Proceedings of Machine Learning Research vol 98:1–23, 2019 30th International Conference on Algorithmic Learning Theory

Optimal Collusion-Free Teaching

David Kirkpatrick KIRK@CS.UBC.CA
Department of Computer Science, University of British Columbia

Hans U. Simon HANS.SIMON@RUB.DE
Department of Mathematics, Ruhr-University Bochum

Sandra Zilles ZILLES@CS.UREGINA.CA
Department of Computer Science, University of Regina

Editors: Aurélien Garivier and Satyen Kale

Abstract

Formal models of learning from teachers need to respect certain criteria to avoid collusion. The most commonly accepted notion of collusion-freeness was proposed by Goldman and Mathias (1996), and various teaching models obeying their criterion have been studied. For each model M and each concept class $\mathcal{C}$, a parameter $M\text{-TD}(\mathcal{C})$ refers to the *teaching dimension* of concept class $\mathcal{C}$ in model M—defined to be the number of examples required for teaching a concept, in the worst case over all concepts in $\mathcal{C}$.

This paper introduces a new model of teaching, called no-clash teaching, together with the corresponding parameter $\text{NCTD}(\mathcal{C})$. No-clash teaching is provably optimal in the strong sense that, given *any* concept class $\mathcal{C}$ and *any* model M obeying Goldman and Mathias's collusion-freeness criterion, one obtains $\text{NCTD}(\mathcal{C}) \leq M\text{-TD}(\mathcal{C})$. We also study a corresponding notion NCTD^+ for the case of learning from positive data only, establish useful bounds on NCTD and NCTD^+, and discuss relations of these parameters to the VC-dimension and to sample compression.

In addition to formulating an optimal model of collusion-free teaching, our main results are on the computational complexity of deciding whether $\text{NCTD}^+(\mathcal{C}) = k$ (or $\text{NCTD}(\mathcal{C}) = k$) for given $\mathcal{C}$ and k. We show some such decision problems to be equivalent to the existence question for certain constrained matchings in bipartite graphs. Our NP-hardness results for the latter are of independent interest in the study of constrained graph matchings.

Keywords: machine teaching, constrained graph matchings, sample compression

1. Introduction

Models of machine learning from carefully chosen examples, i.e., from teachers, have gained increased interest in recent years, due to various application areas, such as robotics (Argall et al., 2009), trustworthy AI (Zhu et al., 2018), and pedagogy (Shafto et al., 2014). Machine teaching is also related to inverse reinforcement learning (Ho et al., 2016), to sample compression (Moran et al., 2015; Doliwa et al., 2014), and to curriculum learning (Bengio et al., 2009). The paper at hand is concerned with abstract notions of teaching, as studied in computational learning theory.

A variety of formal models of teaching have been proposed in the literature, for example, the classical teaching dimension model (Goldman and Kearns, 1995), the optimal teacher model (Balbach, 2008), recursive teaching (Zilles et al., 2011), or preference-based teaching (Gao et al., 2017).

In each of these models, a mapping T (the *teacher*) assigns a finite set $T(C)$ of correctly labelled examples to a concept C in a concept class $\mathcal{C}$ in a way that the learner can reconstruct C from $T(C)$.

Intuitively, unfair collusion between the teacher and the learner should not be allowed in any formal model of teaching. For example, one would not want the teacher and learner to agree on a total order over the domain and a total order over the concept class and then to simply use the ith instance in the domain for teaching the ith concept, irrespective of the actual structure of the concept class.

However, there is no general definition of what constitutes collusion, and of what constitutes desirable or undesirable forms of learning. In this manuscript, we focus on a notion of collusion that was proposed by Goldman and Mathias (1996) and that has been adopted by the majority of teaching models studied in the literature. In a nutshell, Goldman and Mathias's model demands that, (i) the examples in $T(C)$ are labelled consistently with C, and (ii) if the learner correctly identifies C from $T(C)$, then it will also identify C from any superset S of $T(C)$ as long as the sample set S remains consistent with C. In other words, adding more information about C to $T(C)$ will not divert the learner to an incorrect hypothesis.

Most existing abstract models of machine teaching are collusion-free in this sense. Historically, some of these models were designed in order to overcome weaknesses of the previous models. For example, the optimal teacher model by Balbach (2008) is designed to overcome limitations of the classical teaching dimension model, and was likewise superseded by the recursive teaching model (Zilles et al., 2011). The latter again was inapplicable to many interesting infinite concept classes, which gave rise to the model of preference-based teaching (Gao et al., 2017). Each model strictly dominates the previous one in terms of the *teaching complexity*, i.e., the worst-case number of examples needed for teaching a concept in the underlying concept class $\mathcal{C}$. In this context, one quite natural question has been ignored in the literature to date: what is the smallest teaching complexity that can be achieved under Goldman and Mathias's condition of collusion-freeness? This is exactly the question addressed in this paper.

Our first contribution is the formal definition of a collusion-free teaching model that has, for every concept class $\mathcal{C}$, the provably smallest teaching complexity among all collusion-free teaching models. We call this model *no-clash teaching*, since its core property, which turns out to be characteristic for collusion-freeness, requires that no pair of concepts are consistent with the union of their teaching sets. A similar property was used once in the literature in the context of sample compression schemes (Kuzmin and Warmuth, 2007), and dubbed the *non-clashing* property.

For example, consider a concept class (i.e., set system) $\mathcal{C}$ over the instance space $\{1,2,3,4\}$, consisting of the four concepts of the form $\{i,(i+1) \bmod 4\}$ for $1 \leq i \leq 4$. Then no-clash teaching is possible by assigning the singleton set $\{(i,1)\}$ (interpreted as the information "i belongs to the target concept") as a teaching set to the concept $\{i,(i+1) \bmod 4\}$; no two distinct concepts are consistent with the union of their assigned teaching sets. Thus, in the no-clash setting, each concept in $\mathcal{C}$ can be taught with a single example. By comparison, consider the classical teaching dimension model, in which a teaching set for a given concept is required to be inconsistent with all other concepts in the concept class (Goldman and Kearns, 1995). It is not hard to see that, under such constraints, no concept in $\mathcal{C}$ can be taught with a single example; a smallest teaching set for concept $\{i,(i+1) \bmod 4\}$ would then be $\{(i,1),((i+1) \bmod 4,1)\}$.

We call the worst-case number of examples needed for non-clashing teaching of any concept C in a given concept class $\mathcal{C}$ the *no-clash teaching dimension* of $\mathcal{C}$, abbreviated $\mathrm{NCTD}(\mathcal{C})$, and we study a variant $\mathrm{NCTD}^+(\mathcal{C})$ in which teaching uses only positive examples. In the example above, $\mathrm{NCTD} = \mathrm{NCTD}^+ = 1$, while the classical teaching dimension is 2.

The value $\mathrm{NCTD}(\mathcal{C})$ being the smallest collusion-free teaching complexity parameter of $\mathcal{C}$ makes it interesting for several reasons.

