevant / Non-Critical problem, given an oracle access to a deterministic private approximation algorithm for minVC.

Claim 3.12 *Let $\mathcal{A}$ be a deterministic private approximation algorithm for* minVC*, let $G = \langle V, E\rangle$ be a graph, and denote $W = \mathcal{A}(G)$. Then for any two different vertices $v_1, v_2 \in V \backslash W$, the vertex v_1 is not critical for G (and, similarly, v_2 is not critical).*

Proof: As v_1 and v_2 are not in W, and W is a cover of G, we infer that $(v_1, v_2) \notin E$. Let $E^* = E \cup \{(v_1, v_2)\}$, and define $G^* = \langle V, E^*\rangle$. We now consider a hypothetical execution of the algorithm $\mathcal{A}$ on G^* and denote $W^* = \mathcal{A}(G^*)$. The set W^* must cover the edge (v_1, v_2) and thus $W^* \neq W$.

If v_1 is critical for G, then, by Claim 3.8, the sets of minimum-size vertex cover of G and G^* are equal. However, since $\mathcal{A}$ is deterministic and private, and $W \neq W^*$, the set of minimal vertex covers of G and G^* must be different. Therefore, v_1 is not critical for G. □

In Figure 4, we present the algorithm `Relevant-Non-Critical`. It takes a graph $G = \langle V, E\rangle$ and a vertex $v \in V$ as inputs, and uses a private $n^{1-\epsilon}$-approximation algorithm $\mathcal{A}$ to solve the Relevant / Non-Critical problem.

The correctness of the algorithm `Relevant-Non-Critical` steams from the following claims:

Claim 3.13 *If $W_2 \neq W\binom{v'}{\wedge}$, then v is not critical for G.*

Proof: Assume towards contradiction that v is critical for G. By Claim 3.10, the graphs G_2 and $G\binom{v'}{\wedge}$ have the same set of minimal vertex covers. Hence, from the privacy of $\mathcal{A}$, we get that $W_2 = \mathcal{A}(G_2) = \mathcal{A}(G\binom{v'}{\wedge}) = W\binom{v'}{\wedge}$, contradicting $W_2 \neq W\binom{v'}{\wedge}$. □

Claim 3.14 *If $W_2 = W\binom{v'}{\wedge}$, then v is relevant for G.*

Proof: As $W_2 = W\binom{v'}{\wedge}$, and $v' \notin W_2$, we get $v' \notin W\binom{v'}{\wedge}$. Hence, by Claim 3.12, the vertex v' is not critical for $G\binom{v'}{\wedge}$. Applying Claim 3.11, we infer that v is relevant for G. □

To conclude the proof of Part 1 of Theorem 3.2, we show that there is a vertex we can choose in step (4) of the algorithm.

Algorithm `Relevant-Non-Critical`

INPUT AND OUTPUT: See Definition 3.6.

1. Let I be a set of vertices of size $(4n)^{1/\epsilon} - 2n$.
2. Construct the graph G_2 from G and I as in Definition 3.9.
3. Execute $\mathcal{A}$ on G_2 and denote $W_2 = \mathcal{A}(G_2)$.
4. Choose any vertex $v' \in I \backslash W_2$ (there must be at least two such vertices as the approximation algorithm cannot return all the set I).
5. Construct $G(\genfrac{}{}{0pt}{}{v'}{\wedge})$ from G, I, and v' as in Definition 3.9.
6. Execute $\mathcal{A}$ on $G(\genfrac{}{}{0pt}{}{v'}{\wedge})$ and denote $W(\genfrac{}{}{0pt}{}{v'}{\wedge}) = \mathcal{A}(G(\genfrac{}{}{0pt}{}{v'}{\wedge}))$.
7. If $W_2 \neq W(\genfrac{}{}{0pt}{}{v'}{\wedge})$ return "NOT CRITICAL." Else return "RELEVANT."

Figure 4: Algorithm `Relevant-Non-Critical` for a private deterministic $\mathcal{A}$.

Claim 3.15 *Let $\epsilon > 0$. There is a vertex $v' \in I$ such that $v' \in I \setminus W_2$.*

Proof: Let $N = |I| + 2n = (4n)^{1/\epsilon}$ be the number of vertices in G_2. The size of the minimum vertex cover of G_2 is twice the size of the minimum vertex cover of G, thus, it is at most $2n$. Since $\mathcal{A}$ is an $N^{1-\epsilon}$-approximation algorithm for vertex cover, the size of $\mathcal{A}(G_2)$ is at most

$$2n \cdot N^{1-\epsilon} \; = \; 2n \cdot ((4n)^{1/\epsilon})^{1-\epsilon} \; = \; \frac{(4n)^{1/\epsilon}}{2} \; < \; (4n)^{1/\epsilon} - 2n \; = \; |I|.$$

Consequently, there is at least one vertex $v' \in I \setminus W$. □

We extend the techniques described in this section to get an impossibility result with respect to a weaker notion of private approximation defined in Section 4 (see Theorem 5.1).

3.2 Impossibility Results for Private Approximation of Exact 3SAT

Similar results are obtained for private approximation of maxE3SAT as stated in the following theorem:

Theorem 3.16 *Let $\epsilon > 0$ be a constant.*

1. *If $\mathcal{P} \neq \mathcal{NP}$, then there is no* deterministic *private $1/n^{1-\epsilon}$-approximation algorithm for the search problem of* maxE3SAT.
2. *If $\mathcal{RP} \neq \mathcal{NP}$, then there is no* randomized *private $1/n^{1-\epsilon}$-approximation algorithm for the search problem of* maxE3SAT.

See Appendix C for a proof of the deterministic case. The proof of the probabilistic case follows, with similar modifications as done in Appendix B for minVC.

4 Algorithms that Leak Little Information

As demonstrated in the previous section, in some cases it is impossible to design an efficient algorithm which is private with respect to a privacy structure $\mathcal{R}$. However, letting $\mathcal{R}'$ be a refinement of $\mathcal{R}$, we view a private algorithm with respect to $\mathcal{R}'$ as a private algorithm with respect to $\mathcal{R}$ that *leaks* information. The amount of information leaked is quantified according to the relation between $\mathcal{R}$ and $\mathcal{R}'$.

Definition 4.1 (k-Refinement) *Let $\mathcal{R}$ and $\mathcal{R}'$ be two privacy structures over $\{0,1\}^*$ and $k : \mathbb{N} \to \mathbb{N}$. We say that $\mathcal{R}'$ is a $k(n)$-refinement of $\mathcal{R}$ if $\mathcal{R}' \subseteq \mathcal{R}$ and for every $n \in \mathbb{N}$ every equivalence class of $\mathcal{R}$ of strings of size n is a union of at most $2^{k(n)}$ equivalence classes of $\mathcal{R}'$.*

Definition 4.2 (Algorithms Leaking $k(n)$-bits) *Let $\mathcal{R}$ be a privacy structure. A probabilistic polynomial time algorithm $\mathcal{A}$* leaks at most $k(n)$ bits *with respect to $\mathcal{R}$ if there exists a privacy structure $\mathcal{R}'$ such that (i) $\mathcal{R}'$ is a $k(n)$-refinement of $\mathcal{R}$, and (ii) $\mathcal{A}$ is private with respect to $\mathcal{R}'$.*

4.1 Solution-List Algorithms

Solution-list algorithms are algorithms whose outcome is always in a small predetermined set. With respect to privacy, solution-list algorithms are valuable as they leak only a few bits with respect to any privacy structure – at most logarithmic in the number of their possible outcomes.

Definition 4.3 (Solution-List Algorithm) *We say that a deterministic algorithm $\mathcal{A}$ is a $K(n)$-*solution-list *algorithm if for every $n \in \mathbb{N}$*

$$|\{\, y \,:\, \exists x \in \{0,1\}^n \text{ such that } \mathcal{A}(x) = y \,\}| \leq K(n).$$

I.e., a solution-list algorithm is an algorithm such that, for every input size n, "chooses" its outputs from a set of at most $K(n)$ possible outcomes.

We define the universal relation, denoted $\mathcal{U}^* = \cup_{n \in \mathcal{N}} \mathcal{U}_n^*$, as the privacy structure where every two instances of the same size are equivalent. Note that any privacy structure is a refinement of $\mathcal{U}^*$, hence if an algorithm is private with respect to $\mathcal{U}^*$ it is also private with respect to any privacy structure,[2] and similarly, if an algorithm leaks at most k bits with respect to $\mathcal{U}^*$, then so is the case with respect to any privacy structure.

Observation 4.4 *Any $K(n)$-solution-list algorithm leaks at most $\log K(n)$ bits with respect to $\mathcal{U}^*$.*

4.2 Solution-List Algorithms for Exact 3SAT

In this section we present a $(7/8 - \epsilon)$-approximation algorithm for maximum satisfiability on exact 3CNF formulae that leaks little information, i.e., $O(\log \log n)$ bits. The algorithm is a solution-list algorithm as defined in Definition 4.3. The approximation in our algorithm is nearly optimal, as by the result of Håstad [16], if $\mathcal{P} \neq \mathcal{NP}$, then there is no polynomial-time $(7/8 + \epsilon)$-approximation algorithm for this problem.[3]

We start with a simple motivating example. Consider the following simple algorithm for the max-SAT problem:

[2] An algorithm $\mathcal{A}$ is private with respect to $\mathcal{U}^*$ if only the instance size may be learned from its outcome.

[3] Any $(7/8 + \epsilon)$-approximation solution-list algorithm that uses $\mathrm{poly}(n)$ solutions would imply $\mathcal{NP} \subseteq \mathcal{P}/\mathrm{poly}$ even if the list cannot be efficiently constructed.

If 0^n satisfies at least half of the clauses in ϕ, then return 0^n. Otherwise, return 1^n.

For every clause, either 0^n or 1^n satisfy the clause. Thus, 0^n or 1^n satisfy at least half of the clauses in ϕ, and this is a $1/2$-approximation of max-SAT. Since there are only two possible answers, this algorithm leaks at most one bit.

Claim 4.5 *There is a* $7/8$*-approximation algorithm for* maxE3SAT *that leaks at most* $O(\log n)$ *bits with respect to* $\mathcal{R}_{\text{maxE3SAT}}$. *Furthermore, for every* $\epsilon > 0$*, there is a* $(7/8 - \epsilon)$*-approximation algorithm for* maxE3SAT *that leaks at most* $O(\log\log n)$ *bits with respect to* $\mathcal{R}_{\text{maxE3SAT}}$.

