On the Contradiction Graph of a Hypothesis Class

Jesse Campbell
University of Illinois – Chicago

Daniel Ibaibarriaga
University of Illinois – Chicago

Abstract

The contradiction graph was introduced by Alon et al. (2024), and was shown to characterize approximate and pure differentially private PAC learnability via the clique and fractional clique dimensions, respectively. We partially answer a question of Alon et al. by utilizing large cliques in the contradiction graph to construct a distribution over labeled examples over which it is difficult to learn privately. Furthermore, we introduce a weighted generalization of the contradiction graph which encodes the number of contradictions between datasets, and show that it characterizes PAC learnability via *balanced cliques* satisfying a certain normalizing edge weight condition. We also give a sufficient condition for efficient SQ learnability in-terms of cliques with large pairwise contradictions in the weighted contradiction graph. We leave as an open question whether this sufficient condition is necessary.

1 Introduction

Since the question of learning under differential privacy was raised in 2008 by Kasiviswanathan et al. [KLN$^+$11], it has become one of the most important and well-studied in learning theory. Recently in 2024, Alon et al. [AMSY24] introduced the *contradiction graph* of a hypothesis class, and remarkably showed that its graph theoretic properties characterize private PAC learnability. In this note, we continue the study of the contradiction graph, showing that it unifies several different notions of learnability beyond private PAC.

In **Section 2**, we state necessary definitions and lemmas that are key to our results.

In **Section 3**, we partially answer a question of Alon et al. [AMSY24], who asked for direct proofs between large (fractional) cliques in the contradiction graph and the hardness of private learnability. In particular, they asked for distributions over labeled examples over which are hard to privately learn given large cliques. First, **Theorem 3.1** directly gives an algorithm for PAC learning in the pure DP setting given finite fractional clique dimension, without a reduction to representation dimension. This result immediately implies a lower bound for fractional clique dimension in-terms of representation dimension. **Theorem 3.2** shows the converse direction in the approximate DP setting. That is, if a concept class is approximately privately PAC learnable, then clique dimension is finite. We achieve this result by explicitly constructing a distribution over examples over which it is difficult to learn privately.

In **Section 4**, we generalize the contradiction graph to a weighted graph, where the edge weights count the number of disagreements between two realizable datasets. Unlike the contradiction graph, where [AMSY24] conjecture PAC learnability is not able to be determined, **Theorem 4.1** proves that the weighted contradiction graph characterizes PAC learnability via the existence of large *balanced cliques*, where each vertex in the clique is adjacent to a fixed number of edges with a certain weight. Moreover, **Theorem 4.2** proves that if there is no large clique in the weighted contradiction graph where all the edge weights between vertices in the clique are large, then a concept class is efficiently SQ learnable. The proof of **Theorem 4.2** relies on the characterization of efficient SQ learnability by SQ dimension. Lastly, **Theorem 4.4** verifies that our characterizations of PAC and efficient SQ learnability via the weighted contradiction graph are consistent with the known relationship between the two learnability classes, namely, that SQ $\subset$ PAC. We leave it as an open question whether our sufficient condition for efficient SQ learnability is necessary.

Finally, **Section 5** states several open questions and interesting directions for future research.

1.1 Previous Work

The study of differentially private PAC was introduced in 2008 by Kasiviswanathan et al. [KLN$^+$11]. Since then, it has received extensive study, and remains an active area of research in learning theory today. The characterization of pure differentially private PAC learning was first studied by Beimel et al. in 2010, where they noted that proper private PAC learning is strictly harder than improper private PAC learning [BKN10]. Later in 2013, Beimel et al. introduced the representation dimension of a concept class, and showed that it characterizes when private learning is possible, as well as the sample complexity of private learning [BNS13].

By contrast, the class of approximately private PAC learnable concepts was shown by Feldman and Xiao in 2014 to be characterized by Littlestone dimension [FX14]. In 2013, another paper of Beimel et al. showed that the representation dimension does not characterize approximate private PAC learning, distinguishing pure and approximate DP PAC as two separate learnability classes. In 2015, Bun et al. gave the first lower bound for the sample complexity of approximately private PAC learners [BNSV15], which was extended into the improper case in 2018 by Alon et al. who showed approximately private PAC implies finite Littlestone dimension [ALMM19]. Later in 2022, Alon et al. proved that finite Littlestone dimension and approximately private PAC are equivalent [ABL$^+$22]. Moreover, Bun et al. gave a sample complexity upper bound for approximate private PAC learning in the improper case [BLM20b], which was improved and extended to the proper case by Ghazi et al. in 2021 [GGKM21].

The *contradiction graph* of a hypothesis class was introduced by Alon, Moran, Schefler, and Yehudayoff in 2024 [AMSY24]. They defined new combinatorial dimensions based on the cliques in the contradiction graph, and related them to the representation and Littlestone dimensions, thus giving an equivalence to private learnability.

2 Preliminaries

2.1 Learning Theory

First, we state the definition of *PAC learnability*, which is the most fundamental notion of learning. Roughly speaking, a concept class is PAC learnable if there is an algorithm which, with high probability, outputs a hypothesis which is within a small approximation factor of the target concept.

Definition 2.1 (PAC Learning, [Val84])**.** *A concept class $\mathcal{C}$ is PAC-learnable using hypothesis class $\mathcal{H}$ if there exists an algorithm $\mathcal{A} : \mathcal{D}^m \to \mathcal{H}$ and a polynomial poly$(\cdot,\cdot,\cdot,\cdot)$ such that for any $\alpha, \beta > 0$, for all distributions $\mathcal{D}$ on $\mathcal{X}$ and for any target concept $c \in \mathcal{C}$, the following holds for any sample size $m \geq$ poly$(1/\alpha, 1/\beta, n,$ size$(h))$:*

$$\Pr_{S \sim \mathcal{D}^m}[R(h_S) \leq \alpha] \geq 1 - \beta$$

where $h_S = \mathcal{A}(S)$.

We also make explicit the distinction between proper (or realizable) and improper (or unrealizable) learning.

Definition 2.2. *An algorithm satisfying* ***Definition 2.1*** *with $\mathcal{H} \subseteq \mathcal{C}$ is called a proper PAC learner; otherwise it is called an improper PAC learner.*

Before we continue, we address the ambiguity in the characterization of a concept class as *PAC-learnable*. In this note and related works regarding private PAC learning, a concept class is said to be *PAC-learnable* if *there exists* some hypothesis class $\mathcal{H}$ satisfying **Definition 2.1**. In general, it is not the case that $\mathcal{H} \subseteq \mathcal{C}$. It was first asked by Bousquet et al. [BLM20a] whether proper learning in the approximate DP setting was possible for classes with finite Littlestone dimension, and was answered positively by Ghazi et al. [GGKM21]. However, the question remains open in the pure DP setting whether finite representation dimension (or, equivalently, fractional clique dimension) implies a concept class can be properly PAC learned.

The *Littlestone dimension* of a concept class is defined as the maximum depth of a shattered binary tree, and is known to characterize approximate DP PAC learnability [ABL+22]. In particular, the following lemma characterizes the sample complexity of approximate DP PAC learning in the *improper case* in-terms of Littlestone dimension.