(1) NCTD represents the limit of data efficiency in teaching when obeying Goldman and Mathias's notion of collusion-freeness. Therefore the study of NCTD has the potential to further our understanding how collusion-freeness constrains teaching. It will also help to compare other notions of collusion-freeness (see, e.g., (Zilles et al., 2011)) to that of Goldman and Mathias.

(2) An open question in computational learning theory is whether the VC-dimension (VCD), (Vapnik and Chervonenkis, 1971), which characterizes the sample complexity of learning from randomly chosen examples, also characterizes teaching complexity for some reasonable notion of teaching. Recently, the first strong connections between teaching and VCD were established, culminating in an upper bound on the recursive teaching dimension (RTD) that is quadratic in VCD (Hu et al., 2017), but it remains open whether this bound can be improved to be linear in VCD. Obviously, now NCTD is a much stronger candidate for a linear relationship with VCD than RTD is. In fact, there is no concept class known yet for which NCTD exceeds VCD.

(3) The problem of relating teaching complexity to VCD is connected to the famous open problem of determining whether VCD is an upper bound on the size of the smallest possible sample compression scheme (Littlestone and Warmuth, 1986; Floyd and Warmuth, 1995) of a concept class. Some interesting relations between sample compression and teaching have been established for RTD (Moran et al., 2015; Doliwa et al., 2014; Darnstädt et al., 2016). The study of NCTD can potentially strengthen such relations.

In addition, an important contribution of our paper is to link NCTD to the extensively developed theory of constrained graph matching. We show that the question whether $\mathrm{NCTD}^+ = 1$ is equivalent to a very natural constrained bipartite matching problem which has apparently not yet been studied in the literature. We proceed by proving that this particular matching problem is NP-hard—a result that generalizes to larger values of NCTD^+ as well as to NCTD. By comparison, the question whether $\mathrm{RTD}^+ = 1$ or $\mathrm{RTD} = 1$ can be answered in linear time; see the expanded version of this paper (Kirkpatrick et al., 2019) for details.

To sum up, our new notion of optimal collusion-free teaching is of relevance to the study of important open problems in computational learning theory as well as of fundamental graph-theoretic decision problems, and therefore appears to be worth studying in more detail.

2. Preliminaries

Given a domain $\mathcal{X}$, a concept over $\mathcal{X}$ is a subset $C \subseteq \mathcal{X}$, and we usually denote by $\mathcal{C}$ a *concept class* over $\mathcal{X}$, i.e., a set of concepts over $\mathcal{X}$. Implicitly, we identify a concept C over $\mathcal{X}$ with a mapping $C : \mathcal{X} \to \{0, 1\}$, where $C(x) = 1$ iff $x \in C$. By $\mathrm{VCD}(\mathcal{C})$, we denote the VC-dimension of $\mathcal{C}$.

A labelled example is a pair $(x, \ell) \in \mathcal{X} \times \{0, 1\}$, and it is consistent with a concept C if $C(x) = \ell$. Likewise, a set S of labelled examples over $\mathcal{X}$, which is also called a *sample set*, is consistent with C, if every element of S is consistent with C. An example with the label $\ell = 1$ is a positive example, while $\ell = 0$ is the label of a negative example.

Intuitively, the notion of teaching refers to compressing any concept in a given concept class to a consistent sample set.

Definition 1 *Let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$. A* teacher mapping *for $\mathcal{C}$ is a mapping T on $\mathcal{C}$ such that, for all $C \in \mathcal{C}$, $T(C)$ is a finite sample set $S \subseteq \mathcal{X} \times \{0, 1\}$ that is consistent with C.*

The first model of teaching that was proposed in the literature required from a teacher mapping T that the concept $C \in \mathcal{C}$ be the only concept in $\mathcal{C}$ that is consistent with $T(C)$, for any $C \in \mathcal{C}$

(Shinohara and Miyano, 1991; Goldman and Kearns, 1995). This led to the definition of the well-known teaching dimension parameter.

Definition 2 (Shinohara and Miyano (1991); Goldman and Kearns (1995)) *Let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$ and $C \in \mathcal{C}$ be a concept. A* teaching set *for C (with respect to $\mathcal{C}$) is a sample set S such that C is the only concept in $\mathcal{C}$ consistent with S. The* teaching dimension *of C in $\mathcal{C}$, denoted by $\mathrm{TD}(C, \mathcal{C})$, is the size of the smallest teaching set for C with respect to $\mathcal{C}$. The teaching dimension of $\mathcal{C}$ is then defined as $\mathrm{TD}(\mathcal{C}) = \sup\{\mathrm{TD}(C, \mathcal{C}) \mid C \in \mathcal{C}\}$.*

For example, let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$ of exactly m elements, containing the empty concept, all singleton concepts over $\mathcal{X}$, and no other concepts. Then $\mathrm{TD}(\{x\}, \mathcal{C}) = 1$ for each singleton concept $\{x\}$, since $\{(x, 1)\}$ serves as a teaching set for $\{x\}$. By comparison, $\mathrm{TD}(\emptyset, \mathcal{C}) = m$, since any set of up to $m - 1$ negative examples is consistent with some singleton concept, so that all m negative examples need to be presented in order to identify the empty concept. Consequently, $\mathrm{TD}(\mathcal{C}) = \sup\{\mathrm{TD}(C, \mathcal{C}) \mid C \in \mathcal{C}\} = m$.

As mentioned in the introduction, various notions of teaching have been proposed in the literature. The one that is most relevant to our work is the model of preference-based teaching. In this model, intuitively, a preference relation on $\mathcal{C}$ is used to reduce the size of teaching sets. In particular, a concept C need no longer be the only concept consistent with its teaching set $T(C)$; it suffices if C is the unique most preferred concept in $\mathcal{C}$ that is consistent with $\mathcal{C}$. In order to avoid cyclic preferences, the preference relation is required to form a partial order over $\mathcal{C}$.

Definition 3 (Gao et al. (2017)) *Let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$ and $\succ$ any binary relation that forms a strict (possibly non-total) order over $\mathcal{C}$. We say that concept C is preferred over concept C' (with respect to $\succ$), if $C \succ C'$. The* preference-based teaching dimension *of C with respect to $\mathcal{C}$ and $\succ$, denoted by $\mathrm{PBTD}(C, \mathcal{C}, \succ)$, is the size of the smallest sample set S such that*

1. *S is consistent with C, and*
2. *$C \succ C'$ for all $C' \in \mathcal{C} \setminus \{C\}$ such that S is consistent with C'.*

We write $\mathrm{PBTD}(\mathcal{C}, \succ) = \sup\{\mathrm{PBTD}(C, \mathcal{C}, \succ) \mid C \in \mathcal{C}\}$. Finally, the preference-based teaching dimension of $\mathcal{C}$, denoted by $\mathrm{PBTD}(\mathcal{C})$, is defined by

$$\mathrm{PBTD}(\mathcal{C}) = \min\{\mathrm{PBTD}(\mathcal{C}, \succ) \mid \succ \subseteq \mathcal{C} \times \mathcal{C} \text{ and } \succ \text{ forms a strict order on } \mathcal{C}\}\,.$$

An interesting variant of preference-based teaching is obtained when disallowing negative examples in teaching. Learning from positive examples only has been studied extensively in the computational learning theory literature, see, e.g., (Denis, 2001; Angluin, 1980) and is motivated by studies on language acquisition (Wexler and Culicover, 1980) or, more recently, by problems of learning user preferences from a user's interactions with, say, an e-commerce system (Schwab et al., 2000), as well as by problems in bioinformatics (Wang et al., 2006).