Proof: We first describe the $7/8$-approximation algorithm. Towards this goal, we construct for every n a list of $\text{poly}(n)$ assignments such that for every exact 3CNF formula with n variables there is an assignment in the list that satisfies at least $7/8$ of the clauses of the formula. Furthermore, there is an efficient algorithm that generates this list. Thus, the $7/8$-approximation algorithm, with input ϕ – a formula with n variables – constructs this list, and chooses the first assignment in the list that satisfies at least $7/8$ of the clauses in ϕ.

We next explain how to construct the list, using ideas of the randomized $7/8$-approximation algorithm of Johnson [18]. Fix a clause with three different literals. If we pick an assignment at random, then with probability at least $7/8$ it satisfies the clause. Now, fix any exact 3CNF formula. If we pick an assignment at random, then the expected fraction of satisfied clauses is at least $7/8$. Thus, there exists at least one assignment that satisfies a fraction of at least $7/8$ of the clauses in the formula. This is true even if we pick the assignments from a 3-wise independent space. As there is a 3-wise independent space of size $O(n^3)$, this implies the existence of the list. To generate the assignments we can use any of the constructions of 3-wise independent spaces, e.g., the construction based on polynomials (see, e.g., [20, 1, 6]).

We next describe the $(7/8 - \epsilon)$-approximation algorithm. As in the previous case, it suffices to show how to efficiently construct, for every n and $\epsilon > 0$, a list of $\text{poly}(\frac{\log n}{\epsilon})$ assignments such that for every exact 3CNF formula with n variables there is an assignment in the list that satisfies at least $7/8 - \epsilon$ of the clauses of the formula. To construct the list, notice that if we pick an assignment from an $(\epsilon, 3)$-wise independent space, then the probability that a given clause is satisfied is at least $7/8 - \epsilon$. Thus, the expected fraction of satisfied clauses is at least $7/8 - \epsilon$, and there exists at least one assignment that satisfies a fraction of at least $7/8 - \epsilon$ of the clauses in the formula. There are $(\epsilon, 3)$-wise independent spaces of size $\text{poly}(\frac{\log n}{\epsilon})$. To generate the assignments we can use any of the constructions of [22, 2]. □

Claim 4.6 *Every solution-list algorithm for* maxE3SAT *that achieves approximation ratio better than* $1/2$ *uses at least* $\log n - 1$ *solutions.*

Proof: Assume a solution-list algorithm for maxE3SAT, and let $a_1, \ldots, a_t$, where $t < \log n - 1$, be its list of possible output assignments on formulae over n variables $x_1, \ldots, x_n$. To each variable x_i we assign a *label* that is the concatenation of the truth values assigned to x_i by the t assignments $\langle a_1(x_i), \ldots, a_t(x_i)\rangle$. As there are at most 2^t different labels and $n/2^t > 2$, there exist three distinct variables $x_{i_1}, x_{i_2}, x_{i_3}$ that share the same label. I.e., for all $1 \leq j \leq t$ it holds that $a_j(x_{i_1}) = a_j(x_{i_2}) = a_j(x_{i_3})$.

Consider the formula $\phi = (x_{i_1} \vee x_{i_2} \vee x_{i_3}) \wedge (\neg x_{i_1} \vee \neg x_{i_2} \vee \neg x_{i_3})$. It is easy to see that ϕ is satisfied by exactly those assignment which do not assign the same truth value to all three variables $x_{i_1}, x_{i_2}, x_{i_3}$. However, as this is not the case for any of the t assignments, each of them satisfies exactly one clause in ϕ, achieving approximation factor at most $1/2$. □

4.3 Solution-List Algorithms for Vertex Cover

In this section we present almost private search algorithms for minimum vertex cover. These algorithms are significant, but not as dramatic as the algorithms for maxE3SAT. For any $0 < \epsilon < 1$, there is an $n^{1-\epsilon}$-approximation algorithm that leaks $O(n^\epsilon)$ bits. The algorithms are solution-lists algorithms, and we will prove that no solution-list algorithm can do better for this problem.

Claim 4.7 *For every $0 < \epsilon < 1$, there is an $n^{1-\epsilon}$-approximation algorithm for the minimum vertex cover problem which leaks at most $2n^\epsilon$ bits.*

Proof: The algorithm proceeds as follows:

> INPUT: A graph G with n vertices.
>
> 1. Execute any 2-approximation algorithm for VC on G, and get a cover C.
> 2. Let $\ell \leftarrow 2n^\epsilon$.
> 3. Partition the n vertices into ℓ fixed sets, $V_1, \ldots, V_\ell$, each of size $n/\ell = n^{1-\epsilon}/2$.
> 4. Let $C' \leftarrow \bigcup_{\{i:V_i \cap C \neq \emptyset\}} V_i$.
> 5. Return C'.

The algorithm first finds a small cover. Then, if V_i contains at least one vertex in this cover, the algorithm returns the entire set V_i. This implies that the size of C' is at most $|C|n^{1-\epsilon}/2$, and since $|C|$ is at most twice the size of the minimum vertex cover, this algorithm is an $n^{1-\epsilon}$-approximation algorithm.

Notice that the algorithm has 2^ℓ possible outputs (it only chooses which of the sets V_i is in its output). That is, this is a solution-list algorithm with a list of size 2^ℓ, thus, it leaks at most $\ell = 2n^\epsilon$ bits. □

The private algorithms for maxE3SAT and for minVC that we presented are solution-list algorithms. For maxE3SAT, the algorithm generated the entire list, and chooses the best candidate in the list. For minVC this is not possible as the size of the list is big. The algorithm generates a "good" candidate from the list without generating the entire list.

We next claim that any solution-list algorithm for minVC cannot use a shorter list than the algorithm we presented (up-to a constant factor).

Claim 4.8 *Any solution-list algorithm that $n^{1-\epsilon}$-approximates* minVC*, uses at least $2^{n^\epsilon/6}$ solutions.*

Proof: Assume that there is a list of covers that $n^{1-\epsilon}$-approximates minVC, that is, for every graph with n vertices and minimum vertex cover of size d, there exists a cover of size at most $n^{1-\epsilon}d$ in the list. We construct a "big" family of graphs, and show that every cover in the list covers "few" graphs in the family, thus the size of the list must be "big."

Consider the following family of graphs. Each graph is defined by a subset I of size $d \stackrel{\text{def}}{=} n^\epsilon/6$. The graph G_I contains all edges between I and $V \setminus I$. Notice that the number of graphs in this family is

$$\binom{n}{d} \geq (n/d)^d. \tag{1}$$

The set I of size d is a cover of the graph G_I. Since we assume that the list $n^{1-\epsilon}$-approximates minVC, there is a cover in the list that covers G_I and contains at most $n^{1-\epsilon}d = n^{1-\epsilon}n^\epsilon/6 = n/6$ vertices. However, every vertex cover of G_I contains either all vertices in I (and possibly vertices from $V \setminus I$) or all vertices of $V \setminus I$ (and possibly vertices from I). Since $|V \setminus I| = n - d > n/6$, this cover must contain I. The number of graphs in the family that a given cover of size at most $n/6$ covers is at most

$$\binom{n/6}{d} \leq \left(\frac{en/6}{d}\right)^d. \tag{2}$$

Thus, by (1) and (2), the number of covers in the list is at least

$$\frac{(n/d)^d}{(en/6d)^d} \geq 2^d = 2^{n^\epsilon/6}.$$

□

We emphasize that Claim 4.8 applies only to the size of lists used by solution lists algorithms, and does not imply that there is no polynomial $n^{1-\epsilon}$-approximation algorithms that leaks $o(n^\epsilon)$ bits.[4]

5 Impossibility Result for Vertex Cover Approximation that Leaks $\log n$ Bits

In this section we show it is unlikely that there is an efficient approximation algorithm for minVC that leaks $\log n$ bits of information. Specifically, if there is such an $n^{1-\epsilon}$-approximation algorithm $\mathcal{A}$ that leaks at most $\frac{\epsilon}{6}\log n$ bits, then $\mathcal{RP} = \mathcal{NP}$.

Theorem 5.1 *Let $\epsilon > 0$ be a constant. If $\mathcal{RP} \neq \mathcal{NP}$, then there is no randomized $n^{1-\epsilon}$-approximation algorithm for the search problem of* minVC *that leaks at most $\frac{\epsilon}{6}\log n$ bits.*

As in the proof of Theorem 3.2, we assume the existence of such an approximation algorithm and deduce an algorithm that solves vertex cover. Again, we do this by designing an algorithm that solves the Relevant / Non-Critical problem (see Definition 3.6), and thus, by Claim 3.7, solves vertex cover. This algorithm, given the input G and v, and an oracle access to an approximation algorithm $\mathcal{A}$ that leaks k bits, applies $\mathcal{A}$ on a set of inputs, and decides whether v is relevant or non-critical for G according to the results. Note that Algorithm `Relevant-Non-Critical` for the (perfectly) private case is not applicable; here, even assuming $\mathcal{A}$ is deterministic, the fact that $\mathcal{A}(G_1) \neq \mathcal{A}(G_2)$ does not directly imply that G_1 and G_2 are not equal under $\mathcal{R}_{\text{minVC}}$. However, as $\mathcal{A}$ leaks at most k bits, if there are $2^k + 1$ graphs with different outputs, then at least two of them are not equivalent.

We first deal in Section 5.1 with a deterministic $\mathcal{A}$ that leaks at most one bit, then we deal in Section 5.2 with a deterministic $\mathcal{A}$ that leaks at most $O(\log n)$ bits. The proof of the impossibility result for a randomized $\mathcal{A}$ that leaks at most $O(\log n)$ bits is deferred to Appendix B. We believe that presenting the proof through these stages is helpful for the reader.

Our inputs for the `Relevant/Non Critical` algorithm are generally constructed from a number of copies of the original input graph G and big set of isolated vertices I. In each such graph we connect the different copies of the input vertex v with vertices from I, and sometimes connect two different vertices from I. We will use the following notation to address these graphs.

[4]It can be proved that the algorithm we presented leaks $\Omega(n^\epsilon)$ bits.