Lemma 2.1 ([BLM20b])**.** *Let $\mathcal{H} \subseteq \{\pm 1\}^X$ be a class with Littlestone dimension d, let $\varepsilon, \delta \in (0, 1)$ be privacy parameters, and let $\alpha, \beta \in (0, 1/2)$ be accuracy parameters. For*

$$n = O\left(\frac{2^{\tilde{O}(2^d)} + \log(1/(\beta\delta))}{\alpha\varepsilon}\right) = O_d\left(\frac{\log(1/(\beta\delta))}{\alpha\varepsilon}\right)$$

there exists an (ε, δ)-DP learning algorithm such that for every realizable distribution D, given an input sample $S \sim D^n$, the output hypothesis $f = A(S)$ satisfies $\mathrm{loss}_D(f) \leq \alpha$ with probability at least $1 - \beta$, where the probability is taken over $S \sim D^n$ as well as the internal randomness of A.

Finally, we establish an alternative notion of learning called *efficient statistical query learning*, or efficient SQ learning. An efficient SQ learning algorithm works by prompting an oracle for expectations of queries over a target dataset, and similar to PAC, aims to output a hypothesis which is a good approximation of the target concept.

Definition 2.3 (Efficient SQ leaning, [Kea98])**.** *Let $\mathcal{C}$ be a class of boolean functions $c : X \to \{-1, 1\}$. We say that $\mathcal{C}$ is efficiently SQ-learnable if there exists an algorithm A such that for every $c \in \mathcal{C}$, any probability distribution $\mathcal{D}$, and any $\epsilon > 0$, there is a polynomial $p(\cdot, \cdot, \cdot)$ such that:*

1. *A makes at most $p(1/\epsilon, n, |c|)$ calls to the SQ oracle,*
2. *the smallest τ that A uses satisfies $1/\tau \leq p(1/\epsilon, n, |c|)$,*
3. *the queries q are evaluable in time $p(1/\epsilon, n, |c|)$,*

and A outputs a hypothesis h satisfying $\mathrm{err}_{\mathcal{D}}(h) \leq \epsilon$.

SQ learnability is characterized by the SQ dimension of a concept class, which is defined as follows.

Definition 2.4 (Statistical query dimension, [BFJ+94])**.** *For a concept class C and distribution D, the statistical query dimension of C with respect to D, denoted $\text{SQ-DIM}_D(C)$, is the largest number d such that C contains d functions $f_1, f_2, \ldots, f_d$ such that for all $i \neq j$, $|\langle f_i, f_j \rangle_D| \leq 1/d$, where $\langle f_i, f_j \rangle_D = \mathbb{E}_D[f_i \cdot f_j]$. When we leave out the distribution D as a subscript, we refer to the statistical query dimension with respect to the worst-case distribution*

$$\text{SQ-DIM}(C) = \max_{D \in \mathcal{D}} \left(\text{SQ-DIM}_D(C)\right).$$

2.2 Differential Privacy

Differential privacy solved the problem of privacy, giving a mathematically rigorous notion of privacy with strong composition properties without requiring excessive noise to be added to data. Differential privacy is defined with respect to neighboring datasets which differ in the contribution of a single individual, and requires that the output of an algorithm is statistically indistinguishable over neighboring inputs.

Definition 2.5 (Differential Privacy, [DR14]). *Let $\mathcal{X}$ be a data universe and let $\mathcal{D} \subseteq \mathcal{X}^n$ be the set of databases of size n. Two databases $D, D' \in \mathcal{D}$ are called* neighbors *(written $D \sim D'$) if they differ in the data of at most one individual. A randomized algorithm (mechanism) $\mathcal{A} : \mathcal{D} \to \mathcal{Y}$ is said to satisfy* (ε, δ)-differential privacy *if for all neighboring databases $D \sim D'$ and for all measurable subsets $S \subseteq \mathcal{Y}$,*

$$\Pr[\mathcal{A}(D) \in S] \;\le\; e^{\varepsilon} \Pr[\mathcal{M}(D') \in S] + \delta.$$

When $\delta = 0$, we say that $\mathcal{M}$ satisfies pure *(or ε-) differential privacy and write simply ε-DP. Otherwise, we say the algorithm satisfies approximate differential privacy.*

There are several fundamental tools in the design of differentially private algorithms, among which is the exponential mechanism. Give a list of candidates, the exponential mechanism selects a candidate which is nearly optimal under pure differential privacy. In particular, the utility of the exponential mechanism degrades logarithmically in the number of candidates.

Lemma 2.2 (Exponential Mechanism, [MT07]). *Let $\mathcal{X}$ be a data universe and $\mathcal{D} \subseteq \mathcal{X}^n$ the set of databases of size n. Let $\mathcal{R}$ be a (finite or countable) range of possible outputs, and let $q : \mathcal{D} \times \mathcal{R} \to \mathbb{R}$ be a* quality score function. *The (global) sensitivity of q is defined as*

$$\Delta q \;:=\; \max_{D \sim D'} \max_{r \in \mathcal{R}} \bigl| q(D, r) - q(D', r) \bigr|,$$

where $D \sim D'$ denotes that D and D' are neighboring databases. For a privacy parameter $\varepsilon > 0$, the exponential mechanism $\mathcal{M}_q$ *outputs an element $r \in \mathcal{R}$ with probability*

$$\Pr[\mathcal{M}_q(D) = r] \;\propto\; \exp\!\left(\frac{\varepsilon\, q(D, r)}{2\Delta q} \right),$$

and satisfies ε-differential privacy. Moreover, for every $\beta \in (0, 1)$ and every database $D \in \mathcal{D}$, with probability at least $1 - \beta$ over the randomness of $\mathcal{M}_q(D)$, the exponential mechanism returns a function q^* satisfying,

$$q(D, \mathcal{M}_q(D)) \;\ge\; q^*(D) \;-\; \frac{2\Delta q}{\varepsilon} \Bigl(\log |\mathcal{R}| + \log (1/\beta) \Bigr)$$

One reason for the success of differential privacy is its strong composition properties. That is, the composition of several DP mechanisms maintains differential privacy, albeit with degrading privacy parameters. One fundamental composition result is advanced or adaptive composition, which is stated in the following lemma.

Lemma 2.3 (Advanced Composition, [DRV10]). *Let $\varepsilon, \delta \ge 0$ and let $\delta' \in (0, 1)$. For $k \in \mathbb{N}$, suppose $\mathcal{M}_1, \ldots, \mathcal{M}_k$ are (possibly adaptive) randomized mechanisms such that for each $i \in \{1, \ldots, k\}$, the mechanism $\mathcal{M}_i$ satisfies (ε, δ)-differential privacy. Then, the composed mechanism $\mathcal{M}(D) := (\mathcal{M}_1(D), ..., \mathcal{M}_k(D))$ satisfies (ε', δ'')-differential privacy with*

$$\varepsilon' = \sqrt{2k \ln\bigl(1/\delta'\bigr)}\varepsilon + k\varepsilon(e^{\varepsilon} - 1) \quad \textit{and} \quad \delta'' = k\delta + \delta'$$

2.3 Contradiction Graphs and Clique Dimensions

The contradiction graph was introduced by [AMSY24] as a graph encoding realizable datasets as vertices and contradictory relationships as edges.

Definition 2.6 (Contradiction Graph)**.** *Let $\mathcal{H} \subseteq \{0,1\}^X$ be a concept class and let $m \in \mathbb{N}$. The* contradiction graph *of order m of $\mathcal{H}$ is an undirected graph $G_m = G_m(\mathcal{H})$ whose vertices are datasets of size m that are consistent with $\mathcal{H}$. Two datasets are connected by an edge whenever they contradict each other. That is, $\forall (S,T) \in V^2$,*

$$(S,T) \in E \iff \exists (x,y) \in S \ s.t. \ (x,-y) \in T$$

Using the contradiction graph, [AMSY24] further introduced two new combinatorial dimensions associated with a concept class, based on the existence of large cliques and fractional cliques.