Definition 4 (Gao et al. (2017)) *Let $\mathcal{C}$ be a concept class over a domain $\mathcal{X}$. The* positive preference-based teaching dimension *of $\mathcal{C}$, denoted by $\mathrm{PBTD}^+(\mathcal{C})$, is defined analogously to $\mathrm{PBTD}(\mathcal{C})$, where the sets S in Definition 3 are required to contain only positive examples.*

In the same way, one can define the notion TD^+. The following property, proven by Gao et al. (2017), is crucial when computing the PBTD and PBTD^+ of finite classes.

Proposition 5 (Gao et al. (2017)) *Let $\mathcal{C}$ be a finite concept class. If $\mathrm{PBTD}(\mathcal{C}) = d$, then $\mathcal{C}$ contains some C with $\mathrm{TD}(C, \mathcal{C}) \leq d$. If $\mathrm{PBTD}^+(\mathcal{C}) = d$, then $\mathcal{C}$ contains some C with $\mathrm{TD}^+(C, \mathcal{C}) \leq d$.*

This result immediately implies that PBTD and the well-known notion of RTD[1] coincide for finite concepts classes, and so do PBTD^+ and RTD^+.

3. Collusion-free Teaching and the Non-Clashing Property

While there is no objective measure of how "reasonable" a formal model of teaching is, the literature offers some notions of what constitutes an "acceptable" model of teaching, i.e., one in which the teacher and learner do not collude. So far, the notion of collusion-free teaching that found the most positive resonance in the literature is the one defined by Goldman and Mathias.

Definition 6 (Goldman and Mathias (1996)) *Let $\mathcal{C}$ be a concept class over $\mathcal{X}$ and T a teacher mapping on $\mathcal{C}$. Let L be a learner mapping that assigns to each set of labelled examples a concept over $\mathcal{X}$. The pair (T, L) is successful on $\mathcal{C}$ if $L(T(C)) = C$ for all $C \in \mathcal{C}$. The pair (T, L) is* collusion-free *on $\mathcal{C}$ if $L(S) = L(T(C))$ for any $C \in \mathcal{C}$ and any set S of labelled examples such that S is consistent with C and S contains $T(C)$.*

Intuitively, Goldman and Mathias's definition captures the idea that a learner conjecturing concept C will not change its mind when given additional information consistent with C.

For example, teacher-learner pairs following the classical teaching dimension model, Balbach's optimal teacher model, the recursive teaching model, or the preference-based teaching model are always collusion-free according to Definition 6. Of these models, the classical teaching dimension model is the one imposing the most constraints on the mapping T, followed by Balbach's optimal teaching, recursive teaching, and preference-based teaching in that order. Consequently, the "teaching complexity" among these models is lowest for preference-based teaching; if every concept in a concept class $\mathcal{C}$ can be taught with at most z examples in any of these models, then every concept in $\mathcal{C}$ can be taught with at most z examples in the preference-based model.

One can still argue that the preference-based model is unnecessarily constraining. Preference-based teaching of a concept class $\mathcal{C}$ relies on a preference relation that induces a strict order on $\mathcal{C}$. However, this strict order is used by the learner only after the teaching set has been communicated, since the learner chooses the unique most preferred concept among those *consistent with the set of examples provided by the teacher*. One might consider loosening the constraints by, for example, demanding only that the set of concepts consistent with any chosen teaching set be ordered under the chosen preference relation (rather than requiring acyclic preferences over the whole concept class). In the same vein, one could relax more conditions—every relaxation might result in a more powerful model of teaching satisfying the collusion-free property.

In this manuscript, we will define the provably most powerful model of teaching that is collusion-free in the sense proposed by Goldman and Mathias (1996), namely a model that adheres to no other constraints on the teacher-learner pairs (T, L) than those given by Goldman and Mathias: (i) T is a teacher mapping; (ii) (T, L) is successful on $\mathcal{C}$; and (iii) (T, L) is collusion-free on $\mathcal{C}$.

1. The RTD, short for "recursive teaching dimension, is a well-studied teaching parameter defined by Zilles et al. (2011).

Before we define this model formally, we introduce a crucial property that was originally proposed by Kuzmin and Warmuth (2007) in the context of unlabeled sample compression.

Definition 7 *Let $\mathcal{C}$ be a concept class and T be a teacher mapping on $\mathcal{C}$. We say that T is* non-clashing *(on $\mathcal{C}$) if and only if there are no two distinct $C, C' \in \mathcal{C}$ such that both $T(C)$ is consistent with C' and $T(C')$ is consistent with C.*

It turns out that, for a teacher mapping T, the non-clashing property is equivalent to the existence of a learner mapping L such that (T, L) is successful and collusion-free:

Theorem 8 *Let $\mathcal{C}$ be a concept class over the instance space $\mathcal{X}$. Let T be a teacher mapping on $\mathcal{C}$. Then the following two conditions are equivalent:*

1. *T is non-clashing on $\mathcal{C}$.*
2. *There is a mapping $L : 2^{\mathcal{X}\times\{0,1\}} \to \mathcal{C}$ such that (T, L) is successful and collusion-free on $\mathcal{C}$.*

Proof First, suppose T is a non-clashing teacher mapping, and define L as follows. Given any set S of labelled examples as input, L checks for the existence of a concept $C \in \mathcal{C}$ such that $T(C) \subseteq S$ and C is consistent with S. If such a concept C is found, L returns an arbitrary such C; otherwise L returns some default concept in $\mathcal{C}$.

To show that (T, L) is successful and collusion-free, suppose there is some concept $C \in \mathcal{C}$ such that a given set S of labelled examples is consistent with C and contains $T(C)$. We claim that then such C is uniquely determined. For if there were two distinct concepts $C, C' \in \mathcal{C}$ consistent with S such that $T(C) \cup T(C') \subseteq S$, then $T(C')$, being a subset of S, would be consistent with C and, likewise, $T(C)$ would be consistent with C'—in contradiction to the non-clashing property of T. From the definition of L, it then follows that (T, L) is successful and collusion-free.