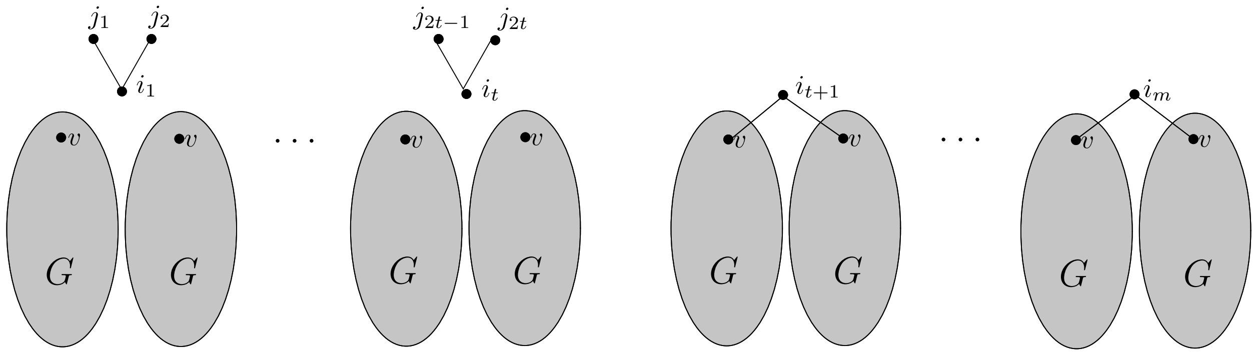

Figure 5: The graph $G(\substack{j_1 j_2 \\ \vee \\ i_1} \cdots \substack{j_{2t-1} j_{2t} \\ \vee \\ i_t} \; \substack{i_{t+1} \\ \wedge} \cdots \substack{i_m \\ \wedge})$.

Definition 5.2 *Let $G = \langle V, E\rangle$ be a graph, I be a set of vertices, t, m be indices such that $0 \leq t \leq m$, and $i_1, \ldots, i_m, j_1, \ldots, j_{2t} \in I$ be distinct vertices. The graph*

$$G(\substack{j_1 j_2 \\ \vee \\ i_1} \cdots \substack{j_{2t-1} j_{2t} \\ \vee \\ i_t} \; \substack{i_{t+1} \\ \wedge} \cdots \substack{i_m \\ \wedge})$$

is defined as follows: its vertex set is $(V \times \{1, \ldots, 2m\}) \cup I$, and its edges are $E = E_m \cup E_\vee \cup E_\wedge$, where

$$\begin{aligned}
E_m &\stackrel{\text{def}}{=} \{(\langle u, \ell\rangle, \langle w, \ell\rangle) : (u, w) \in E, \ell \in \{1, \ldots, m\}\}, \\
E_\vee &\stackrel{\text{def}}{=} \{(i_\ell, j_{2\ell-1}), (i_\ell, j_{2\ell}) : \ell \in \{1, \ldots, t\}\}, \textit{ and} \\
E_\wedge &\stackrel{\text{def}}{=} \{(\langle v, 2\ell-1\rangle, i_\ell), (\langle v, 2\ell\rangle, i_\ell) : \ell \in \{t+1, \ldots, m\}\}.
\end{aligned}$$

Informally, the graph has $2m$ copies of G (see Figure 5). It has t "vees," where the ℓth vee is associated with copies $2\ell - 1$ and 2ℓ of G. It has $m - t$ "wedges," where the ℓth wedge connects i_ℓ to the copies of v in copies $2\ell - 1$ and 2ℓ of G, for $t < \ell \leq m$.

We next present two combinatorial lemmas, whose role is similar to the Claims 3.10 and 3.11 in the case where $\mathcal{A}$ was perfectly private.

Claim 5.3 *Let $1 \leq t \leq m$, and $i_1, \ldots, i_m, j_1, \ldots, j_{2t}, i'_{t+1}, \ldots, i'_m \in I$ be distinct vertices. Furthermore, let*

$$H \stackrel{\text{def}}{=} G(\substack{j_1 j_2 \\ \vee \\ i_1} \cdots \substack{j_{2t-1} j_{2t} \\ \vee \\ i_t} \; \substack{i_{t+1} \\ \wedge} \cdots \substack{i_m \\ \wedge}) \quad \text{and} \quad H' \stackrel{\text{def}}{=} G(\substack{j_1 j_2 \\ \vee \\ i_1} \cdots \substack{j_{2t-1} j_{2t} \\ \vee \\ i_t} \; \substack{i'_{t+1} \\ \wedge} \cdots \substack{i'_m \\ \wedge}).$$

If v is critical for G, then $H \equiv_{\mathcal{R}_{\text{minVC}}} H'$.

Proof: By Claim 3.10, the graphs H and H' are unions of graphs the are equivalent, thus, H and H' are equivalent. □

Claim 5.4 *Let $0 \leq t' < t \leq m$, and $i_1, \ldots, i_m, j_1, \ldots, j_{2t} \in I$ be distinct vertices. Furthermore, let*

$$H \stackrel{\text{def}}{=} G(\substack{j_1 j_2 \\ \vee \\ i_1} \cdots \substack{j_{2t'-1} j_{2t'} \\ \vee \\ i_{t'}} \cdots \substack{j_{2t-1} j_{2t} \\ \vee \\ i_t} \; \substack{i_{t+1} \\ \wedge} \cdots \substack{i_m \\ \wedge}) \quad \text{and} \quad H' \stackrel{\text{def}}{=} G(\substack{j_1 j_2 \\ \vee \\ i_1} \cdots \substack{j_{2t'-1} j_{2t'} \\ \vee \\ i_{t'}} \; \substack{i_{t'+1} \\ \wedge} \cdots \substack{i_m \\ \wedge}).$$

If v is non-relevant for G, then $H \equiv_{\mathcal{R}_{\text{minVC}}} H'$.

Proof: Note that the vertices $i_1, \dots, i_m$ are critical for both H and H': It is straightforward that i_ℓ is critical for $\overset{j_{2\ell-1}j_{2\ell}}{\underset{i_\ell}{\vee}}$. By Claim 3.11, vertex i_ℓ is critical for $\overset{i_\ell}{\wedge}$. Therefore, the two graphs have the same set of minimum vertex covers, hence they are equivalent under $\mathcal{R}_{\text{minVC}}$. □

In both claims we have the same graph H. In the Claim 5.3, the graph H' has the same "vees" as H, however, the "wedges" have different vertices from I (namely, $i'_1, \dots, i'_{t-1}$). In Claim 5.4, the sequence of i's is the same in H and H', however, there are "vees" in H' in some places where there are "wedges" in H.

5.1 Handling One Bit Leakage

To simplify our presentation, we first present `Alg. RelNonCrit − one bit` in Figure 5.5 . This algorithm assumes that the deterministic algorithm $\mathcal{A}$ leaks at most one bit. This case is simpler and describes some of the ideas used for the $\log n$ bits leakage case.

`Alg. RelNonCrit − one bit`

INPUT: A graph $G = \langle V, E\rangle$ and a vertex $v \in V$.
OUTPUT: One of the following: (i) The vertex v is relevant for G. (ii) The vertex v is not critical for G.

1. Execute $\mathcal{A}$ on all the graphs of the form $G(\overset{i_1}{\wedge}\ \overset{i_2}{\wedge})$, where $i_1, i_2 \in I$.
   (a) If these executions result in more than two different answers, return "Non Critical,"
   (b) Else, remove from I all the vertices that appeared in any of the above results.
2. Pick arbitrary $i_1, j_1, j_2 \in I$ and execute $\mathcal{A}$ on all the graphs of the form $G(\overset{j_1 j_2}{\underset{i_1}{\vee}}\ \overset{i_2}{\wedge})$, where $i_2 \in I \setminus \{i_1, j_1, j_2\}$.
   (a) If these executions result in more than two different answers, return "Non Critical."
   (b) Else, return "Relevant."

Figure 6: `Alg. RelNonCrit − one bit`.

Claim 5.5 `Alg. RelNonCrit − one bit` *is correct.*

Proof: By Claim 5.3, if v is critical for G, then all graphs considered in step (1) of the algorithm are equivalent. As $\mathcal{A}$ leaks at most 1 bit, there can be at most 2 different answers on equivalent graphs. Hence, if there are more than two different answers in step (1), vertex v is non critical for G. Similarly, if there are more than two different answers in step (2), vertex v is non critical for G. Thus, if the algorithm outputs "Non critical," vertex v is non critical for G, and the algorithm is correct.

Else, fix any $i_2, j_3, j_4 \in I$ that did not appear in any of the results of $\mathcal{A}$ in `Alg. RelNonCrit − one bit`, and note that:

1. $\mathcal{A}(G(\overset{i_1}{\wedge}\ \overset{i_2}{\wedge}))$ *does not contain* any of $\{i_1, j_1, j_2\}$, and *does not contain* any of $\{i_2, j_3, j_4\}$.
2. $\mathcal{A}(G(\overset{j_1 j_2}{\underset{i_1}{\vee}}\ \overset{i_2}{\wedge}))$ *contains* at least one of $\{i_1, j_1, j_2\}$, and *does not contain* any of $\{i_2, j_3, j_4\}$.
3. $\mathcal{A}(G(\overset{j_1 j_2}{\underset{i_1}{\vee}}\ \overset{j_3 j_4}{\underset{i_2}{\vee}}))$ *contains* at least one of $\{i_2, j_3, j_4\}$.

Thus, these 3 results of $\mathcal{A}$ are all different. However, by Claim 5.4, if vertex v is not relevant for G, then the three graphs are equivalent. Since $\mathcal{A}$ leaks at most one bit, there cannot be three different answers on three equivalent graphs. Thus, if the algorithm outputs "Relevant," vertex v is relevant for G, and the algorithm is correct. $\square$

Similarly to the case where $\mathcal{A}$ is perfectly private, it is enough to set $|I| = O((4n)^{1/\epsilon} - 2n)$ to ensure there are enough vertices in I that are not returned by $\mathcal{A}$ (see Claim 3.15). As I is polynomial in n, the number of calls to $\mathcal{A}$ in the algorithm is polynomial. Thus, we have proved that if $\mathcal{P} \neq \mathcal{NP}$, then there is no deterministic $n^{1-\epsilon}$-approximation algorithm for vertex cover that leaks at most 1 bit.

5.2 Handling $\log n$-Bits Leakage

We next want to accommodate a deterministic algorithm $\mathcal{A}$ that leaks $\log n$ bits. Using a similar algorithm as in the one bit leakage case with graphs that contain more copies of G, we can handle Algorithms $\mathcal{A}$ that leaks more bits. However, if $\mathcal{A}$ leaks $\omega(1)$ bits, the number of executions of $\mathcal{A}$ will be too big. To reduce the number of calls to $\mathcal{A}$, we choose the vertices from I at random.

In Figure 7, we present Alg. `RelNonCrit` $- \log n$ `bits`, which assumes that Algorithm $\mathcal{A}$ is a deterministic $n^{1-\epsilon}$-approximation algorithm that leaks at most $\frac{\epsilon}{6}\log n$ bits. Alg. `RelNonCrit` $- \log n$ `bits` is a randomized algorithm which returns a correct answer with probability at least $1-\delta$, for some $0 < \delta < 1$. Given a graph G with n vertices, we construct the graphs from Definition 5.2. These graphs have $N \stackrel{\text{def}}{=} (12n/\delta)^{2/\epsilon}$ vertices, $2m$ copies of G, and a disjoint set of vertices I. We choose the number of copies to be $2m$, where

$$m \stackrel{\text{def}}{=} N^{\epsilon/6} = (12n/\delta)^{1/3}. \tag{3}$$

The size of the set I is $|I| = N - 2mn$. In the proof of Claim 5.7 it would become clear why we made these choices.