Definition 2.7 (Clique Dimension)**.** *The clique dimension of a concept class $\mathcal{H}$, denoted* $\mathrm{CD}(\mathcal{H})$*, is defined as follows:*

$$\mathrm{CD}(\mathcal{H}) := \sup\{m : \omega_m = 2^m\} \in \mathbb{N} \cup \{\infty\},$$

where ω_m is the clique number of the contradiction graph $G_m(\mathcal{H})$.

Definition 2.8 (Fractional Clique Dimension)**.** *The fractional clique dimension of a concept class $\mathcal{H}$, denoted* $\mathrm{CD}(\mathcal{H})$*, is defined as follows:*

$$\mathrm{CD}(\mathcal{H}) := \sup\{m : \omega_m^* = 2^m\} \in \mathbb{N} \cup \{\infty\},$$

where ω_m^* is the fractional clique number of the contradiction graph $G_m(\mathcal{H})$.

Another major result of [AMSY24] was to show that the clique and fractional clique dimensions satisfy a Sauer-Shelah-Perles-type exponential polynomial dichotomy.

Lemma 2.4 (SSP for (Fractional) Clique Dimension)**.** *Let $\mathcal{H}$ be a class and let ω_m denote the clique number of $G_m(\mathcal{H})$. Then, exactly one of the following statements holds:*

1. *$\omega_m = 2^m$ for all m.*
2. *$\omega_m \leq P(m)$ for all m, where $P(m)$ is a polynomial.*

Moreover, an identical statement holds for ω_m^*, the fractional clique number of $G_m(\mathcal{H})$.

Finally, they gave a probabilistic interpretation to the fractional clique number of $G_m(\mathcal{C})$. This probabilistic interpretation is due to the strong duality in contradiction graphs, and therefore the equivalence of fractional clique and fractional chromatic numbers.

Lemma 2.5. *Let ω_m^* denote the fractional clique number of $G_m(\mathcal{C})$. Then there exists a distribution $\mu^\star$ over hypotheses such that for every realizable dataset S of size m,*

$$\Pr_{h \sim \mu^\star}\big[\, h \text{ is consistent with } S \,\big] \geq \frac{1}{\omega_m^*}.$$

3 Direct Proofs

In this section, we give "direct proofs" that finite fractional clique dimension implies pure DP PAC learnability, and approximate DP PAC learnability implies approximate DP PAC learnability. [AMSY24] showed that clique dimension is finite if and only if Littlestone dimension is finite, and fractional clique dimension is finite if and only if representation dimension is finite, indirectly proving the equivalence between private PAC learnability and finite (fractional) clique dimension. Our proofs are direct in the sense that they do not reduce to showing finiteness of Littlestone or representation dimension.

More precisely, in the approximate DP case, we construct a distribution over examples for which private learning is hard if the contradiction graph of a concept class elicits a large clique, a question posed by [AMSY24]. In particular, we show that the uniform distribution over the set of examples in any vertex in a large clique is hard over which to learn privately.

3.1 Finite Fractional Clique Dimension Implies Private PAC Learnability

In this section, we prove that finite fractional clique dimension implies pure DP PAC learnability. Our proof is straightforward, and works as follows: given a sample and using the distribution from **Lemma 2.5**, we sample polynomially many (in $\mathrm{CD}^*(\mathcal{C})$) hypotheses, and argue that one is consistent with the sample with high probability. Then, we use the exponential mechanism to privately select a near-consistent hypothesis. Standard generalization bounds for consistent hypothesis algorithms complete the argument.

Theorem 3.1. *Let $\mathcal{C}$ be a concept class for which $\mathrm{CD}^*(\mathcal{C})$ is finite. Then, $\mathcal{C}$ is ε-differentially privately PAC-learnable using hypothesis class $\mathcal{H} = \{0,1\}^{\mathcal{X}}$.*

Proof. Let $\alpha = \frac{1}{\omega^*(G_m(\mathrm{CD}^*(\mathcal{C})))} - \frac{1}{2^{\mathrm{CD}^*(\mathcal{C})}}$ and $d = \log(1/\alpha)/\alpha^2$. Then, in particular, $\omega^*(G_m(\mathcal{C})) < m^d$ for all $m \in \mathbb{Z}^+$ (Theorem 3 in [AMSY24]). Let μ_m be a distribution over hypotheses from $\mathcal{H}$ such that for any realizable dataset S of size m,

$$\Pr_{h \sim \mu_m}[h \text{ is consistent with } S] \geq \frac{1}{m^d} \tag{1}$$

Then, we give the following ε-DP algorithm for PAC learning $\mathcal{C}$.

Algorithm 1: Algorithm for ε-DP PAC learning $\mathcal{H}$ when $\mathrm{CD}^*(\mathcal{H}) < \infty$

```
Input: $S = ((x_1, y_1), ..., (x_m, y_m)) \sim \mathcal{D}^m$
Output: $\hat{h} \in \mathcal{H}$
Sample $H \sim \mu_m^N$ where $N = m^d \log(3/\beta)$;
Compute $u(S,h) = \sum_{i=1}^m \mathbf{1}_{h(x_i) \neq y_i}$ for each $h \in H$;
return $\hat{h} \sim \mathcal{P}$, where $\mathcal{P}$ is the probability distribution over $H$ which selects $h$ with
  probability proportional to $\exp(-\varepsilon \cdot u(S,h)/2)$;
```

Privacy Analysis.

In **Algorithm 1**, sampling from μ is done independent of the input S. Hence, H is public. Moreover, the sensitivity of $u(S,h)$ is 1, therefore selecting h with probability proportional to,

$$\exp\left(-\frac{\varepsilon \cdot u(S,h)}{2}\right)$$

is ε-DP by **Lemma 2.2**.

Accuracy analysis.
First, we bound the probability that H contains a hypothesis that is consistent on S. By (1), the probability that some hypothesis in H is consistent with some dataset S is,

$$\Pr[\exists h \in H : h \text{ is consistent with } S] \geq 1 - \left(1 - \frac{1}{m^d}\right)^N \geq 1 - \beta/3 \tag{2}$$

Conditioning on the event that there is some $h \in H$ that is consistent with our sample, pick $\alpha, \beta > 0$. The standard generalization bound for finite hypothesis classes in the inconsistent case says that for any $\delta > 0$, with probability $1 - \beta/3$ the following inequality holds,

$$\forall h \in H,\ R(h) \leq \hat{R}(h) + \sqrt{\frac{\log N + \log(3/\beta)}{2m}} \tag{3}$$

We wish to find a lower bound for m such that,

$$\Pr[\forall h \in H,\ |R(h) - \hat{R}(h)| \leq \alpha/2] \geq 1 - \beta/3 \tag{4}$$

From (3), we have,

$$\sqrt{\frac{\log N + \log(3/\beta)}{2m}} \leq \frac{\alpha}{2} \implies m \geq O\left(\frac{1}{\alpha^2}(\log N + \log(1/\beta))\right)$$

Next, let $h^* \in H$ be consistent on S. The standard utility bound for the exponential mechanism guarantees,

$$\Pr\left[\hat{R}(\hat{h}) \leq \hat{R}(h^*) + \frac{2}{\varepsilon m}(\log N + \log(3/\beta))\right] \geq 1 - \beta/3 \tag{5}$$