Second, suppose T is a teacher mapping and there is a mapping L such that (T, L) is successful and collusion-free, i.e., for all $C \in \mathcal{C}$, we have $L(S) = L(T(C)) = C$ whenever S is consistent with C and contains $T(C)$. To see that T is non-clashing, suppose two concepts $C, C' \in \mathcal{C}$ are both consistent with $T(C) \cup T(C')$. Then $C = L(T(C)) = L(T(C) \cup T(C')) = L(T(C')) = C'$. ■

Consequently, teaching with non-clashing teacher mappings is, in terms of the worst-case number of examples required, the most efficient model that obeys Goldman and Mathias's notion of collusion-freeness. We hence define the notion of no-clash teaching dimension as follows.

Definition 9 *Let $\mathcal{C}$ be a concept class over the instance space $\mathcal{X}$. Let $T : \mathcal{C} \to (\mathcal{X} \times \{0,1\})^*$ be a non-clashing teacher mapping. The* order *of T on $\mathcal{C}$, denoted by $\mathrm{ord}(T, \mathcal{C})$, is then defined by $\mathrm{ord}(T, \mathcal{C}) = \sup\{|T(C)| \mid C \in \mathcal{C}\}$. The* No-Clash Teaching Dimension *of $\mathcal{C}$, denoted by $\mathrm{NCTD}(\mathcal{C})$, is defined as $\mathrm{NCTD}(\mathcal{C}) = \min\{\mathrm{ord}(T, \mathcal{C}) \mid T$ is a non-clashing teacher mapping for $\mathcal{C}\}$.*

From Theorem 8 we obtain that, for every concept class $\mathcal{C}$,

$$\mathrm{NCTD}(\mathcal{C}) = \min\{\mathrm{ord}(T, \mathcal{C}) \mid \text{ there exists an } L \text{ s.t. } (T, L) \text{ is successful and collusion-free on } \mathcal{C}\}\,.$$

T' is said to be an *extension* of T if $T(C) \subseteq T'(C)$ holds for every $C \in \mathcal{C}$. Clearly, if T' is an extension of T and T is non-clashing, then T' is non-clashing. Thus, without loss of generality, we can assume that all sets $T(C)$ used by a non-clashing teacher mapping on the concept class $\mathcal{C}$ are of the same size:

Proposition 10 *Let T be a non-clashing teacher mapping for $\mathcal{C}$. Then there is a non-clashing teacher mapping T' for $\mathcal{C}$ such that $\mathrm{ord}(T,\mathcal{C}) = \mathrm{ord}(T',\mathcal{C}) = |T'(C)|$ for all $C \in \mathcal{C}$.*

As in the case of preference-based teaching, it is natural to study a variant of non-clashing teaching that uses positive examples only.

Definition 11 *Let $\mathcal{C}$ be a concept class over the domain $\mathcal{X}$. A teacher mapping T is called positive on $\mathcal{C}$ if $T(C) \subseteq \mathcal{X} \times \{1\}$ for all $C \in \mathcal{C}$. We then define $\mathrm{NCTD}^+(\mathcal{C}) = \min\{\mathrm{ord}(T,\mathcal{C}) \mid T$ is a positive non-clashing teacher mapping for $\mathcal{C}\}$.*

While many of our definitions and results apply to both finite and infinite concept classes, we will hereafter assume that $\mathcal{X}$ (and $\mathcal{C}$) are finite.

4. Lower Bounds on NCTD and NCTD$^+$

To establish lower bounds on NCTD and NCTD$^+$ for finite concept classes, we first show that $\mathrm{NCTD}(\mathcal{C})$ must be at least as large as the smallest d satisfying $|\mathcal{C}| \leq 2^d\binom{|\mathcal{X}|}{d}$. A similar statement then follows for NCTD$^+$. In fact, we prove a slightly stronger result, replacing $|\mathcal{X}|$ with a potentially smaller value:

Definition 12 *We define $\mathcal{X}_T \subseteq \mathcal{X}$ as the set of instances that are part of a labelled example in a teaching set $T(C)$ for some $C \in \mathcal{C}$. Moreover, we define*

$$X(\mathcal{C}) = \min\{|\mathcal{X}_T| : T \text{ is a non-clashing teacher mapping for } \mathcal{C} \text{ with } \mathrm{ord}(T,\mathcal{C}) = \mathrm{NCTD}(\mathcal{C})\} \ .$$

Intuitively, $X(\mathcal{C})$ is the smallest number of instances that must be employed by any optimal non-clashing teacher mapping for $\mathcal{C}$. Likewise, we define $X^+(\mathcal{C})$ for positive non-clashing teaching.

Theorem 13 *Let $\mathcal{C}$ be any concept class.*

1. *If $\mathrm{NCTD}(\mathcal{C}) = d$, then $|\mathcal{C}| \leq 2^d\binom{X(\mathcal{C})}{d}$.*
2. *If $\mathrm{NCTD}^+(\mathcal{C}) = d$, then $|\mathcal{C}| \leq \sum_{i=0}^{d}\binom{X^+(\mathcal{C})}{i}$.*

Proof To prove statement 1, let $\mathcal{X}'$ be a subset of size $X(\mathcal{C})$ of $\mathcal{X}$. Let $C \mapsto T(C) \subseteq \mathcal{X}' \times \{0,1\}$ be a consistent and non-clashing mapping which witnesses that $\mathrm{NCTD}(\mathcal{C}) = d$, and let L be the mapping such that $L(T(C)) = C$ for all $C \in \mathcal{C}$. By Proposition 10, one may assume without loss of generality that $|T(C)| = d$ for all $C \in \mathcal{C}$. Since T is an injective mapping and there are only $2^d\binom{X(\mathcal{C})}{d}$ labelled teaching sets at our disposal, the claim follows.

Statement 2 is proven analogously, taking into consideration that, in the NCTD$^+$ case, we do not have an analogous statement to Proposition 10, since a concept does not in general contain d or more elements. Note that the formula has no factors 2^i since there are no options for labelling the instances in any set $T(C)$. ■

We will next establish a useful lower bound on $\mathrm{NCTD}(\mathcal{C})$ based on the number of neighbors of any concept in $\mathcal{C}$, as well as a related lower bound on $\mathrm{NCTD}^+(\mathcal{C})$. A concept $C' \in \mathcal{C}$ is a *neighbor* of concept $C \in \mathcal{C}$ if it differs from C on exactly one instance, i.e., if the symmetric difference

$C \Delta C' := (C \setminus C') \cup (C' \setminus C)$ has size one. The *degree* of $C \in \mathcal{C}$, denoted as $\deg_{\mathcal{C}}(C)$, is defined as the number of neighbors of C in $\mathcal{C}$. The average degree of concepts in $\mathcal{C}$ is then denoted by

$$\deg_{avg}(\mathcal{C}) := \frac{1}{|\mathcal{C}|} \cdot \sum_{C \in \mathcal{C}} \deg_{\mathcal{C}}(C) \,.$$

The *dominance* of $C \in \mathcal{C}$, denoted as $\mathrm{dom}_{\mathcal{C}}(C)$, is defined as the number of smaller neighbors of C in $\mathcal{C}$, i.e. neighbors that contain exactly one fewer instance than C.

Theorem 14 *Every concept class $\mathcal{C}$ over a finite domain satisfies* $\mathrm{NCTD}(\mathcal{C}) \geq \left\lceil \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}) \right\rceil$.