Alg. `RelNonCrit` $- \log n$ `bits`

INPUT: A graph $G = \langle V, E\rangle$, a vertex $v \in V$, and a number $0 < \delta < 1$.
OUTPUT: One of the following: (i) The vertex v is relevant for G. (ii) The vertex v is not critical for G. The algorithms errs with probability at most δ.

1. Choose distinct $i_1, \dots, i_m, j_1, \dots, j_{2m}$ at random from I.
2. For $t = 0$ to m do:
   (a) Let $G_t = G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$.
   (b) If $\mathcal{A}(G_t) \cap \{i_{t+1}, j_{2t+1}\} \neq \emptyset$, then return "Non Critical."
3. ($*$ all $m+1$ sets W_t do not contain the two vertices $*$)
   RETURN "Relevant."

Figure 7: Alg. `RelNonCrit` $- \log n$ `bits`.

In the following we assume that, on graphs with n vertices, $\mathcal{A}$ leaks at most $k(n) = \frac{\epsilon}{6}\log n$ bits. Recall that we execute $\mathcal{A}$ on graphs with N vertices, thus, the approximation ratio is $N^{1-\epsilon}$ and the leakage is bounded by $k(N) = \frac{\epsilon}{6}\log N = \log m$. The next sequence of claims asserts the correctness of Alg. `RelNonCrit` $- \log n$ `bits`.

Claim 5.6 *If v is non-relevant, then Alg.* `RelNonCrit` $-\log n$ `bits` *does not return "Relevant."*

Proof: If the algorithms returns "Relevant," then the for loop finishes. Thus, the sets $\mathcal{A}(G_0), \ldots, \mathcal{A}(G_m)$ are $m+1 = 2^k+1$ different sets; for $t' < t$ the cover $\mathcal{A}(G_t)$ must contain at least one of the vertices $i_{t'+1}, j_{2t'+1}$ (as it must cover the edge $(i_{t'+1}, j_{2t'+1})$), while $\mathcal{A}(G_{t'})$ contains none of them. Thus, there are 2^k+1 graphs with pair-wise different answers of $\mathcal{A}$, and, since $\mathcal{A}$ leaks at most k bits, at least two of the graphs are not equivalent according to $\mathcal{R}_{\text{minVC}}$. Thus, by Claim 5.4, the vertex v is relevant for G. $\square$

Claim 5.7 *Let $0 \leq t \leq m$. For every $i_1, \ldots, i_t, j_1, \ldots, j_{2t} \in I$, consider the following experiment: choose $i_{t+1}, \ldots, i_m$ and j_{2t+1} at random with uniform distribution from I and return 1 if*

$$\mathcal{A}(G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})) \cap \{i_{t+1}, j_{2t+1}\} \neq \emptyset.$$

If v is critical and $N = (12n/\delta)^{2/\epsilon}$, then the probability that this experiment returns 1 is at most δ/m.

Proof: We first prove that we chose N, the number of vertices in the graphs we construct, such that the size of each cover $\mathcal{A}$ returns is at most $|I|/\alpha$, where $\alpha \stackrel{\text{def}}{=} 2m^2/\delta$. We will need the following requirement for our computations:

$$|I| = N - 2mn \geq N/2. \tag{4}$$

That is, we need to show that $N/2 \geq 2mn$. As $m = N^{\epsilon/6} \leq N^{1/6}$, it suffices to require $N \geq (4n)^{6/5}$. By the choice of N

$$N = (12n/\delta)^{2/\epsilon} \geq (12n)^2 \geq (4n)^{6/5}$$

(since $0 < \delta, \epsilon \leq 1$), thus (4) holds.

We next upper-bound the size of the covers that $\mathcal{A}$ outputs. The size of a minimum vertex cover of these graphs is at most $(2n+1)m \leq 3nm$. Since $\mathcal{A}$ is an $N^{1-\epsilon}$-approximation algorithm for vertex cover, the size of its output is at most $3 \cdot nm \cdot N^{1-\epsilon}$. Recall that $N = (\frac{12n}{\delta})^{2/\epsilon}$ and $m = N^{\epsilon/6}$, and thus, by (3),

$$12nm^3 = \delta(12n/\delta) \cdot N^{\epsilon/2} = \delta((12n/\delta)^{2/\epsilon})^{\epsilon/2} \cdot N^{\epsilon/2} = \delta N^{\epsilon/2} N^{\epsilon/2} = \delta N^{\epsilon}. \tag{5}$$

Therefore, the size of the cover returned by $\mathcal{A}$ is at most

$$3 \cdot nm \cdot N^{1-\epsilon} = \frac{3nmN}{N^{\epsilon}} \leq \frac{6nm|I|}{N^{\epsilon}} = \frac{12nm^3}{\delta N^{\epsilon}} \cdot \frac{\delta|I|}{2m^2} = \frac{\delta|I|}{2m^2} = \frac{|I|}{\alpha},$$

where the inequality above follows from (4) and the last equality follows (5). Thus, with probability at most $1/\alpha$, a random vertex from I is in a given answer.

We now analyze the probability that the experiment returns 1. If v is critical, then, by Claim 5.3, every two choices of $i_{t+1}, \ldots, i_m$ result in an equivalent graphs according to $\mathcal{R}_{\text{minVC}}$. Since $\mathcal{A}$ leaks at most $k(N) = \frac{\epsilon}{6}\log N = \log m$ bits, there are at most $2^{k(N)} = m$ different answers for all the different choices of $i_{t+1}, \ldots, i_m$. Thus, the size of the union of all the answers is at most $\frac{m|I|}{\alpha}$. The probability that any of the two vertices i_{t+1}, j_{2t+1} are in the union of the m answers is, by the union bound, at most $\frac{2m}{\alpha}$. Thus, the probability of the experiment returning 1 is at most $\frac{2m^2}{\alpha} = \delta/m$, as required. $\square$

Claim 5.8 *If vertex v is critical, then the probability that Alg.* `RelNonCrit` $-\log n$ `bits` *returns "Non Critical" is at most δ.*

Proof: Alg. $\texttt{RelNonCrit} - \log n\ \texttt{bits}$ repeats $2^k + 1$ the experiment of Claim 5.7 and returns "Non Critical" if one of the experiments returns 1. Thus, by Claim 5.7 and the union-bound, if vertex v is critical, then probability that Alg. $\texttt{RelNonCrit} - \log n\ \texttt{bits}$ returns "Non Critical" is at most δ. $\square$

Alg. $\texttt{RelNonCrit} - \log n\ \texttt{bits}$ executes m times the polynomial algorithm $\mathcal{A}$ on graphs with N vertices, where $N = (12n/\delta)^{2/\epsilon}$. Thus,

Claim 5.9 *If $\mathcal{A}$ runs in polynomial time and $0 \leq \epsilon < 1$ is a constant, then the running time of Alg.* $\texttt{RelNonCrit} - \log n\ \texttt{bits}$ *is* $\mathrm{poly}(n/\delta)$.

We summarize the results of this section in the next lemma.

Lemma 5.10 *Let $\epsilon > 0$ be a constant. If $\mathcal{RP} \neq \mathcal{NP}$, then there is no deterministic $n^{1-\epsilon}$-approximation algorithm for the search problem of* vertex cover *that leaks at most $\frac{\epsilon}{6} \log n$ bits.*

Proof: `Algorithm Greedy Vertex Cover`, which solves vertex cover, executes at most n times Alg. $\texttt{RelNonCrit} - \log n\ \texttt{bits}$ with graphs of size at most n. We execute these calls with $\delta = \frac{1}{4n}$ (where n is the original number of vertices in G), thus, all together the error is at most $1/4$. By Claim 5.9, the running time of `Algorithm Greedy Vertex Cover` is polynomial. Thus, if there is an $n^{1-\epsilon}$-approximation algorithm for the search problem of vertex cover that leaks at most $k(n) = \frac{\epsilon}{6} \log n$ bits, then there is a polynomial time randomized algorithm for minimum vertex cover that errs with probability $1/4$. This implies that $\mathcal{NP} \subseteq \mathcal{BPP}$.

To contradict $\mathcal{RP} \neq \mathcal{NP}$, this algorithm is transformed to a one-sided error algorithm for the decision problem of vertex cover: Given $\langle G, s\rangle$ decide if G has a vertex cover of size at most s. The transformation is simple; execute the algorithm for the search problem of vertex cover. If this algorithm returns a set that covers G and its size is at most s return "yes." $\square$

6 Impossibility Result for Private Computing of a Search Problem in $\mathcal{P}$

An algorithm solving a search problem can return any solution to the problem. An algorithm solving a search problem *privately* has additional requirements on the solutions that it returns. In this section, we show that these additional requirements can make the problem much harder. That is, we show that there is a polynomial relation Q whose search problem is in $\mathcal{P}$, however, unless $\mathcal{NP} \subseteq \mathcal{P}/\mathrm{poly}$, there is no polynomial time private algorithm for Q with respect to $\mathcal{R}_Q$.

Definition 6.1 *Let G be a graph and $C_1, C_2 \subseteq V$. We define the relation Q as follows: $\langle G, C_1\rangle$ and C_2 are in Q (that is, $\langle\langle G, C_1\rangle, C_2\rangle \in Q$) if $|C_1| = |C_2|$ and C_1 and C_2 are cliques in G.*

Clearly, the search problem of Q is easy, given $\langle G, C_1\rangle$ return C_1. Assume that there is a private algorithm for $\mathcal{R}_Q$. That is, if C_1 and C_2 are two disjoint cliques of the same size in a graph G, then a private algorithm has to return the same output distribution on $\langle G, C_1\rangle$ and $\langle G, C_2\rangle$. Intuitively, this implies that given a clique in the graph, there is an efficient algorithm that finds another clique. We will prove that this is impossible unless $\mathcal{NP} \subseteq \mathcal{P}/\mathrm{poly}$.

Theorem 6.2 *If $\mathcal{NP} \not\subseteq \mathcal{P}/\mathrm{poly}$, then there is no polynomial-time private algorithm for the search problem of Q with respect to the privacy structure $\mathcal{R}_Q$.*

Proof: We assume towards contradiction that there exists a polynomial-time private algorithm $\mathcal{A}$ for the search problem of Q with respect to the privacy structure $\mathcal{R}_Q$. We will use $\mathcal{A}$ to prove that the $\mathcal{NP}$-complete problem CLIQUE is in $\mathcal{P}$/poly. That is, we construct a sequence of polynomial-length advice strings $\langle a_n \rangle_{n\in\mathbb{N}}$ and a polynomial time algorithm $\mathcal{B}$ such that, given a graph G with n vertices, an integer k, and the advice a_n, Algorithm $\mathcal{B}$ decides if G contains a clique of size k.