Setting $2/(\varepsilon m)(\log N + \log(3/\beta)) = \alpha/2$ yields,

$$m \geq O\left(\frac{1}{\varepsilon\alpha}(\log N + \log(1/\beta)\right)$$

By a union bound over (2), (4) and (5), with probability at least $1 - \beta$,

$$R(\hat{h}) \leq \hat{R}(\hat{h}) + \frac{\alpha}{2} \leq \hat{R}(h^*) + \alpha = \alpha$$

In other words, our algorithm $\mathcal{A}$ is ε-DP and satisfies,

$$\Pr_{S \sim \mathcal{D}^m}[R(\mathcal{A}(S)) \leq \alpha] \geq 1 - \beta$$

therefore it is ε-privately PAC-learnable with sample complexity,

$$\begin{aligned} m &= \max\left\{O\left(\frac{1}{\alpha^2}(\log N + \log(1/\beta)\right), O\left(\frac{1}{\varepsilon\alpha}(\log N + \log(1/\beta)\right), \mathrm{CD}^*(\mathcal{C})\right\} \\ &= \max\left\{O\left(\left(\frac{1}{\alpha^2} + \frac{1}{\varepsilon\alpha}\right)(d\log m + \log(1/\beta)\right), \mathrm{CD}^*(\mathcal{C})\right\} \end{aligned}$$

which is polynomial in all the relevant parameters. In particular, we may solve for m yielding,

$$m = \max\left\{\widetilde{O}\left(\left(\frac{1}{\alpha^2} + \frac{1}{\varepsilon\alpha}\right)(d + \log(1/\beta)\right), \mathrm{CD}^*(\mathcal{C})\right\}$$

where $\widetilde{O}$ suppresses polylogarithmic factors. □

We end by noting that Beimel et al. [BNS13] showed that $\Theta(\textsc{RepDim}(\mathcal{C}))$ samples are necessary and sufficient for PAC learning in the ε-DP setting, where $\textsc{RepDim}(\cdot)$ is the representation dimension. Hence, our result immediately shows that $\mathrm{CD}^*(\mathcal{C}) = \Omega(\textsc{RepDim}(\mathcal{C}))$. Another interesting direction is to find an upper bound for $\mathrm{CD}^*(\mathcal{C})$ in-terms of $\textsc{RepDim}(\mathcal{C})$.

3.2 Approximately Private PAC Learning Implies Finite Clique Dimension

In this section, we prove that if a concept class is PAC learnable under approximate DP, then it has finite clique dimension. In particular, given the largest clique in $G_m(\mathcal{C})$, we consider the uniform distribution over samples in an arbitrary vertex in the clique. We show that the existence of an approximate DP learner over this distribution forces the number of vertices in the clique to be smaller than 2^m for large m. This partially answers a question of Alon et al. [AMSY24] who asked for hard distributions for private learners given large cliques in the contradiction graph. Our proof uses the sample complexity bound for approximate DP learners (**Lemma 2.1**) and advanced composition of DP mechanisms (**Lemma 2.3**).

Theorem 3.2. *Let $\mathcal{C}$ be a concept class which is (ε, δ)-differentially privately PAC learnable using hypothesis class $\{0,1\}^{\mathcal{X}}$. Then $CD(\mathcal{C})$ is finite.*

Proof. Let $\mathcal{A}$ be a (ε, δ)-DP PAC learner for $\mathcal{C}$. Fix some value of m and let $(S_1, S_2, ..., S_{N(m)})$ be the vertices of the largest clique in $G_m(\mathcal{C})$, ties broken arbitrarily. We wish to show that $N(m) < 2^m$ when m is large enough. For each $S_i = ((x_1^i, y_1^i), (x_2^i, y_2^i), ..., (x_m^i, y_m^i))$, let $H_i \subset \{0,1\}^{\mathcal{X}}$ be the set of concepts that agree with the labelings of S_i on its m examples. Moreover, let $\mathcal{D}_i$ be the uniform distribution over the x_j^i's. We may choose $\alpha = 1/(2m)$ such that, by the PAC learning guarantees, there is some $n = \mathrm{poly}(1/\varepsilon, 1/\delta, 1/\alpha, 1/\beta)$ for which,

$$\Pr_{S\sim\mathcal{D}_i^n}[R(\mathcal{A}(S)) \leq 1/(2m)] \geq 1 - \beta \tag{6}$$

Notice that for any hypothesis $h \in \mathcal{C}$,

$$R_{\mathcal{D}_i}(h) = \sum_{j=1}^{m} \Pr_{x\sim\mathcal{D}_i}[x = x_j^i] \cdot \mathbf{1}\{h(x) \neq y_j^i\}$$

Therefore, $R(\mathcal{A}(S)) \leq 1/(2m)$ if and only if h agrees with the labelings of S_i on all of it's m examples. That is,

$$R(\mathcal{A}(S)) \leq 1/(2m) \iff \mathcal{A}(S) \in H_i$$

Next, let $\mathbf{x}_i \sim \mathcal{D}_i^n$. Since $\#\mathbf{x}_i = n$, $\mathbf{x}_i$ and $\mathbf{x}_j$ can differ on at most n points for any $i, j \in [N(m)]$, that is, $\mathbf{x}_i$ and $\mathbf{x}_j$ are n-neighboring inputs to $\mathcal{A}$. By iteratively applying advanced composition of DP mechanisms n times (**Lemma 2.3**) and (6), for any $\delta' > 0$,

$$1 - \beta \leq \Pr[\mathcal{A}(\mathbf{x}_i) \in H_i] \leq (\exp(\varepsilon\sqrt{2n\log(1/\delta')}) + n\varepsilon(e^{\varepsilon} - 1))\Pr[\mathcal{A}(\mathbf{x}_j) \in H_i] + n\delta + \delta'$$

That is,

$$\Pr[\mathcal{A}(\mathbf{x}_j) \in H_i] \geq \frac{1 - \beta - n\delta - \delta'}{\exp(\varepsilon\sqrt{2n\log(1/\delta')}) + n\varepsilon(e^{\varepsilon} - 1)} \tag{7}$$

Moreover, since the S_i are the vertices of a clique in the contradiction graph, $H_i \cap H_j = \emptyset$ for all $1 \leq i < j \leq N(m)$. That is, there are no hypotheses that are consistent with any two vertices in a clique in $G_m(\mathcal{C})$. Hence, the events $(\mathcal{A}(\mathbf{x}) \in H_i)$ and $(\mathcal{A}(\mathbf{x}) \in H_j)$ are disjoint. By normalization of probability measures, it follows that,

$$\sum_{i=1}^{N(m)} \Pr[\mathcal{A}(\mathbf{x}_j) \in H_i] \leq 1 \tag{8}$$

Combining (7) and (8) yields,

$$N(m) \leq \frac{\exp(\varepsilon\sqrt{2n\log(1/\delta')}) + n\varepsilon(e^{\varepsilon} - 1)}{1 - \beta - n\delta - \delta'} \tag{9}$$

In particular, fixing $\varepsilon, \delta', \beta > 0$ as small constants, and choosing $\delta = 1/m^2$ we may bound the growth of $N(m)$ by considering only the exponential term in the numerator. Notice that our choice of δ combined with **Lemma 2.1** avoids a pathological case where $N(m) \leq 0$ for large enough m. That is,

$$N(m) = O(e^{\sqrt{n}}) \tag{10}$$

Recalling that $\alpha = 1/(2m)$, **Lemma 2.1** states,

$$n = O_d\left(\frac{\log(1/(\beta\delta))}{\alpha\varepsilon}\right) = O_{d,\varepsilon,\beta}(m\log m) \tag{11}$$

Combining (10) and (11) yields,

$$N(m) = O(e^{O(\sqrt{m\log m})})$$

therefore $N(m) < 2^m$ for large enough m. Hence, by definition of clique dimension, $\mathrm{CD}(\mathcal{C}) < \infty$. □

We note that the question of constructing distributions over which private learning is hard is still open in the case of large fractional cliques in $G_m(\mathcal{C})$. This remains an interesting research direction, and will further clarify our understanding of the contradiction graph, as well as the growth rate of the fractional clique number in contradiction graphs. Unlike in the case of clique dimensions, there is no canonical distribution, like the uniform distribution over a dataset, in the fractional case, making the question more difficult.