Proof Let T be any non-clashing teacher mapping for $\mathcal{C}$. If C_1 and C_2 are neighbors, say $C_1 \Delta C_2 = \{x_i\}$, then at least one of the sets $T(C_1), T(C_2)$ must contain x_i. We obtain $\sum_{C \in \mathcal{C}} |T(C)| \geq \frac{1}{2} \cdot \sum_{C \in \mathcal{C}} \deg_{\mathcal{C}}(C) = |\mathcal{C}| \cdot \frac{1}{2} \cdot \deg_{avg}(\mathcal{C})$. According to the pigeon-hole principle, there must exist a concept $C \in \mathcal{C}$ such that $|T(C)| \geq \left\lceil \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}) \right\rceil$, which concludes the proof of the theorem. ■

Theorem 15 *Every concept class $\mathcal{C}$ over a finite domain satisfies* $\mathrm{NCTD}^+(\mathcal{C}) \geq \max_{C \in \mathcal{C}} \mathrm{dom}_{\mathcal{C}}(C)$.

Proof If the smaller neighbor C' of $C \in \mathcal{C}$ differs from C on instance x_i, then $(x_i, 1)$ must be used in teaching C. Hence, every $C \in \mathcal{C}$ must have a positive teaching set of size at least $\mathrm{dom}_{\mathcal{C}}(C)$. ■

Although the lower bounds in Theorems 14 and 15 are not expected to be attained very often, the following example shows that they are sometimes tight:

Example 1 *Let $\mathcal{P}_2$ be the powerset of $\{a, b\}$. Every concept in $\mathcal{P}_2$ has degree 2, so that $\deg_{avg}(\mathcal{P}_2) = 2$. It follows from Theorem 14 that $\mathrm{NCTD}(\mathcal{P}_2) \geq \frac{1}{2} \cdot \deg_{avg}(\mathcal{P}_2) = 1$. As the mapping T given by*

$$\emptyset \mapsto \{(a,0)\}, \{a\} \mapsto \{(b,0)\}, \{b\} \mapsto \{(b,1)\}, \{a,b\} \mapsto \{(a,1)\} \ ,$$

is non-clashing for $\mathcal{P}_2$, it follows that $\mathrm{NCTD}(\mathcal{P}_2) \leq 1$. Furthermore, since $\mathrm{dom}_{\mathcal{P}_2}(\{a,b\}) = 2$, it follows from Theorem 15 that $\mathrm{NCTD}^+(\mathcal{P}_2) \geq 2$. But the positive mapping T that maps $S \in \mathcal{P}_2$ to $S \times \{1\}$ is trivially non-clashing, and hence $\mathrm{NCTD}^+(\mathcal{P}_2) \leq 2$.

5. Sub-additivity of NCTD and NCTD$^+$

In this section, we will show that the NCTD is sub-additive with respect to the free combination of concept classes. As an application of this result, we will determine the NCTD of the powerset over any finite domain $\mathcal{X}$. While the powerset is a rather special concept class, knowing its NCTD will turn out useful to obtain a variety of further results.

Definition 16 *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be concept classes over disjoint domains $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Then the* free combination *$\mathcal{C}_1 \sqcup \mathcal{C}_2$ of $\mathcal{C}_1$ and $\mathcal{C}_2$ is a concept class over the domain $\mathcal{X}_1 \cup \mathcal{X}_2$ defined by $\mathcal{C}_1 \sqcup \mathcal{C}_2 = \{C_1 \cup C_2 | \ C_1 \in \mathcal{C}_1$ and $C_2 \in \mathcal{C}_2\}$.*

Lemma 17 *Let $\mathcal{C} = \mathcal{C}_1 \sqcup \mathcal{C}_2$ be the free combination of $\mathcal{C}_1$ and $\mathcal{C}_2$. Moreover, for $i = 1, 2$, let T_i be a non-clashing mapping for $\mathcal{C}_i$. Then, for $T(\mathcal{C}_1 \sqcup \mathcal{C}_2)$ defined by setting $T(C_1 \cup C_2) = T_1(C_1) \cup T_2(C_2)$, we have that T is a non-clashing teacher mapping for $\mathcal{C}_1 \sqcup \mathcal{C}_2$. Moreover, as witnessed by T, NCTD acts sub-additively on $\sqcup$, i.e.,*

$$\mathrm{NCTD}(\mathcal{C}_1 \sqcup \mathcal{C}_2) \leq \mathrm{NCTD}(\mathcal{C}_1) + \mathrm{NCTD}(\mathcal{C}_2) \ . \tag{1}$$

Proof Suppose that concepts $C_{i_1}, C_{j_1} \in \mathcal{C}_1$ and $C_{i_2}, C_{j_2} \in \mathcal{C}_2$ give rise to distinct concepts $C_{i_1} \cup C_{i_2}$ and $C_{j_1} \cup C_{j_2} \in \mathcal{C}_1 \sqcup \mathcal{C}_2$ that clash under T. (Without loss of generality we can assume that $i_1 \neq j_1$.) Then $C_{j_1} \cup C_{j_2}$ is consistent with $T_1(C_{i_1}) \cup T_2(C_{i_2})$ and $C_{i_1} \cup C_{i_2}$ is consistent with $T_1(C_{j_1}) \cup T_2(C_{j_2})$. Hence C_{j_1} is consistent with $T_1(C_{i_1})$ and C_{i_1} is consistent with $T_1(C_{j_1})$, that is concepts C_{i_1} and C_{j_1} in $\mathcal{C}_1$ clash under the mapping T_1. ■

Remark 18 In Lemma 17, if T_1 and T_2 are positive non-clashing mappings, then the same proof shows that T (a positive non-clashing mapping) witnesses the fact that NCTD^+ also acts sub-additively on $\sqcup$, i.e.,

$$\mathrm{NCTD}^+(\mathcal{C}_1 \sqcup \mathcal{C}_2) \leq \mathrm{NCTD}^+(\mathcal{C}_1) + \mathrm{NCTD}^+(\mathcal{C}_2) \ . \tag{2}$$

Furthermore, since $\sqcup$ is associative it follows immediately that, for any concept class $\mathcal{C}$

$$\mathrm{NCTD}(\mathcal{C}^k) \leq k \cdot \mathrm{NCTD}(\mathcal{C}) \ \text{ and } \ \mathrm{NCTD}^+(\mathcal{C}^k) \leq k \cdot \mathrm{NCTD}^+(\mathcal{C}) \tag{3}$$

where $\mathcal{C}^k := \mathcal{C}_1 \sqcup \ldots \sqcup \mathcal{C}_k$ and $\mathcal{C}_i := \{C \times \{i\} | \ C \in \mathcal{C}\}$ for $i = 1, \ldots, k$.

These sub-additivity results can be applied in order to determine the NCTD and NCTD^+ of the powerset over an arbitrary finite domain.