Given two graph $G_0 = \langle V_0, E_0 \rangle$ and $G_1 = \langle V_1, E_1 \rangle$, where $V_1 \cap V_2 = \emptyset$, we define their disjoint union $G_0 \cup G_1$ as the graph $G = \langle V, E \rangle$ where $V = V_0 \cup V_1$ and $E = E_0 \cup E_1$. Every clique in $G_0 \cup G_1$ is either a clique in G_0 or a clique in G_1. Assume that C_0 and C_1 are cliques of size k in G_0 and G_1 respectively. Then, $\langle G_0 \cup G_1, C_0 \rangle$ and $\langle G_0 \cup G_1, C_1 \rangle$ have the same set of cliques of size k, that is, they are in $\mathcal{R}_Q$.

Algorithm $\mathcal{A}'$

INPUT: Two graphs G_0, G_1 with disjoint sets of vertices and a set C.
PROMISE: C is a clique in $G_0 \cup G_1$.

1. Execute $\mathcal{A}$ on $\langle G_0 \cup G_1, C \rangle$
   (a) Let j be the index such that $\mathcal{A}$ returns a clique in G_j.
   (b) Return j.

Figure 8: Algorithm $\mathcal{A}'$ which gets two graphs and a clique, executes $\mathcal{A}$ on their union, and checks in which graph $\mathcal{A}$ returns a clique.

As a first step, we construct in Figure 8 an algorithm $\mathcal{A}'$ that will be used to construct Algorithm $\mathcal{B}$ and the advice strings. This algorithm gets two graphs G_0, G_1 and a clique C in one of them, executes $\mathcal{A}$ on their union, and checks in which graph $\mathcal{A}$ returns a clique. If only one graph G_i has a clique of size $|C|$, then, by the correctness requirement of $\mathcal{A}$, Algorithm $\mathcal{A}'$ must return i. However, if both graphs have a clique of size $|C|$, then, by the privacy requirement of $\mathcal{A}$, the output distribution of $\mathcal{A}'$ is approximately the same when C is a clique in G_0 and when C is a clique in G_1. That is, there is an integer n_0 such that for every pair of graphs with at least n_0 vertices, the difference between the probabilities that $\mathcal{A}'$ returns 0 in the two cases is small (for concreteness, at most $1/3$).

We construct the advice string a_n as the concatenation of advice strings $a_{n,1}, \ldots, a_{n,n}$ where $a_{n,k}$ is used to decide if a graph G with n vertices has a clique of size k.[5] Fix an integer n and an integer k, where $n \geq n_0$ and $1 \leq k \leq n$. For every graph G with n vertices that has a clique of size k, we fix some clique C_G of size k in G. Let G_0 and G_1 be two graphs that have a clique of size k.

Definition 6.3 *We say that G_0 loses to G_1 if Algorithm $\mathcal{A}'$ returns 1 on input $\langle G_0, G_1, C_{G_0} \rangle$ with probability at least $1/3$.*

That is, G_0 loses to G_1 if, when given a clique in G_0, Algorithm $\mathcal{A}$ "magically" manages to return with probability $1/3$ a clique in the graph G_1; thus, G_0, C_{G_0} can aid in finding a clique in G_1.

If G_0 does not lose to G_1, then $\mathcal{A}'$ returns 0 with probability at least $2/3$, and by the privacy requirement, with probability at least $1/3$ Algorithm $\mathcal{A}'$ returns 0 on input $\langle G_0, G_1, C_{G_1} \rangle$. That is,

Claim 6.4 *Let G_0 and G_1 be two graphs with n vertices that have a clique of size k. If G_0 does not lose to G_1, then G_1 loses to G_0. (It is possible that G_0 loses to G_1 and G_1 loses to G_0.)*

[5] We assume that any two graphs with n vertices have disjoint sets of vertices. E.g., we can order all graphs with n vertices according to some order, and the vertices of the ith graph in this order are $\{1, \ldots, n\} \times \{i\}$.

Thus, for every set of graphs with n vertices and a clique of size k, there exists a graph G_0 in the set that loses to at least half the graphs in the set. This is the idea of constructing the advice string $a_{n,k}$.

Construction of $a_{n,k}$

1. $a_{n,k} \leftarrow \emptyset$.
2. Initialize L as the set of all graphs with n vertices that have a clique of size k.
3. While $L \neq \emptyset$ do
   (a) Choose a graph G in L that loses to at least half of the graphs in L.
   (b) $a_{n,k} \leftarrow a_{n,k} \cup \{\langle G, C_G \rangle\}$.
   (c) $L \leftarrow L \setminus \{G_1 : G \text{ loses to } G_1\}$.

Since there are $2^{O(n^2)}$ graphs with n vertices, the advice $a_{n,k}$ contains $O(n^2)$ graphs. We are ready to describe the non-uniform algorithm $\mathcal{B}$ that, given a graph G with n vertices, an integer k, and the advice a_n, decides if G contains a clique of size k.

Algorithm $\mathcal{B}$

INPUT: G, k, and $a_{n,k}$.

1. For every G_0 in $a_{n,k}$ execute $\mathcal{A}'$ on $\langle G_0, G, C_{G_0} \rangle$.
2. If there exists an execution of $\mathcal{A}'$ which returns 1, return "G has a clique of size k,"
3. Otherwise, return "FAIL."

If G does not have a clique of size k, then, by the correctness of $\mathcal{A}$, Algorithm $\mathcal{A}'$ always returns 0, and $\mathcal{B}$ always returns "FAIL." If G has a clique of size k, then there exists a graph G_0 in $a_{n,k}$ that loses to G, thus, with probability at least $1/3$, Algorithm $\mathcal{A}'$ returns 1 on input $\langle G_0, G, C_{G_0} \rangle$, and, thus, with probability at least $1/3$, Algorithm $\mathcal{B}$ returns "G has a clique of size k." That is, $\mathcal{B}$ is a randomized algorithm with advice strings that decides if $\langle G, k \rangle \in$ CLIQUE with a one-sided error of $2/3$. By a standard amplification and union-bound arguments, we can get a deterministic polynomial-time non-uniform algorithm that decides if $\langle G, k \rangle \in$ CLIQUE without error. $\square$

Acknowledgments. We thank Yinnon Haviv for helpful discussions and Robi Krauthgamer for pointing out that we can use almost k-independent spaces.

References

[1] N. Alon, L. Babai, and A. Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms*, 7(4):567 – 583, 1986.

[2] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3:289–304, 1992.

[3] R. Bar-Yehuda, B. Chor, E. Kushilevitz, and A. Orlitsky. Privacy, additional information, and communication. *IEEE Trans. on Information Theory*, 39(6):1930–1943, 1993.

[4] R. Bar-Yehuda and S. Even. A local-ratio theorem for approximating the weighted vertex cover problem. *Annals of Disc. Math.*, 25:27–46, 1985.

[5] M. Bellare and E. Petrank. Making zero-knowledge provers efficient. In *Proc. of the 24th ACM Symp. on the Theory of Computing*, pages 711–722, 1992.

[6] B. Chor, J. Friedmann, O. Goldreich, J. Hastad, S. Rudich, and R. Smolansky. The bit extraction problem or t-resilient functions. In *Proc. of the 26th IEEE Symp. on Foundations of Computer Science*, pages 396–407, 1985.

[7] I. Dinur and S. Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, 162(1), 2005.

[8] J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. J. Strauss, and R. N. Wright. Secure multiparty computation of approximations. In P. G. Spirakis and J. van Leeuven, editors, *Proc. of the 28th International Colloquium on Automata, Languages and Programming*, volume 2076 of *Lecture Notes in Computer Science*, pages 927–938. Springer-Verlag, 2001.

[9] M. J. Freedman, K. Nissim, and B. Pinkas. Efficient private matching and set intersection. In C. Cachin and J. Camenisch, editors, *Advances in Cryptology – EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, pages 1–19. Springer-Verlag, 2004.

[10] O. Goldreich, S. Micali, and A. Wigderson. How to play any mental game. In *Proc. of the 19th ACM Symp. on the Theory of Computing*, pages 218–229, 1987.

[11] O. Goldreich, R. Ostrovsky, and E. Petrank. Computational complexity and knowledge complexity. *SIAM J. on Computing*, 27(4):1116–1141, 1998.

[12] O. Goldreich and E. Petrank. Quantifying knowledge complexity. *Computational Complexity*, 8(1):50–98, 1999. Preliminary version appeared in FOCS 91.

[13] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof systems. *SIAM J. on Computing*, 18(1):186–208, 1989.

[14] S. Halevi, R. Krauthgamer, E. Kushilevitz, and K. Nissim. Private approximation of NP-hard functions. In *Proc. of the 33th ACM Symp. on the Theory of Computing*, pages 550–559, 2001.

[15] E. Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. In *Proc. of the 11th ACM-SIAM Symp. on Discrete Algorithms*, pages 329–337, 2000.

[16] J. Håstad. Some optimal inapproximability results. *J. of the ACM*, 48(4):798–859, 2001.

[17] P. Indyk and D. Woodruff. Polylogarithmic private approximations and efficient matching. In S. Halevi and T. Rabin, editors, *Proc. of the Third Theory of Cryptography Conference – TCC 2006*, volume 3876 of *Lecture Notes in Computer Science*, pages 245–264. Springer-Verlag, 2006.

[18] D. S. Johnson. Approximation algorithms for combinatorial problems. *J. of Computer and System Sciences*, 9:256–278, 1974.

[19] E. Kiltz, G. Leander, and J. Malone-Lee. Secure computation of the mean and related statistics. In J. Kilian, editor, *Proc. of the Second Theory of Cryptography Conference – TCC 2005*, volume 3378 of *Lecture Notes in Computer Science*, pages 283–302. Springer-Verlag, 2005.

[20] M. Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM J. on Computing*, 15(4):1036–1055, 1986.

[21] B. Monien and E. Speckenmeyer. Ramsey numbers and an approximation algorithm for the vertex cover problem. *Acta Inf.*, 22:115–123, 1985.

[22] J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. on Computing*, 22(4):838–856, 1993.

[23] C. H. Papadimitriou and M. Yannakakis. On limited nondeterminism and the complexity of the V-C dimension. *J. of Computer and System Sciences*, 53(2):161–170, 1996.

[24] E. Petrank and G. Tardos. On the knowledge complexity of NP. *Combinatorica*, 22(1):83–121, 2002.