4 Expressivity of Weighted Contradiction Graphs

Alon et al. [AMSY24] raised the question of the *expressivity* of contradiction graphs. In particular, they asked whether the contradiction graph can characterize PAC learnability, conjecturing that it cannot. We introduce a weighted generalization of the contradiction graph, encoding not only which datasets are contradictory, but moreover by how much they contradict one another. With this added information, we show that the weighted contradiction graph is highly expressive, encoding not just private PAC learnability, but also PAC learnability and sufficient conditions for efficient SQ learnability. We begin this section by introducing key definitions, including the weighted contradiction graph and *balanced cliques*.

Definition 4.1. *The weighed contradiction graph $G_m^{(\omega)}(\mathcal{C}) = (V, E, \omega)$ of a concept class $\mathcal{C}$ is the weighted graph obtained from $G_m(\mathcal{C})$ by letting $\omega : E \to [m]$ be the number of contradctions between two realizable datasets on m examples. Formally, for an edge $(S, T) \in E$, we define,*

$$\omega((S,T)) = |\{x \in S, T : (x, y) \in S \text{ and } (x, -y) \in T\}|$$

Notice that the weighted contradiction graph is indeed a true generalization of the contradiction graph, as it can be obtained by disregarding the weight function ω. That is, any property of $\mathcal{C}$ that is characterized by the contradiction graph may also be characterized by the weighted contradiction graph.

Next, we introduce a special type of weighted graph called a *balanced clique*. We define balanced cliques towards the goal of showing that they characterize PAC learnability in the weighted contradiction graph.

Definition 4.2. *A balanced clique of size $n = 2^k$ for some $k \in \mathbb{Z}^+$ is a weighted complete graph $K_n^{(\omega)} = (V, E, \omega)$ such that, for every vertex $u \in V$, u is adjacent to precisely $\binom{k}{i}$ edges of weight i for $1 \le i \le k$. Formally,*

$$|\{v \in V : \omega(u, v) = i\}| = \binom{k}{i}$$

For example, consider the Hamming cube H_d which is obtained from the hypercube Q_d by labeling the vertices in such a way that the Hamming distance between two vertex labels is precisely the distance between those two vertices, where d is a power of 2. We may construct a balanced clique from H_d by completing the edge set of the graph and weighing the edges according to the Hamming distance between labels of pairs of vertices. Interestingly, a balanced clique is not unique up to isomorphism, as demonstrated by the following example.

Example 4.1. *The following two non-isomorphic graphs on $2^3 = 8$ vertices are both balanced cliques.*

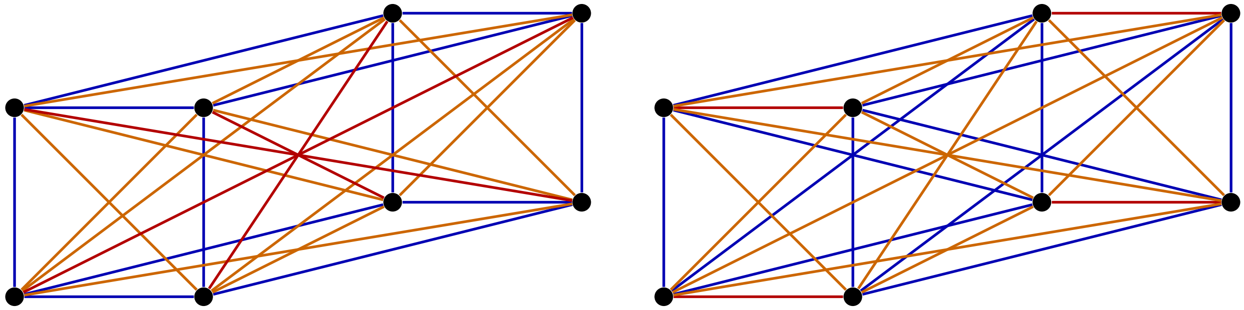

Figure 1: Let blue edges be of weight 1, green edges weight 2, and red edges weight 3. Then both graphs are balanced cliques. However, notice that while the subgraph induced by the blue edges in the left graph is bipartite, the subgraph induced by the blue edges in the right graph contains triangles, therefore they are non-isomorphic.

In **Appendix A**, we prove structural conditions on balanced cliques in weighted contradiction graphs. In particular, we prove a graph such as the right graph in **Example 4.1** cannot be realized as a balanced clique in a weighted contradiction graph as the subgraph induced by the edges of weight 1 has to be connected.

4.1 PAC Learnability and Balanced Cliques

In this section we show that a concept class is PAC learnable if and only if there are no maximal balanced cliques in the weighted contradiction graph for m large enough. In particular, we use the characterization of PAC learnability in-terms of finite VC dimension, which is defined as the maximum size of a set of examples for which all possible labelings are realizable.

Lemma 4.1 ([BEHW89])**.** *A concept class $\mathcal{C}$ is PAC learnable if and only if* $\mathrm{VC}\textsc{dim}(\mathcal{C}) < \infty$*.*

We are now ready to state our main result of this section. **Example 4.1** shows that the existence of a balanced clique does not immediately imply the existence of a shattered set, so a more precise argument is needed.

Theorem 4.1. *Let $\mathcal{C}$ be a concept class. Then $\mathrm{VC}\textsc{dim}(\mathcal{C}) \geq m$ if and only if $G_m^{(\omega)}(\mathcal{C})$ contains a balanced clique of size 2^m. In particular, $\mathcal{C}$ is PAC learnable if and only if there exists some $m_0 \in \mathbb{Z}^+$ such that $G_m^{(\omega)}(\mathcal{C})$ contains no balanced clique on 2^m vertices whenever $m > m_0$.*

Proof. Suppose $\mathrm{VC}\textsc{dim}(\mathcal{C}) = \infty$. That is, for any $m \in \mathbb{Z}^+$, there is a set of examples $X = (x_1, x_2, ..., x_m)$ such that all of the 2^m labels of X are realizable by $\mathcal{C}$. Let C_X be the set of 2^m realizable labelings of X in $V(G_m^{(\omega)}(\mathcal{C}))$. Clearly, $G_m^{(\omega)}[C_X]$ is a clique, since each labeling of X contradicts each other labeling. Moreover, it is not hard to check that $G_m^{(\omega)}[C_X]$ is a balanced clique.