Theorem 19 *Let $\mathcal{P}_m$ be the powerset over the domain $\{x_1, \ldots, x_m\}$. Then $\mathrm{NCTD}(\mathcal{P}_m) = \lceil m/2 \rceil$ and $\mathrm{NCTD}^+(\mathcal{P}_m) = m$.*

Proof Since every concept in $\mathcal{P}_m$ has degree m, the average degree of concepts in $\mathcal{P}_m$ equals m as well. Furthermore, the concept $\{x_1, \ldots, x_m\}$ clearly has dominance m in $\mathcal{P}_m$. Now $\mathrm{NCTD}(\mathcal{P}_m) \geq \lceil m/2 \rceil$ and $\mathrm{NCTD}^+(\mathcal{P}_m) \geq m$ follow from Theorems 14 and 15 respectively.

Obviously $\mathrm{NCTD}^+(\mathcal{P}_m) \leq m$. To show that $\mathrm{NCTD}(\mathcal{P}_m) \leq \lceil m/2 \rceil$ it suffices to verify this upper bound for even m. When m is even, we have $\mathcal{P}_m = \mathcal{P}_2^{m/2}$. Now $\mathrm{NCTD}(\mathcal{P}_m) \leq \lceil m/2 \rceil$ follows from $\mathrm{NCTD}(\mathcal{P}_2) = 1$ (compare with Example 1) and from (3). ■

Since the NCTD of any concept class over a domain $\mathcal{X}$ is trivially upper bounded by the NCTD of the power set over $\mathcal{X}$, this result in particular implies that $\lceil |\mathcal{X}|/2 \rceil$ is an upper bound on the NCTD of any concept class over a domain $\mathcal{X}$.

A further consequence of Theorem 19 is that NCTD is sometimes strictly subadditive with respect to free combination, i.e., that inequality (1) is sometimes strict. An example for that is the free combination $\mathcal{P}_m \sqcup \mathcal{P}_m$ of two copies of $\mathcal{P}_m$ for odd m. Since the domain of $\mathcal{P}_m \sqcup \mathcal{P}_m$ has size $2m$, we obtain $\mathrm{NCTD}(\mathcal{P}_m \sqcup \mathcal{P}_m) = m$, while $\mathrm{NCTD}(\mathcal{P}_m) + \mathrm{NCTD}(\mathcal{P}_m) = 2\lceil \frac{m}{2} \rceil = 2\frac{m+1}{2} = m + 1$.

A situation (that we will exploit later) where NCTD^+ acts strictly additively on $\sqcup$ is captured in the following:

Lemma 20 *Let $\mathcal{P}_m$ be the powerset over the domain $\{x_1, \ldots, x_m\}$ and let $\mathcal{C}$ be a concept class with domain $\mathcal{X}$ disjoint from $\{x_1, \ldots, x_m\}$. Then, $\mathrm{NCTD}^+(\mathcal{P}_m \sqcup \mathcal{C}) = \mathrm{NCTD}^+(\mathcal{P}_m) + \mathrm{NCTD}^+(\mathcal{C})$.*

Proof By (2) it suffices to show that $\mathrm{NCTD}^+(\mathcal{P}_k \sqcup \mathcal{C}) \geq \mathrm{NCTD}^+(\mathcal{P}_k) + \mathrm{NCTD}^+(\mathcal{C})$. Theorem 15 implies that, for each $C_i \in \mathcal{C}$, any positive non-clashing mapping T for $\mathcal{P}_k \sqcup \mathcal{C}$ must use $k = \mathrm{NCTD}^+(\mathcal{P}_k)$ examples from $\{x_1, \ldots, x_k\}$ to teach the single concept $\{x_1, \ldots, x_k\} \sqcup C_i$ within the

concept class $\mathcal{P}_k \sqcup C_i$. So the only way that T could use fewer than $k + \mathrm{NCTD}^+(\mathcal{C})$ examples in total for each concept in $\{x_1, \ldots, x_k\} \sqcup \mathcal{C}$ is if each such concept is taught with exactly k examples from $\{x_1, \ldots, x_k\}$, and hence fewer than $\mathrm{NCTD}^+(\mathcal{C})$ examples from $\mathcal{X}$, a contradiction. ■

Furthermore, it is easily seen that the average degree acts additively on $\sqcup$:

Lemma 21 *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be concept classes over disjoint and finite domains. Then the following holds:*

$$\deg_{avg}(\mathcal{C}_1 \sqcup \mathcal{C}_2) = \deg_{avg}(\mathcal{C}_1) + \deg_{avg}(\mathcal{C}_2) \ . \tag{4}$$

Proof Let $\mathcal{C} := \mathcal{C}_1 \sqcup \mathcal{C}_2$. The concepts in $\mathcal{C}$ that are neighbors of $C_1 \cup C_2 \in \mathcal{C}$ are precisely the concepts of the form $C_1 \cup C_2'$ or $C_1' \cup C_2$ where C_2' is a neighbor of C_2 in $\mathcal{C}_2$ and C_1' is a neighbor of C_1 in $\mathcal{C}_1$. Hence

$$\deg_{\mathcal{C}}(C_1 \cup C_2) = \deg_{\mathcal{C}_1}(C_1) + \deg_{\mathcal{C}_2}(C_2) \ .$$

Moreover $|\mathcal{C}| = |\mathcal{C}_1| \cdot |\mathcal{C}_2|$. It follows that

$$\sum_{C \in \mathcal{C}} \deg_{\mathcal{C}}(C) = \sum_{C_1 \in \mathcal{C}_1} \sum_{C_2 \in \mathcal{C}_2} \deg_{\mathcal{C}}(C_1 \cup C_2) = |\mathcal{C}_2| \cdot \sum_{C_1 \in \mathcal{C}_1} \deg_{\mathcal{C}_1}(C_1) + |\mathcal{C}_1| \cdot \sum_{C_2 \in \mathcal{C}_2} \deg_{\mathcal{C}_2}(C_2) \ .$$

Division by $|\mathcal{C}_1| \cdot |\mathcal{C}_2|$ immediately yields (4). ■

The free combination of classes with a tight degree lower bound is again a class with a tight degree lower bound:

Corollary 22 *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two concept classes over disjoint and finite domains, and let $\mathcal{C} = \mathcal{C}_1 \sqcup \mathcal{C}_2$. Then $\mathrm{NCTD}(\mathcal{C}_i) = \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}_i)$ for $i = 1, 2$ implies that $\mathrm{NCTD}(\mathcal{C}) = \frac{1}{2} \cdot \deg_{avg}(\mathcal{C})$.*

Proof The assertion is evident from the chain of inequalities:

$$\mathrm{NCTD}(\mathcal{C}) \overset{(1)}{\leq} \mathrm{NCTD}(\mathcal{C}_1) + \mathrm{NCTD}(\mathcal{C}_2) = \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}_1) + \frac{1}{2} \cdot \deg_{avg}(\mathcal{C}_2) \overset{(4)}{=} \frac{1}{2} \cdot \deg_{avg}(\mathcal{C})$$

and Theorem 14. ■

6. Relation to Other Learning-theoretic Parameters

In this section, we set NCTD in relation to PBTD and VCD, as well as to the smallest possible size of a sample compression scheme for a given concept class.