[25] A. C. Yao. Protocols for secure computations. In *Proc. of the 23th IEEE Symp. on Foundations of Computer Science*, pages 160–164, 1982.

A Semantic-Security Flavored Definitions of Private Algorithms

In Definition 2.2, private algorithms are defined as those whose outputs on $\mathcal{R}$-equivalent instances are computationally indistinguishable. To shed a little more light on this definition we now compare it with two other definitions. Definition A.1 is of semantically-secure private algorithms, where the outcome of the private algorithm $\mathcal{A}$ on x is simulated by an efficient simulator that is given access to an arbitrary representative y of the equivalence class of x. Definition A.2 is a (seemingly) weakening of this definition, where the simulator is potentially much more powerful. Here, the simulator is given access to an arbitrary oracle O of its choice, that answers queries about the equivalence class of x. For example, one may choose an oracle O that enumerate in lexicographic order the values y such that $x \equiv_{\mathcal{R}} y$, hence empowering the simulator to perform computations that may be otherwise intractable. It turns out that Definitions 2.2, A.1, and A.2 are equivalent. This gives a further indication that Definition 2.2 is rather a minimal notion of privacy.

Definition A.1 (Private Algorithm – Semantic Security) *Let $\mathcal{R}$ be a privacy structure. A probabilistic polynomial time algorithm $\mathcal{A}$ is* semantically private *with respect to $\mathcal{R}$ if there exists a probabilistic polynomial time simulator $\mathcal{S}$, such that for every polynomial-time algorithm $\mathcal{D}$ and for every positive polynomial $p(\cdot)$, there exists some $n_0 \in \mathbb{N}$ such that for every $x, y \in \{0,1\}^*$ where $x \equiv_{\mathcal{R}} y$ and $|x| = |y| \geq n_0$*

$$|\Pr[\mathcal{D}(\mathcal{A}(x), x, y) = 1] - \Pr[\mathcal{D}(\mathcal{S}(y), x, y) = 1]| \leq \frac{1}{p(|x|)}.$$

That is, the simulator on input y, outputs a distribution that is indistinguishable from the output of $\mathcal{A}$ on x.

For a string $y \in \{0,1\}^*$ define $Eq(y) = \{z : y \equiv_{\mathcal{R}} z\}$ to be its equivalence class. Let $O : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}^*$ be an oracle. We use the notation $O(Eq(y), \cdot)$ to denote the oracle obtained from O by "hardwiring" its first input to $Eq(y)$. Informally, O is a collection of oracles – one per each equivalence class. The oracle $O(Eq(y), \cdot)$ answers queries about the equivalence class of y. For an oracle machine S, we use $S^{O(Eq(y),\cdot)}$ to denote the Turing Machine S with oracle access to $O(Eq(y), \cdot)$. This way, S may use O to learn facts about the equivalence class of y.

Definition A.2 (Private Algorithm – Weak Semantic Security) *Let $\mathcal{R}$ be a privacy structure. A probabilistic polynomial time algorithm $\mathcal{A}$ is* weakly semantically private *with respect to $\mathcal{R}$ if there exists an oracle $O : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}^*$ and a probabilistic polynomial time oracle machine*[6] *$\mathcal{S}$ (the simulator), such that for every polynomial-time algorithm $\mathcal{D}$ and for every positive polynomial $p(\cdot)$, there exists some $n_0 \in \mathbb{N}$ such that for every $x, y \in \{0,1\}^*$ such that $|x| = |y| \geq n_0$ and $x \equiv_{\mathcal{R}} y$:*

$$\left|\Pr[\mathcal{D}(\mathcal{A}(x), x, y) = 1] - \Pr[\mathcal{D}(\mathcal{S}^{O(Eq(y),\cdot)}, x, y) = 1]\right| \leq \frac{1}{p(|x|)}.$$

That is, the simulator with oracle access to $O(Eq(y), \cdot)$, outputs a distribution that is indistinguishable from the output of $\mathcal{A}$ on x.

Theorem A.3 *Definitions 2.2, A.1, and A.2 are equivalent.*

Proof: Let $\mathcal{A}$ be an algorithm that is private with respect to a privacy structure $\mathcal{R}$ according to Definition 2.2. Taking The simulator $\mathcal{S}$ to be $\mathcal{A}$ itself shows $\mathcal{A}$ is private according to Definition A.1 as well. Privacy under Definition A.1 implies privacy under Definition A.2, as any simulator $\mathcal{S}$ in view of Definition A.1 can be transformed into an oracle simulator that asks for a representative of the class and proceeds like $\mathcal{S}$.

Finally, we prove that privacy under Definition A.2 implies privacy under Definition 2.2. Let $\mathcal{A}$ be an algorithm that is private under Definition A.2. Let x and y be inputs such that $x \equiv_{\mathcal{R}} y$, $\mathcal{S}$ be the corresponding simulator and $\mathcal{D}$ be a distinguishing algorithm. By privacy under Definition A.2, the advantage of $\mathcal{D}$ in distinguishing $(\mathcal{A}(x), x, y)$ and $(\mathcal{S}^{O(Eq(y),\cdot)}, x, y)$ is negligible. Furthermore, replacing the roles of x and y, the advantage of $\mathcal{D}$ in distinguishing $(\mathcal{A}(y), x, y)$ and $(\mathcal{S}^{O(Eq(x),\cdot)}, x, y)$ is negligible. Therefore, as $Eq(x) = Eq(y)$, we get that the advantage of $\mathcal{D}$ in distinguishing $(\mathcal{A}(x), x, y)$ and $(\mathcal{A}(y), x, y)$ is also negligible, satisfying Definition 2.2. □

B Impossibility Result for Randomized Private Approximation of Vertex Cover

In this section, we generalize the inapproximability results of Section 5.2 to *randomized* private protocols that leak $O(\log n)$ bits. We follow a similar strategy as in the proof of Lemma 5.10 for deterministic private protocols that leak $O(\log n)$ bits. In that proof we checked if $\mathcal{A}(G_t) \cap \{i_{t+1}, j_{2t+1}\} \neq \emptyset$. Here, we execute $\mathcal{A}(G_t)$ many times and estimate the probability that $\mathcal{A}(G_t) \cap \{i_{t+1}, j_{2t+1}\} \neq \emptyset$.

In Figure 9, we present `Alg. RelNonCrit` – `randomized` $\log n$ `bits`, which assumes that Algorithm $\mathcal{A}$ is a possibly randomized $n^{1-\epsilon}$-approximation algorithm that leaks at most $\frac{\epsilon}{6}\log n$ bits. `Alg. RelNonCrit` – `randomized` $\log n$ `bits` is a randomized algorithm which returns a correct answer with probability at least $1 - \delta$, for a given $0 < \delta < 1$. Given a graph G with n vertices, we construct the graphs from Definition 5.2. These graphs have $N = (120n/\delta)^{2/\epsilon}$ vertices, $2m$ copies of G, and a disjoint set of vertices I. We choose the number of copies to be $2m$, where

$$m \stackrel{\text{def}}{=} N^{\epsilon/6} = (120n/\delta)^{1/4}. \tag{6}$$

Therefore, the size of the set I is $|I| = N - 2mn$. In the proof of Claim B.4 it would become clear why we made these choices.

[6] the choice of probabilistic polynomial time is arbitrary, as S can delegate computations to the oracle O.

`Alg. RelNonCrit` $-$ `randomized` $\log n$ `bits`

INPUT: A graph $G = \langle V, E \rangle$, a vertex $v \in V$, and a number $0 < \delta < 1$.
OUTPUT: One of the following: (i) The vertex v is relevant for G. (ii) The vertex v is not critical for G. The algorithms errs with probability at most δ.

1. Choose distinct $i_1, \ldots, i_m, j_1, \ldots, j_{2m}$ at random from I.
2. For $t = 0$ to $m - 1$ do:
   (a) Let $G_t = G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$.
   (b) Execute n times $\mathcal{A}(G_t)$.
   (c) If one of the vertices i_{t+1}, j_{2t+1} appears in more than $0.3n$ of the n answers of $\mathcal{A}$, then return "Non Critical."
3. RETURN "Relevant."

Figure 9: `Alg. RelNonCrit` $-$ `randomized` $\log n$ `bits`.

In the following we assume that, on graphs with n vertices, $\mathcal{A}$ leaks at most $k(n) = \frac{\epsilon}{6} \log n$ bits. Recall that we execute $\mathcal{A}$ on graphs with N vertices, thus, the approximation ratio is $N^{1-\epsilon}$ and the leakage is bounded by $k(N) = \frac{\epsilon}{6} \log N = \log m$. The next sequence of claims asserts the correctness of Alg. `RelNonCrit` $-\log n$ `bits`.

Let $0 \le p \le 1$. We say that a vertex v is p-light for a graph H if $\Pr[v \in \mathcal{A}(H)] < p$ and v is p-heavy for H if $\Pr[v \in \mathcal{A}(H)] \ge p$. Let $\mathcal{R}'$ be a privacy structure that is a $k(N)$-refinement of $\mathcal{R}_{\text{minVC}}$ such that $\mathcal{A}$ is private with respect to $\mathcal{R}'$. The next observation follows from Chernoff's inequality.

Observation B.1 *Assume we execute n times $\mathcal{A}(H)$.*

- *If a vertex u is 0.4-heavy for H, then the probability that u appears in at most $0.3n$ of the n answers of $\mathcal{A}$ is $2^{-O(n)}$.*
- *If a vertex u is 0.2-light for H, then the probability that u appears in more than $0.3n$ of the n answers of $\mathcal{A}$ is $2^{-O(n)}$.*

A distinguishing algorithm can approximate the probability that a vertex is in $\mathcal{A}(H)$. Hence, for every two graphs H_0, H_1 such that $H_0 \equiv_{\mathcal{R}'} H_1$ and for every vertex u the probability that $u \in \mathcal{A}(H_0)$ is approximately the same as the probability that $u \in \mathcal{A}(H_1)$. This is formalized in the following claim.

Claim B.2 *Let $0 \le p < 1$ and $0 < \gamma < 1 - p$ be constants. Then, there exists $n_0 \in \mathbb{N}$ such that if H_0 and H_1 are two graphs with $n \ge n_0$ vertices and there exists a vertex w that is p-light for H_0 and is $(p + \gamma)$-heavy for H_1, then H_0 and H_1 are not equivalent according to $\mathcal{R}'$.*

Proof: Consider the follwing distuinguising algorithm $\mathcal{D}$:

INPUT: Two graphs Γ_0, Γ_1 and a cover $C = \mathcal{A}(\Gamma_j)$ for some $j \in \{0,1\}$.