We continue to prove the converse direction. Let C be a balanced clique in $G_m^{(\omega)}(\mathcal{C})$. We will show that every vertex in C is supported on the same set of examples, thus exhibiting a shattering of m points and proving the claim. To start, let $U = ((x_i, y_i)_{i=1}^m) \in C$ and let $S = \mathrm{supp}(U) = \{x_i\}_{i=1}^m$. Moreover, define $W_k = \{E \in C : \omega(U, E) = k\}$ for every $k \geq 1$. Notice that, by the assumption that C is balanced, $|W_k| = \binom{m}{k}$. We prove by descending induction on k that every vertex in C is supported on S. For the base case of $k = m$, notice that if $\omega(U, E) = m$, then U and E contradict on all m examples, therefore they must contain the same set of unlabeled examples. Next, suppose that the claim holds for all datasets in W_{k+1}, and take any dataset $E \in W_k$. In particular, since $|W_{k+1}| = \binom{m}{k+1}$ and each dataset in W_{k+1} is supported on S, we may associate each $(k+1)$-subset of S, call it Z to a unique dataset F_Z in W_{k+1} such that $Z = \{i : \ell_{F_Z}(x_i) \neq \ell_U(x_i)\}$, where $\ell_U(x)$ denotes the label of x in dataset U. Define $I(E) = S \cap \mathrm{supp}(E)$ and $A(E) = \{x \in I(E) : \ell_E(x) \neq \ell_U(x)\}$. By way of contradiction, suppose that $I(E) \neq S$ and pick $x \in S \backslash I(E)$. Then $B = A(E) \cup \{x\}$ is a $(k+1)$-subset of S, therefore there exists a unique $F_B \in W_{k+1}$ for which $A(F_B) = B$. Notice that the common examples between E and F_B are precisely,

$$\mathrm{supp}(E) \cap \mathrm{supp}(F_B) = \mathrm{supp}(E) \cap S = I(E)$$

by the induction hypothesis. For any $y \in I(E)$, if $y \in A(E)$, then $y \in B$, so both E and F_B disagree with U at y. In particular $\ell_E(y) = \ell_{F_B}(y)$. Likewise, if $y \notin A(E)$, then $y \notin B$, since $B = A(E) \cup \{x\}$, and $x \in S \backslash I(E)$. So, both E and F_B agree with U at y. Hence, once again, $\ell_E(y) = \ell_{F_B}(y)$. So, we have proven that, on all of the common examples between E and F_B, the labelings in E and F_B are the same. It follows that $\omega(E, F_B) = 0$, and E and F_B are non-contradictory datasets, contradicting that E and F_B are vertices of a complete graph in $G_m^{(\omega)}(\mathcal{C})$. □

An alternative proof of **Theorem 4.1** is provided in **Appendix A**.

Our result motivates the study of the contradiction graph, and in particular showing that it *cannot* capture PAC learnability, answering a question of Alon et al. [AMSY24] and thus proving that the weighted contradiction graph is a strictly stronger combinatorial object from the perspective of learning theory than the contradiction graph.

4.2 Efficient SQ Learnability in the Weighted Contradiction Graph

A concept class is not efficiently SQ learnable if the SQ dimension is high. That is, if there is a large set of concepts which pairwise have a low inner product, or are highly *uncorrelated*. Intuitively, this connects to the weighted contradiction graph in the following way: given a set of pairwise uncorrelated concepts, we may construct realizable datasets which pairwise have *many* contradictions by sampling an unlabeled set of examples and labeling them with the family of uncorrelated concepts. This corresponds to a clique in the weighted contradiction graph with large edge weights between each pair of vertices. **Theorem 4.2** is our main result of this section and formalizes this heuristic argument.

Theorem 4.2. *Let $\mathcal{C}$ be a concept class which is not efficiently SQ learnable. Then, for any $t \in \mathbb{Z}^+$ and $m = \Theta(t \log t)$, the weighted contradiction graph $G_m^{(\omega)}(\mathcal{C})$ contains a clique on t vertices in which all edge weights are $\Omega(t)$.*

Proof. Since $\mathcal{C}$ is not efficiently SQ learnable, $\text{SQ-DIM}(\mathcal{C}) > m$. That is, there exists a $\mathcal{D}$ such that we may find a set of t hypotheses $(h_i)_{i=1}^t$ for which $\langle h_i, h_j \rangle \leq 1/t$. Consider a set of unlabeled examples $X \sim \mathcal{D}^m$ drawn from the distribution $\mathcal{D}$, where we count the multiplicities of repeated examples. We first show that, with high probability, the empirical inner-product of two hypotheses on X is a good approximation of the true inner-product w.r.t. $\mathcal{D}$. In particular, define,

$$P(h_i, h_j) = \frac{1}{m} \sum_{x \in X} h_i(x) h_j(x)$$

as the empirical inner-product of h_i and h_j. Then, we immediately see that,

$$\mathbb{E}[P(h_i, h_j)] = \langle h_i, h_j \rangle$$

where the expectation is taken over the randomness of the sample X. Hence, by a Hoeffding bound, whenever $m = \Theta(1/\varepsilon^2 \log{(t^2/\delta)})$,

$$\Pr[|P(h_i, h_j) - \langle h_i, h_j \rangle| < \varepsilon] \geq 1 - \frac{\delta}{t^2} \tag{12}$$

By a union bound over all t^2 pairs, the probability that the condition $|P(h_i, h_j) - \langle h_i, h_j \rangle| < \varepsilon$ doesn't hold for some pair $i, j \in \binom{[n]}{2}$ is at most $1 - \delta = 1/4$. Next, we state the following technical lemma.

Lemma 4.2. *Let $\mathcal{D}$ be a distribution for which $\langle f_i, f_j \rangle_D \leq 1/d$ for a family $(f_i)_{i=1}^d$ of functions $f_i : \mathcal{X} \to \{-1, +1\}$. Then, for any $x \in \mathcal{X}$, $\Pr_{y \sim \mathcal{D}}[y = x] < 3/d$.*

Assuming **Lemma 4.2** holds, let $N_X(x)$ be the number of times that x appears in X. Let $N = \max_{x \in \mathcal{X}} N_X(x)$. Then,

$$\Pr[N \geq 100 \log t] \leq \binom{m}{k} \left(\frac{3}{t}\right)^{100 \log t - 1} \leq \frac{t}{3} \left(\frac{3e}{100}\right)^{100 \log t} = \frac{1}{3} t^{1 + 100 \log{(3e/100)}} < 1/4 \tag{13}$$

whenever $t \geq 2$. Moreover, notice that $m \cdot P(h_i, h_j)$ is precisely how many more examples $x \in X$ h_i and h_j agree upon than disagree. That is, let $D_X(h_i, h_j) = |\{x \in X : h_i(x) \neq h_j(x)\}|$, then (12) gives,

$$D_X(h_i, h_j) \geq \frac{m}{2} - m\left(\varepsilon - \frac{1}{t}\right)$$

Next, let S_i be a realizable dataset obtained from X by labeling the examples in X with the hypothesis h_i. Conditioning on the event that each example upon which h_i and h_j disagree occurs at most $100 \log t$ times in X, and choosing $\varepsilon = \frac{1}{\sqrt{t}}$,

$$D_{S_i}(h_i, h_j) \geq \frac{m}{(100 \log t)}\left(\frac{1}{2} - \frac{1}{\sqrt{t}}\right) = \Omega(t)$$

That is, we have constructed t datasets S_i which pairwise contradict on $\Omega(t)$ points. Note that if S_i does not contain m examples due to elements with multiplicities in X, we may arbitrarily add examples from $\mathcal{X}$ and label them with h_i such that S_i is a realizable dataset on m examples without decreasing the value of $D_{S_i}(h_i, h_j)$. Furthermore, if the conditioning in (12) or (13) fails (with probability at most 1/2), we may simply re-sample X until the conditions hold. □