6.1. PBTD and VCD

Since preference-based teaching is collusion-free (Gao et al., 2017), we obtain the following bounds.

Proposition 23 *Let $\mathcal{C}$ be any concept class. Then $\mathrm{NCTD}(\mathcal{C}) \leq \mathrm{PBTD}(\mathcal{C})$ and $\mathrm{NCTD}^+(\mathcal{C}) \leq \mathrm{PBTD}^+(\mathcal{C})$.*

Remark 24 The first inequality in Proposition 23 is sometimes strict, as witnessed by Theorem 19, which states that $\mathrm{NCTD}(\mathcal{P}_m) = \lceil m/2 \rceil$. By comparison, $\mathrm{PBTD}(\mathcal{P}_m) = m$. In particular, this yields a family of concept classes of strictly increasing NCTD for which PBTD exceeds NCTD by a factor of 2. The fact that the second inequality in Proposition 23 is sometimes strict is witnessed by the simple class $\mathcal{C}$ described in the introduction, with $\mathrm{NCTD}^+(\mathcal{C}) = 1$. Since no concept in $\mathcal{C}$ has a positive teaching set of size 1, Proposition 5 implies $\mathrm{PBTD}^+(\mathcal{C}) = 2$. In particular, these examples witness that Proposition 5 does *not* hold for non-clashing teaching.

Results from the literature can now be combined in a straightforward way in order to formulate an upper bound on NCTD in terms of the VC-dimension.

Proposition 25 $\mathrm{NCTD}(\mathcal{C})$ *is upper-bounded by a function quadratic in* $\mathrm{VCD}(\mathcal{C})$.

Proof PBTD is known to lower-bound the recursive teaching dimension (Gao et al., 2017). Hu et al. (2017) proved that, when $\mathrm{VCD}(\mathcal{C}) = d$, the recursive teaching dimension of $\mathcal{C}$ is no larger than $39.3752 \cdot d^2 - 3.6330 \cdot d$. By Proposition 23, the same upper bound applies to NCTD. ■

However, VCD can also be arbitrarily larger than NCTD, a result that follows immediately from the corresponding result for TD:

Proposition 26 (Goldman and Kearns (1995))
Let $k \in \mathbb{N}$, $k \geq 1$. Then there exists a finite concept class $\mathcal{C}$ such that $\mathrm{TD}^+(\mathcal{C}) = \mathrm{TD}(\mathcal{C}) = 1$ *and* $\mathrm{VCD}(\mathcal{C}) = k$.

So far, there is no concept class for which VCD is known to exceed NCTD. Note that any such concept class would have to fulfill PBTD $>$ VCD as well. We tested those classes for which PBTD $>$ VCD is known from the literature, but found that all of them satisfy NCTD $\leq$ VCD.

As an example, here we present "Warmuth's class." This concept class, shown in Table 1, was communicated by Manfred Warmuth and proven by Darnstädt et al. (2016) to be the smallest concept class for which PBTD exceeds VCD. In particular, $\mathrm{VCD}(\mathcal{C}_W) = 2$ while $\mathrm{PBTD}(\mathcal{C}_W) = 3$.

	x_1	x_2	x_3	x_4	x_5		x_1	x_2	x_3	x_4	x_5
C_1	**1**	0	0	0	**1**	C'_1	**1**	0	**1**	0	1
C_2	**1**	**1**	0	0	0	C'_2	1	**1**	0	**1**	0
C_3	0	**1**	**1**	0	0	C'_3	0	1	**1**	0	**1**
C_4	0	0	**1**	**1**	0	C'_4	**1**	0	1	**1**	0
C_5	0	0	0	**1**	**1**	C'_5	0	**1**	0	1	**1**

Table 1: Warmuth's class $\mathcal{C}_W$, with the highlighted entries (in bold) corresponding to the images of a positive non-clashing teacher mapping. The domain of this class is $\{x_1, \ldots, x_5\}$, and it contains 10 concepts, named C_1 through C_5 and C'_1 through C'_5.

Proposition 27 $\mathrm{NCTD}(\mathcal{C}_W) = \mathrm{NCTD}^+(\mathcal{C}_W) = 2$.

Proof The highlighted labels in Table 1 correspond to a positive non-clashing mapping for $\mathcal{C}_W$, which immediately shows that $\mathrm{NCTD}^+(\mathcal{C}_W) \leq 2$ and thus $\mathrm{NCTD}(\mathcal{C}_W) \leq 2$. To show that

$\mathrm{NCTD}(\mathcal{C}_W) \geq 2$, suppose by way of contradiction that $\mathrm{NCTD}(\mathcal{C}_W) = 1$. Then there is a non-clashing teacher mapping T that assigns every concept in $\mathcal{C}_W$ a teaching set of size 1.

Since C_1 and C_1' differ only on the instance x_3, the mapping T must fulfill either $T(C_1) = \{(x_3, 0)\}$ or $T(C_1') = \{(x_3, 1)\}$.

Case 1. $T(C_1) = \{(x_3, 0)\}$. Since C_2 is consistent with $T(C_1)$, the teaching set for C_2 must be inconsistent with C_1. In particular, $T(C_2) \neq \{(x_4, 0)\}$. This implies $T(C_2') = \{(x_4, 1)\}$, since x_4 is the only instance on which C_2 and C_2' disagree. By an analogous argument concerning C_5 and C_5', one obtains $T(C_5') = \{(x_2, 1)\}$. Now T has a clash on C_2' and C_5', which is a contradiction.

Case 2. $T(C_1') = \{(x_3, 1)\}$. One argues as in Case 1, with C_3' and C_4' in place of C_2 and C_5, yielding $T(C_3) = \{(x_5, 0)\}$ and $T(C_4) = \{x_1, 0)\}$. This is a clash, resulting in a contradiction.

As both cases result in a contradiction, we have $\mathrm{NCTD}(\mathcal{C}_W) > 1$ and thus $\mathrm{NCTD}(\mathcal{C}_W) = 2$. Since NCTD^+ is an upper bound on NCTD, we also have $\mathrm{NCTD}^+(\mathcal{C}_W) = 2$. ■

While the general relationship between NCTD and VCD remains open, it is now known that $\mathrm{NCTD}(\mathcal{C})$ is upper-bounded by $\mathrm{VCD}(\mathcal{C})$ when $\mathcal{C}$ is a finite *maximum* class. For a finite instance space $\mathcal{X}$, a concept class $\mathcal{C}$ of VC dimension d is called maximum if its size $|\mathcal{C}|$ meets Sauer's upper bound $\sum_{i=0}^{d} \binom{|\mathcal{X}|}{i}$ (Sauer, 1972) with equality. Recently, Chalopin et al. (2018) showed that every finite maximum class $\mathcal{C}$ admits a so-called *representation map*, i.e., a function r that maps every concept in $\mathcal{C}$ to a set of at most $d(= \mathrm{VCD}(\mathcal{C}))$ instances, in a way that no two distinct concepts $C, C' \in \mathcal{C}$ both agree on all the instances in $r(C) \cup r(C')$. By definition, any representation map is, translated into our setting, simply a non-clashing teacher mapping of order d for $\mathcal{C}$. Therefore, the result by Chalopin et al. implies that $\mathrm{NCTD}(\mathcal{C}) \leq \mathrm{VCD}(\mathcal{C})$ for finite maximum $\mathcal{C}$.