1. For $i \in \{0,1\}$, execute n times $\mathcal{A}(\Gamma_i)$.
2. For every $u \in V$ and $i \in \{0,1\}$, let $\tilde{p}_u^i$ be the fraction of executions in which u appears in the answers of $\mathcal{A}(\Gamma_i)$.
3. Choose a vertex u such that $\tilde{p}_u^1 - \tilde{p}_u^0 > 2\gamma/3$ (if no such u exists return 1).
4. If $u \in C$ return 1 otherwise return 0.

For every $v \in V$ and $i \in \{0,1\}$ define $p_v^i \stackrel{\text{def}}{=} \Pr[v \in \mathcal{A}(\Gamma_i)]$. By Chernoff's inequality, with probability $1 - 2^{-O(n)}$, for every $v \in V$ and $i \in \{0,1\}$,

$$|p_v^i - \tilde{p}_v^i| < \gamma/6. \tag{7}$$

Now consider an execution of $\mathcal{D}$ on the graphs H_0 and H_1, and consider the vertex w guaranteed by the claim. Assume that (7) holds for every vertex in the graph. In particular,

$$\tilde{p}_w^1 - \tilde{p}_w^0 \;=\; p_w^1 - p_w^0 - (p_w^1 - \tilde{p}_w^1) - (\tilde{p}_w^0 - p_w^0) \;\geq\; \gamma - 2\gamma/6 \;=\; 2\gamma/3.$$

Thus, the probability that there is no such u in Step 3 is negligible. Furthermore, for the u chosen in Step 3 with overwhelming probability $p_u^1 - p_u^0 = \tilde{p}_w^1 - \tilde{p}_w^0 - (\tilde{p}_w^1 - p_w^1) - (p_w^0 - \tilde{p}_w^0) \geq \gamma/3$. Therefore,

$$|\Pr[\mathcal{D}(\mathcal{A}(\Gamma_1), \Gamma_0, \Gamma_1) = 1] - \Pr[\mathcal{D}(\mathcal{A}(\Gamma_0), \Gamma_0, \Gamma_1) = 1]| \geq \gamma/3 - 2^{-O(n)}.$$

This implies that H_0 and H_1 are not equivalent according to $\mathcal{R}'$, since $\mathcal{A}$ is private with respect to $\mathcal{R}'$. □

The following claim is analogous to Claim 5.6 for the deterministic case.

Claim B.3 *If a vertex v is non-relevant, then the probability that* `Alg. RelNonCrit` $-$ `randomized` $\log n$ `bits` *returns "Relevant" is negligible.*

Proof: Since the vertex v is non-relevant for G, by Claim 5.4, for every $0 \leq t' < t \leq m$ the graphs $G_{t'}$ and G_t are equivalent according to $\mathcal{R}_{\text{minVC}}$. As $\mathcal{A}$ leaks at most $k(n) = \log m$ bits, every equivalence class of $\mathcal{R}_{\text{minVC}}$ is partitioned into at most m equivalence classes of $\mathcal{R}'$. Since there are $m + 1 = 2^k + 1$ graphs $G_0, \ldots, G_m$, there are two indices $0 \leq t' < t \leq m$, such that $G_{t'}$ and G_t are equivalent according to $\mathcal{R}'$.

As every cover of G_t must cover the edge $(i_{t'+1}, j_{2t'+1})$, at least one of $i_{t'+1}, j_{2t'+1}$ is 0.5-heavy for G_t. Applying Claim B.2 to G_t and $G_{t'}$, at least one of the vertices $i_{t'}, j_{2t'+1}$ is 0.4-heavy for $G_{t'}$. Thus, by Observation B.1, with high probability this vertex will appear at least $0.3n$ times in the answers of $\mathcal{A}(G_{t'})$, and the algorithm will return "Non-Critical." □

To prove that `Alg. RelNonCrit` $-$ `randomized` $\log n$ `bits` is correct when v is critical, we need the following claim.

Claim B.4 *Let $0 \leq t < m$. For every $i_1, \ldots, i_t, j_1, \ldots, j_{2t} \in I$, consider the following experiment: choose $i_{t+1}, \ldots, i_m$ and j_{2t+1} at random and with uniform distribution from I and return 1 if at least one of the vertices i_{t+1}, j_{2t+1} is 0.2-heavy for $G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$. If v is critical, then the probability that this experiment returns 1 is at most δ/m.*

Proof: We chose N, the number of vertices in the graphs we construct, such that the size of each cover $\mathcal{A}$ returns is at most $|I|/(10\alpha)$ for $\alpha \stackrel{\text{def}}{=} 2m^2/\delta$. To show this, we shell need the following requirement:

$$|I| = N - 2mn \geq N/2. \tag{8}$$

That is, we need $N \geq 4mn$. As $m = N^{\epsilon/6} \leq N^{1/6}$, it suffices to require $N \geq (4n)^{6/5}$. By the choice of N,

$$N = (120n/\delta)^{2/\epsilon} \geq (120n)^2 \geq (4n)^{6/5}$$

(since $0 < \delta, \epsilon \leq 1$), thus (8) holds.

We next upper-bound the size of the covers that $\mathcal{A}$ outputs. The size of a minimum vertex cover of these graphs is at most $(2 \cdot n + 1)m \leq 3nm$. Since $\mathcal{A}$ is an $N^{1-\epsilon}$-approximation algorithm for vertex cover, the size of its output is at most $3 \cdot nm \cdot N^{1-\epsilon}$. Recall that $N = (\frac{120n}{\delta})^{2/\epsilon}$ and $m = N^{\epsilon/6}$, and thus, using (6),

$$120nm^3 = \delta(120n/\delta) \cdot N^{\epsilon/2} = \delta((120n/\delta)^{2/\epsilon})^{\epsilon/2} \cdot N^{\epsilon/2} = \delta N^{\epsilon/2} N^{\epsilon/2} = \delta N^{\epsilon}. \tag{9}$$

Therefore, the size of the cover returned by $\mathcal{A}$ is at most

$$3 \cdot nm \cdot N^{1-\epsilon} \;=\; \frac{3nmN}{N^{\epsilon}} \leq \frac{6nm|I|}{N^{\epsilon}} \;=\; \frac{120nm^3}{\delta N^{\epsilon}} \cdot \frac{\delta|I|}{20m^2} \;=\; \frac{\delta|I|}{20m^2} \;=\; \frac{|I|}{10\alpha},$$

where the inequality above follows from (8) and the last two equalities follow from (9) and the choice of α.

Therefore, for every $i_{t+1}, \ldots, i_m, j_{2t+1}, \ldots, j_{2m}$ there are at most $|I|/\alpha$ vertices that are 0.1-heavy for $G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$. We will show that if v is critical, then the number of vertices that are 0.2-heavy for $G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$ for some $i_{t+1}, \ldots, i_m, j_{2t+1}, \ldots, j_{2m}$ is at most $m|I|/\alpha$.

We next prove an upper bound on the number of 0.2 heavy vertices in a given equivalence class of $\mathcal{R}'$. Fix $i_{t+1}, \ldots, i_m, j_{2t+1}, \ldots, j_{2m}$ and let $H_0 = G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$. If a vertex u is 0.2-heavy for H for some H such that $H \equiv_{\mathcal{R}'} H_0$, then by Claim B.2, it is (at least) 0.1-heavy for H_0. Thus, there are at most $|I|/\alpha$ vertices that are 0.2-heavy for some graph in equivalence class of H_0 with respect to $\equiv_{\mathcal{R}'}$.

We next bound the number of 0.2-heavy vertices for any graph of the form $G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$. If v is critical, then, by Claim 5.3, every two choices of $i_{t+1}, \ldots, i_m$ result in an equivalent graphs according to $\mathcal{R}_{\text{minVC}}$. Since $\mathcal{A}$ leaks at most $k(N) = \frac{\epsilon}{6} \log N = \log m$ bits, there are at most $2^{k(N)} = m$ equivalence classes of $\mathcal{R}'$ for the different choices of $i_{t+1}, \ldots, i_m, j_{2t+1}, \ldots, j_{2m}$. Thus, there are at most $m|I|/\alpha$ vertices that are 0.2-heavy for a graph $G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$ for some $i_{t+1}, \ldots, i_m, j_{2t+1}, \ldots, j_{2m}$.

We now analyze the probability that the experiment returns 1. The probability that any of the two vertices i_{t+1}, j_{2t+1} is 0.2-heavy for some graph is, by the union bound, at most $\frac{2m}{\alpha}$. There, the probability of the experiment returning 1 is at most $\frac{2m}{\alpha} = \delta/m$, as required. To conclude, taking $N = (120n/\delta)^{2/\epsilon}$ guarantees that the probability of the experiment returning 1 is at most δ/m.

□

Claim B.5 *If v is critical, then the probability that* `Alg. RelNonCrit` $-$ `randomized` $\log n$ `bits` *returns "Non Critical" is at most δ.*

Proof: Assume v is critical. By Claim B.4, the probability that we choose $i_1, \ldots, i_m, j_1, \ldots, j_{2m}$ such that for some $0 \leq t < m$ at least one of the vertices $i_{t+1}, \ldots, i_m, j_{2t+1}, \ldots, j_{2m}$ is 0.2-heavy for $G(\overset{j_1 j_2}{\underset{i_1}{\vee}} \cdots \overset{j_{2t-1} j_{2t}}{\underset{i_t}{\vee}} \overset{i_{t+1}}{\wedge} \cdots \overset{i_m}{\wedge})$ is at most δ. By Observation B.1, in this case the probability that `Alg. RelNonCrit` $-$ `randomized` $\log n$ `bits` returns "Non Critical" is, therefore, negligible. □

`Alg. RelNonCrit − randomized` $\log n$ `bits` executes nm times the polynomial algorithm $\mathcal{A}$ on graphs with N vertices, where $N = (120n/\delta)^{2/\epsilon}$. Therefore, we have proved the following claim.

Claim B.6 *If $\mathcal{A}$ runs in polynomial time and $0 \le \epsilon < 1$ is a constant, then the running time of* `Alg. RelNonCrit − randomized` $\log n$ `bits` *is* $\text{poly}(n/\delta)$.

We are ready to prove Theorem 5.1.

Theorem 5.1 Let $\epsilon > 0$ be a constant. If $\mathcal{RP} \neq \mathcal{NP}$, then there is no $n^{1-\epsilon}$-approximation algorithm for the search problem of vertexcover that leaks at most $\frac{\epsilon}{6}\log n$ bits.

Proof: `Algorithm Greedy Vertex Cover`, which solves vertex cover, executes at most n times `Alg. RelNonCrit − randomized` $\log n$ `bits` with graphs of size at most n. We execute these calls with $\delta = \frac{1}{4n}$ (where n is the original number of vertices in G), thus, all together the error is at most $1/4$. By Claim B.6, the running time of `Algorithm Greedy Vertex Cover` is polynomial.