*Proof of **Lemma 4.2**.* Fix any $x \in \mathcal{X}$ and let $\mathcal{D}(x) = \Pr_{y \sim \mathcal{D}}[y = x]$ be the probability mass under x in $\mathcal{D}$. By the pigeonhole principle, among the d values $f_1(x), ..., f_d(x) \in \{\pm 1\}$, at least $k = \lceil d/2 \rceil$ share the same sign. Call this subset P and without loss of generality, assume $f_i(x) = +1$ for all $i \in P$. Define $g = \sum_{i \in P} f_i$. Then $g(x) = k$, and hence $\mathbb{E}[g^2] \geq \mathcal{D}(x)k^2$. Moreover,

$$\mathbb{E}[g^2] = \sum_{i \in P} \mathbb{E}[f_i^2] + 2 \sum_{i < j \in P} \langle f_i, f_j \rangle < k + \frac{k(k-1)}{d}$$

Combining these two bounds for $\mathbb{E}[g^2]$,

$$\mathcal{D}(x) < \frac{k + k(k-1)/d}{k^2} < \frac{3}{d}$$

□

We move on to evaluate **Theorem 4.2** in the broader context of learning theory. That is, it is well established that SQ ⊂ PAC or, in other words, everything that is efficiently SQ learnable is also PAC learnable. As such, anything not PAC learnable is also not efficiently SQ learnable. In particular, if our results are correct, whenever we have a large balanced clique in $G_m^{(\omega)}(\mathcal{C})$, standard learning theory results demand that we also have a heavy clique as in the conclusion of **Theorem 4.2**. In what follows, we verify that this is indeed the case. We start by stating one of the most famous results in extremal graph theory: Turán's theorem.

Theorem 4.3 (Turán's Theorem)**.** *Among all K_{r+1}-free graphs on n vertices, the Turán graph $T_r(n)$ has the maximum number of edges. In particular,*

$$\mathrm{ex}(n, K_{r+1}) = e\big(T_r(n)\big) = \left(1 - \frac{1}{r}\right)\frac{n^2}{2} \quad \textit{when } r \mid n.$$

We also state a standard concentration inequality for Bernoulli random variables.

Lemma 4.3. *Let $X = \sum_{i=1}^n X_i$, where each X_i is an independent Bernoulli random variable with parameter p_i. Let $\mu = \mathbb{E}[X] = \sum_{i=1}^n p_i$. Then for every $0 < \delta < 1$,*

$$\Pr\big(X \leq (1-\delta)\mu\big) \;\leq\; \exp\!\left(-\frac{\mu\delta^2}{2}\right).$$

Now, we are ready to state our result that large balanced cliques imply heavy cliques, which in-turn imply a concept class is not efficiently SQ learnable via **Theorem 4.2**.

Theorem 4.4. *Let $\mathcal{C}$ be concept class which is not PAC-learnable. Then, for any $m \geq 60$, the weighted contradiction graph $G_m^{(\omega)}(\mathcal{C})$ contains a clique on m vertices where all edge weights are $\geq m/4$.*

Proof. Let $\mathcal{C}$ be concept class which is not PAC-learnable. By **Theorem 4.1** there is a balanced clique, say C, on $G_m^{(\omega)}(\mathcal{C})$ with 2^m vertices. Consider the subgraph T of C that contains all the 2^m vertices of C and only those edges with weight $\geq m/4$.

By the definition of a balanced clique, the total number of edges of T is,

$$2\,E(T) = \sum_{v \in V(T)} \sum_{i \geq \lfloor m/4 \rfloor} \binom{m}{i} = 2^m \sum_{i \geq \lfloor m/4 \rfloor} \binom{m}{i} = 2^{2m} \Pr(X \geq m/4)$$

where X is the sum of m i.i.d. Bernoulli random variables with parameter $1/2$. Therefore, by **Lemma 4.3**, since $\mathbb{E}[X] = m/2$, we take $\delta = 1/2$ to get that,

$$\Pr(X \leq m/4) \leq \exp\left(-\frac{(m/2)(1/2)^2}{2}\right) = \exp\left(-\frac{m}{16}\right).$$

In particular, we get that

$$E(T) \geq 2^{2m-1}\left(1 - \exp\left(-\frac{m}{16}\right)\right).$$

Now, using **Theorem 4.3** it is enough to prove that T has more edges than the Turan graph $T_m(2^m)$ to conlcude. In order to do so, observe that

$$\left(1 - \frac{1}{m}\right)\frac{(2^m)^2}{2} \leq E(T) \impliedby \left(1 - \frac{1}{m}\right)\frac{2^{2m}}{2} \leq 2^{2m-1}\left(1 - \exp\left(-\frac{m}{16}\right)\right) \iff 1 - \frac{1}{m} \leq 1 - \exp\left(-\frac{m}{16}\right)$$
$$\iff \frac{1}{m} \geq \exp\left(-\frac{m}{16}\right) \iff -\log(m) \geq -\frac{m}{16} \iff 16\log(m) \leq m$$

One can verify that the property $m \geq 60$ is sufficient to verify the last condition. □

5 Future Work

In **Section 3**, we showed how to construct a distribution over labeled examples which, given large cliques in the contradiction graph, is difficult over which to learn with approximate differential privacy. This motivates the same question for fractional cliques and pure differential privacy. Formally,

Question 5.1. *Given a hypothesis with infinite fractional clique dimension, how can we utilize the existence of a large fractional clique to construct a distribution over labeled examples which is difficult for PAC learning in the pure differential privacy setting?*

Furthermore, we explicitly give an upper bound for the sample complexity of pure DP PAC learning in the improper setting, which depends on the polynomial upper bound of the fractional clique number given finite fractional clique dimension. Hence, we restate the question of Alon et al. [AMSY24] to find a better polynomial bound for the fractional clique number.

Question 5.2. *Suppose $CD^*(\mathcal{C}) < \infty$ then is there a number d which depends only on $CD^*(\mathcal{C})$ for which $\omega^*(G_m(\mathcal{C})) < m^d$?*

In **Section 4**, we introduced the weighted contradiction graph, and showed that it characterizes PAC learnability. Towards showing that the weighted contradiction graph is a strictly stronger combinatorial object than the unweighted contradiction graph, we ask the following question.

Question 5.3. *Does there exist two concept classes* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *such that* $\text{VC}\textsc{dim}(\mathcal{C}_1) < \infty$ *and* $\text{VC}\textsc{dim}(\mathcal{C}_2) = \infty$*, but* $G_m(\mathcal{C}_1) \simeq G_m(\mathcal{C}_2)$ *for all* $m \in \mathbb{Z}^+$*?*

Lastly, we gave sufficient graph-theoretic conditions for efficient SQ learnability in the distribution independent setting. A natural question to ask is if this sufficient condition is necessary. That is, does the weighted contradiction graph characterize distribution independent efficient SQ learnability.

Question 5.4. *Given a concept class which, for every* $m = t \log t$*,* $G_m^{(\omega)}(\mathcal{C})$ *contains a clique of size* t *for which every edge has weight* $\Omega(t)$*, is* SQ-DIM $> m$*?*

Towards an answer to **Question 5.4**, for each dataset in the given clique, we may pick some concept that is consistent with the dataset, giving a family (h_i) of hypotheses. Then, it suffices to find some distribution $\mathcal{D}$ over unlabeled examples such that $\langle h_i, h_j \rangle_{\mathcal{D}} < 1/t$. We conjecture that **Question 5.4** or some variant of it is likely true, and will continue to search for a proof.

References

[ABL$^+$22] Noga Alon, Mark Bun, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private and online learnability are equivalent. *J. ACM*, 69(4), August 2022.

[ALMM19] Noga Alon, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private pac learning implies finite littlestone dimension. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 852–860, New York, NY, USA, 2019. Association for Computing Machinery.