Thus, if there is an $n^{1-\epsilon}$-approximation algorithm for the search problem of vertex cover that leaks at most $k(n) = \frac{\epsilon}{6}\log n$ bits, then there is a polynomial time randomized algorithm for minimum vertex cover that errs with probability $1/4$. This implies that $\mathcal{NP} \subseteq \mathcal{BPP}$.

To contradict $\mathcal{RP} \neq \mathcal{NP}$, this algorithm is transformed to a one-sided error algorithm for the decision problem of vertex cover: Given $\langle G, s\rangle$ decide if G has a vertex cover of size at most s. The transformation is simple; execute the algorithm for the search problem of vertex cover. If this algorithm returns a set that covers G and its size is at most s return "yes." □

C Negative Result for Private Approximation of Exact 3SAT

Recall that the privacy structure $\mathcal{R}_{\text{maxE3SAT}}$ contains all pairs of exact 3 CNF formulae ϕ_1, ϕ_2 over n variables for which an assignment a satisfies the maximum number of clauses in ϕ_1 iff it satisfies the maximum number of clauses in ϕ_2.

Definition C.1 (Private Approximation of maxE3SAT**)** *An algorithm $\mathcal{A}$ is a* private $c(n)$ approximation *algorithm for* maxE3SAT *if: (i) $\mathcal{A}$ runs in polynomial time. (ii) $\mathcal{A}$ is a $c(n)$-approximation algorithm for* maxE3SAT*, that is, for every exact 3 CNF formula ϕ over n variables it returns an assignment that satisfies at least $c(n)$ times the maximum number of clauses that are simultaneously satisfiable in ϕ. (iii) $\mathcal{A}$ is private with respect to privacy structure $\mathcal{R}_{\text{maxE3SAT}}$.*

We now prove Theorem 3.16 Part 1.

Theorem 3.16 Part 1 Let $\epsilon > 0$ be a constant. If $\mathcal{P} \neq \mathcal{NP}$, then there is no deterministic private $1/n^{1-\epsilon}$-approximation algorithm for the search problem of maxE3SAT.

Proof sketch: Similarly to the proof of Theorem 3.2, we will assume the existence of a deterministic private $1/n^{1-\epsilon}$-approximation algorithm $\mathcal{A}$, and use $\mathcal{A}$ to construct a deterministic algorithm for deciding the satisfiability of exact 3 CNF formulae. We emphasize that we are solving the decision problem of satisfiability and not the optimization problem of maximum satisfiability.

Definition C.2 (Relevant Assignment to a Variable) *Let ϕ be an exact 3 CNF formula over Boolean variables $x_1, \ldots, x_n$. We say that variable x_i is σ-relevant (where $\sigma \in \{0,1\}$) if there exists a maximum assignment a for ϕ such that $a(x_i) = \sigma$.*

Note that for every satisfiable formula, every variable is 0-relevant or 1-relevant (or both). Furthermore, if a variable is not σ-relevant, then, in each assignment that satisfies the formula, its value is $\neg\sigma$, that is, the variable is "$\neg\sigma$-critical."

We will first assume an algorithm `Relevant Variable-Assignment` that given an exact 3 CNF formula ϕ over variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ outputs an index i and a bit σ such that x_i is σ-relevant. The variables $y_1, \ldots, y_n$ are auxiliary variables that we add to the formula to ensure that the formula remains an exact 3 CNF formula. Algorithm `Relevant Variable-Assignment` returns an index of a variable x_i (it never returns a variable y_i). For technical reasons, `Relevant Variable-Assignment` needs that ϕ includes at least 3 variables from $x_1, \ldots, x_n$ (negated or non-negated). Algorithm `Greedy E3SAT`, described in Figure 10, uses algorithm `Relevant Variable-Assignment` to decide E3SAT.

Algorithm `Greedy E3SAT`

INPUT: An exact 3 CNF formula ϕ on n variables $x_1, \ldots, x_n$.
OUTPUT: 1 if ϕ is satisfiable, 0 otherwise.

1. Let $\hat{\phi} = \phi$.
2. While $\hat{\phi}$ contains at least three variables in $x_1, \ldots, x_n$ do:
   (a) Execute Algorithm `Relevant Variable-Assignment` on $\hat{\phi}$. Denote the answer $\langle i, \sigma \rangle$.
   (b) Assign $a_i \leftarrow \sigma$.
   (c) Modify $\hat{\phi}$ as follows:
      i. Leave in $\hat{\phi}$ every clause that does not include x_i or $\neg x_i$.
      ii. If $\sigma = 0$, then replace every clause $(x_i \vee \ell_j \vee \ell_k)$ in $\hat{\phi}$ by the clauses $(y_i \vee \ell_j \vee \ell_k)$ and $(\neg y_i \vee \ell_j \vee \ell_k)$. Remove all clauses that include $\neg x_i$.
      iii. If $\sigma = 1$, then replace every clause $(\neg x_i \vee \ell_j \vee \ell_k)$ in $\hat{\phi}$ by the clauses $(y_i \vee \ell_j \vee \ell_k)$ and $(\neg y_i \vee \ell_j \vee \ell_k)$. Remove all clauses that include x_i.
3. For every variable x_i that is not in $\hat{\phi}$ and was not assigned a value in step (2b), assign $a_i \leftarrow 0$.
4. Exhaustively check all 4 assignments a that agree with the assignments made in step (2b) and step (3). If at least one of these assignments satisfies ϕ return 1, otherwise return 0.

Figure 10: Algorithm `Greedy E3SAT`.

Note that the algorithm terminates in at most $n - 2$ iterations, as every iteration reduces the number of x variables in $\hat{\phi}$ by one. The final length of $\hat{\phi}$ is at most eight times that of ϕ as the replacement steps 2(c)ii and 2(c)iii are applied only to clauses including x variables, resulting in one less x variable in each of the two replacement clauses. Hence, if algorithm `Relevant Variable-Assignment` runs in polynomial time, so does `Greedy E3SAT`.

We now prove the correctness of algorithm `Greedy E3SAT`. Clearly, the only way `Greedy E3SAT` can err is by outputting 0 on a satisfiable formula ϕ, hence we assume from now on that ϕ is satisfiable, and show that executing `Greedy E3SAT` results in outputting 1. Before proving the correctness of the algorithm, we will try to give the intuition behind it. In each step of the algorithm we have a partial assignment to ϕ that can be extended to an assignment that satisfies ϕ. On one hand, this partial assignment already satisfies some clauses in ϕ and therefore we deleted them from $\hat{\phi}$. On the other hand, some literals in various clauses are not satisfied by the partial assignment. We would have liked to delete these literals from the clauses that they appear in, and continue. However, this might result in a 3 CNF formula that is not an exact 3 CNF formula. We, therefore, replace the literal ℓ_i that is not satisfied by an auxiliary variable y_i. By replacing each clause $(\ell_i \vee \ell_j \vee \ell_k)$ with two clauses $(y_i \vee \ell_j \vee \ell_k)$ and $(\neg(y_i) \vee \ell_j \vee \ell_k)$, one with y_i and one with $\neg y_i$, we ensure that a satisfying assignment to $\hat{\phi}$ must satisfy $\ell_j \vee \ell_k$.

The formal proof is by induction on the number of iterations in `Greedy E3SAT`: After executing the main iteration in algorithm `Greedy E3SAT` for k times (i) $\hat{\phi}$ is satisfiable; (ii) k variables are assigned values in Step 2b; (iii) $\hat{\phi}$ does not contain these variables; and (iv) any assignment that extends the k assigned variables satisfies ϕ iff it satisfies $\hat{\phi}$. $\square$

To conclude the proof, we now present algorithm `Relevant Variable-Assignment`. On input ϕ, the algorithm uses a private $1/n^{1-\epsilon}$-approximation algorithm $\mathcal{A}$ to produce an index i and a bit σ such that x_i is σ-relevant.

Algorithm `Relevant Variable-Assignment`

INPUT: An exact 3 CNF formula ϕ over Boolean variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. Without loss of generality, variables x_1, x_2, x_3 all appear in ϕ. Let m denote the number of clauses in ϕ.
OUTPUT: an index $i \in \{1, 2, 3\}$ and a bit σ such that x_i is σ-relevant.

1. Execute $\mathcal{A}$ on ϕ and denote $a = \mathcal{A}(\phi)$.
2. For $i \in \{1, 2, 3\}$, let $\ell_i = x_i$ if $a(x_i) = 0$ and $\ell_i = \neg x_i$ otherwise.
3. Set $\phi' = \phi \wedge (\ell_1 \vee \ell_2 \vee \ell_3) \wedge \cdots \wedge (\ell_1 \vee \ell_2 \vee \ell_3)$, where the clause $(\ell_1 \vee \ell_2 \vee \ell_3)$ is added $n^{1-\epsilon} \cdot (m+1)$ times.
4. Execute $\mathcal{A}$ on ϕ' and denote $a' = \mathcal{A}(\phi')$.
5. Choose $i \in \{1, 2, 3\}$ such that $a'(x_i) \neq a(x_i)$. Let $\sigma = a(x_i)$.
6. Return $\langle i, \sigma \rangle$.

By our choice of ℓ_1, ℓ_2, ℓ_3, the clause $(\ell_1 \vee \ell_2 \vee \ell_3)$ is not satisfied by the assignment a. Hence, as $\mathcal{A}$ is a $1/n^{1-\epsilon}$ approximation algorithm, at least one of x_1, x_2, x_3 changes assignments between a and a'. Hence, as $\mathcal{A}$ respects the privacy structure $\mathcal{R}_{\text{maxE3SAT}}$, we conclude that the formulae ϕ and ϕ' differ on their sets of maximum assignments. The following claim implies the correctness of algorithm `Relevant Variable-Assignment`:

Claim C.3 *If $a'(x_i) \neq a(x_i)$ for some $i \in \{1, 2, 3\}$, then x_i is $a(x_i)$-relevant.*

Proof: Assume the contrary, i.e. that every maximum assignment for ϕ assigns $a'(x_i)$ to x_i. It is easy to see that every maximum assignment of ϕ is also a maximum assignment for ϕ' as it satisfies the added clauses $(\ell_1 \vee \ell_2 \vee \ell_3)$. In the reverse direction, note that every maximum assignment for ϕ' is also maximum for ϕ as the assignment to the two other variables from x_1, x_2, x_3 does not affect the satisfiability of the clause $(\ell_1 \vee \ell_2 \vee \ell_3)$. We get that $\langle \phi, \phi' \rangle \in \mathcal{R}_{\text{maxE3SAT}}$, contradicting $\mathcal{A}(\phi) \neq \mathcal{A}(\phi')$. $\square$