[AMSY24] Noga Alon, Shay Moran, Hilla Schefler, and Amir Yehudayoff. A unified characterization of private learnability via graph theory. *Proceedings of Machine Learning Research*, 247:94–129, January 2024. Publisher Copyright: © 2024 N. Alon, S. Moran, H. Schefler & A. Yehudayoff.; 37th Annual Conference on Learning Theory, COLT 2024 ; Conference date: 30-06-2024 Through 03-07-2024.

[BEHW89] Anselm Blumer, A. Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the vapnik-chervonenkis dimension. *J. ACM*, 36(4):929–965, October 1989.

[BFJ$^+$94] Avrim Blum, Merrick Furst, Jeffrey Jackson, Michael Kearns, Yishay Mansour, and Steven Rudich. Weakly learning dnf and characterizing statistical query learning using fourier analysis. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '94, page 253–262, New York, NY, USA, 1994. Association for Computing Machinery.

[BKN10] Amos Beimel, Shiva Prasad Kasiviswanathan, and Kobbi Nissim. Bounds on the sample complexity for private learning and private data release. In *Proceedings of the 7th International Conference on Theory of Cryptography*, TCC'10, page 437–454, Berlin, Heidelberg, 2010. Springer-Verlag.

[BLM20a] Olivier Bousquet, Roi Livni, and Shay Moran. Synthetic data generators – sequential and private. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 7114–7124. Curran Associates, Inc., 2020.

[BLM20b] Mark Bun, Roi Livni, and Shay Moran. An Equivalence Between Private Classification and Online Prediction . In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 389–402, Los Alamitos, CA, USA, November 2020. IEEE Computer Society.

[BNS13] Amos Beimel, Kobbi Nissim, and Uri Stemmer. Characterizing the sample complexity of private learners. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS '13, page 97–110, New York, NY, USA, 2013. Association for Computing Machinery.

[BNSV15] Mark Bun, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. Differentially Private Release and Learning of Threshold Functions . In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 634–649, Los Alamitos, CA, USA, October 2015. IEEE Computer Society.

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407, August 2014.

[DRV10] Cynthia Dwork, Guy N. Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 51–60, 2010.

[FX14] Vitaly Feldman and David Xiao. Sample complexity bounds on differentially private learning via communication complexity. In Maria Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 1000–1019, Barcelona, Spain, 13–15 Jun 2014. PMLR.

[GGKM21] Badih Ghazi, Noah Golowich, Ravi Kumar, and Pasin Manurangsi. Sample-efficient proper pac learning with approximate differential privacy. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 183–196, New York, NY, USA, 2021. Association for Computing Machinery.

[Kea98] Michael Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, November 1998.

[KLN$^+$11] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

[MT07] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '07, page 94–103, USA, 2007. IEEE Computer Society.

[Val84] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, November 1984.

A Appendix A: Additional Proof of Theorem 4.1

The proof provided in **Section 4** of **Theorem 4.1**, although sufficient, somewhat undermines the geometric graph-theoretic perspective that we aim to develop. For this reason, we present here a new viewpoint on the theorem, which not only offers an alternative proof but also highlights some of the interesting graph-theoretic properties that balanced cliques possess.

Proposition A.1. *Let $\mathcal{C}$ be a concept class and let T be a balanced clique on $G_m^{(\omega)}(\mathcal{C})$. Then, any datasets $S, S' \in V(T)$ connected by a weight $m-1$ edge have the same example set.*

Proof. Let S and S' be datasets in $V(T)$ connected by an edge of weight $m-1$. By definition, this means that they disagree on $m-1$ examples and may differ on the remaining example at, say, the i-th position.

By the properties of a balanced clique, there exists a realizable sample $\bar{S} \in V(T)$ such that the edge between S and $\bar{S}$ has weight m. This implies that $\bar{S}$ has the same example set as S, but with all labels flipped. Since T is a clique, there must exist an edge between S' and $\bar{S}$, so they must disagree on some example. However, by the previous discussion, S' and $\bar{S}$ agree on every example except possibly the i-th one. Therefore, they must disagree on that example. In particular, they share the same example set, which proves the claim, since the example set of $\bar{S}$ is the same as that of S. □

Consider any balanced clique C, and consider the subgraph formed by the edges of weight $m-1$. The previous proposition shows that all such vertices share the same set of examples. Indeed, the hypothesis of **Proposition A.1** could be relaxed to require only that each vertex has m neighbours of weight $m-1$ and exactly one neighbour of weight m.

Moreover, the same result shows something further. Fix a data set S' in C and consider a data set S connected to S' by an edge of weight $m-1$. The proof of the proposition shows that the edge connecting S' with $\bar{S}$ has weight 1, and that S' and $\bar{S}$ share the same set of examples. By applying this argument to each of the m data sets connected to S' by edges of weight $m-1$, we obtain that these constitute all of the weight-1 edges incident to S' in the balanced graph.

These observations can be summarized in the following theorem.

Theorem A.1. *Let $\mathcal{C}$ be a concept class and let T be a clique on $G_m^{(\omega)}(\mathcal{C})$ of size 2^m. Let W_k be the subgraph obtained from T by including only edges of weight k. Assume W_{m-1} is m-regular and the W_m is a perfect matching on $V(T)$. Then, every connected component in W_{m-1} is supported on the same set of examples.*

Moreover, if the W_1 is m-regular, then every connected component is also supported on the same set of examples.

From here, we describe an inductive process that leads to the following stronger version of **Theorem 4.2**.

Theorem A.2. *Let $\mathcal{C}$ be a concept class and let T be a clique on $G_m^{(\omega)}(\mathcal{C})$ of size 2^m. Assume that W_1 and W_{m-1} are m-regular and that W_m is a perfect matching on $V(T)$. Then, every vertex in T is supported on the same set of examples.*

Proof. Let S be any dataset in $V(T)$ and assume, without loss of generality, that all its labels are 0. If no such data sample exists, then pick any data sample and flip the label convention on every example labeled 1. Our aim is to show that for any $\delta \in \{0,1\}^m$ there exists a dataset in T with the same examples as S and with label vector δ. Since there are 2^m such δ, this will prove the theorem.

By **Theorem A.1**, the neighbourhood of S via edges of weight 1 consists precisely of the m samples obtained by flipping exactly one label among the m examples. As there are exactly m labelings that differ from the "all-zero" labeling in exactly one position, these neighbours correspond to the labelings with all zeros except for a single 1. Let i be the first position where δ has a 1, and let S_1 be the neighbour of S whose label has a 1 in position i and zeros elsewhere.

We now repeat the same procedure with S_1 to obtain a sample S_2 having the same examples and whose label coincides with δ up to the second position where δ has a 1. Iterating this construction at most m times guarantees that we find a data sample in the clique with the same examples and exactly the label vector δ.

The only potential concern is the possibility that all edges of weight 1 from the already explored part of T lead to data samples whose labels we have already seen. However, since we know the labels of all such samples, this situation would imply that we have already reached the data sample with label vector δ. Thus the procedure must succeed. □

This searching procedure actually proves that

Corollary A.1. *Let $\mathcal{C}$ be a concept class and let T be a clique on $G_m^{(\omega)}(\mathcal{C})$ of size 2^m with the properties of the theorem. Then W_1 and W_{m-1} are connected subgraphs of T.*

We have done the construction like usign W_1 because we believe it is easier to understand for the reader. However, by doing the construction with $m-1$ neighborhoods we can drop the assumption that W_1 is m-regular. We also immediately arrive at the following corollary.

Corollary A.2. *Any clique in $G_m^{(\omega)}(\mathcal{C})$ such that W_{m-1} is m-regular and such that W_m is a perfect matching on $V(T)$ is a balanced clique.*