6.2. Sample Compression

Intuitively, a sample compression scheme (Littlestone and Warmuth, 1986) for a (possibly infinite) concept class $\mathcal{C}$ provides a lossless compression of every set S of labeled examples for any concept in $\mathcal{C}$ in the form of a subset of S. It was proven that the existence of a finite upper bound on the size of the compression sets is equivalent to PAC-learnability, i.e., to finite VC-dimension (Moran and Yehudayoff, 2016; Littlestone and Warmuth, 1986). Open for over 30 years now is the question how closely such an upper bound can be related to the VC-dimension.

Formally, a sample compression scheme of size k for a concept class $\mathcal{C}$ over $\mathcal{X}$ is a pair (f, g) of mappings, where, for every sample set S consistent with some concept $C \in \mathcal{C}$, (i) f maps S to a subset $f(S) \subseteq S$ with $|f(S)| \leq k$; and (ii) $g(f(S))$ maps the compressed set to a concept C' over $\mathcal{X}$ (not necessarily in $\mathcal{C}$) that is consistent with S. By $\mathrm{CN}(\mathcal{C})$ we denote the size of the smallest-size sample compression scheme for $\mathcal{C}$. The open question then is whether $\mathrm{CN}(\mathcal{C})$ is upper-bounded by (a function linear in) $\mathrm{VCD}(\mathcal{C})$.

Some connections between sample compression and teaching have been established in the literature (Doliwa et al., 2014; Darnstädt et al., 2016). The non-clashing property bears some similarities to sample compression and has in fact been used in the context of *unlabelled* sample compression (in which $f(S)$ is an unlabelled set) (Kuzmin and Warmuth, 2007). It is thus natural to ask whether CN is an immediate upper or lower bound on NCTD. Below, we answer this question negatively.

Proposition 28

1. *For every $k \in \mathbb{N}$, $k \geq 1$, there is a concept class $\mathcal{C}$ such that $\mathrm{NCTD}(\mathcal{C}) = \mathrm{PBTD}(\mathcal{C}) = 1$ but $\mathrm{CN}(\mathcal{C}) > k$.*

2. *Let $\mathcal{P}_m$ be the powerset over a domain of size m, where $m \geq 5$ is odd. Then* $\mathrm{CN}(\mathcal{P}_m) < \mathrm{NCTD}(\mathcal{P}_m)$ *and* $2\mathrm{CN}(\mathcal{P}_m) < \mathrm{PBTD}(\mathcal{P}_m)$.

Proof Statement 1 is due to Remark 26, which implies the existence of a concept class $\mathcal{C}$ with $\mathrm{NCTD}(\mathcal{C}) = \mathrm{PBTD}(\mathcal{C}) = 1$ and $\mathrm{VCD}(\mathcal{C}) = 5k$. Then $\mathrm{CN}(\mathcal{C}) > k$ follows from a result by Floyd and Warmuth (1995) that states that no concept class of VC-dimension d has a sample compression scheme of size at most $\frac{d}{5}$.

Statement 2 follows from the obvious fact that $\mathrm{PBTD}(\mathcal{P}_m) = m$, in combination with Theorem 19, as well as with a result by Darnstädt et al. (2016) that shows $\mathrm{CN}(\mathcal{P}_m) \leq \lfloor \frac{m}{2} \rfloor$, for any $m \geq 4$ [2]. ■

Note that the compression function f in a sample compression scheme for $\mathcal{C}$ trivially induces a teacher mapping T_f defined by $T_f(C) = f(\{(x, C(x)) \mid x \in X\})$. The decompression mapping g then satisfies $g(T_f(C)) = C$ for all $C \in \mathcal{C}$. Hence (T_f, g) is a successful teacher-learner pair. Proposition 28.2 now states that there are concept classes for which the teacher-learner pairs (T_f, g) induced by any optimal sample compression scheme necessarily display collusion. In other words, optimal sample compression yields collusive teaching.

An interesting problem is to find more examples of concept classes for which optimal sample compression yields collusive teachers and to determine necessary or sufficient conditions on the structure of such classes. Moreover, at present we do not know how large the gap between sample compression scheme size and NCTD can be.

As mentioned above, representation maps, which were proposed by Kuzmin and Warmuth (2007) and Chalopin et al. (2018), yield non-clashing teacher mappings. Clearly, in unlabelled compression, the representation map that compresses any concept in a class $\mathcal{C}$ to a subset of $\mathcal{X}$ must be injective, so that any two concepts in $\mathcal{C}$ remain distinguishable after compression. In other words, the non-clashing teacher mappings induced by representation maps are *repetition-free*, i.e., they do not map any two distinct concepts $C, C' \in \mathcal{C}$ to labelled samples $T(C), T(C')$ for which

$$\{x \in \mathcal{X} \mid (x, l) \in T(C) \text{ for some } l \in \{0, 1\}\} \neq \{x \in \mathcal{X} \mid (x, l') \in T(C') \text{ for some } l' \in \{0, 1\}\} \ .$$

Requiring no-clash teacher mappings to be repetition-free would be a limitation, as the example of the powerset over any set of m instances, $m \geq 2$, shows. In this case, no-clash teaching can be done with teacher mappings of order $\lceil \frac{m}{2} \rceil$, but it is not hard to see that the best possible repetition-free no-clash teacher mapping is of order m.

7. Complexity of Decision Problems Related to No-clash Teaching

In this section, we address the complexity of the problem of deciding whether or not every concept in a given finite concept class can be taught with a non-clashing teaching set of size at most k, for some specified $k \geq 1$. Surprisingly perhaps, such decision problems are NP-hard, even when $k = 1$ and teaching is done using positive examples only. In contrast, all such decision problems have polynomial time solutions in the PBTD (equivalently, RTD) teaching model; see (Kirkpatrick et al., 2019) for details.

2. When $m = 5k$ for some $k \geq 1$, Darnstädt et al. (2016) even show that $\mathrm{CN}(\mathcal{P}_m) \leq 2k$; hence there is a family of concept classes with CN $<$ NCTD for which the gap between CN and NCTD grows linearly with the size of the instance space.

We show an equivalence between the most highly constrained such decision problem (testing if $\mathrm{NCTD}^+ = 1$, for a given concept class) and a natural (but apparently not previously studied) constrained bipartite matching problem that is related to the well-studied notion of induced matchings. We begin by establishing a preliminary result that will allow us to restrict our complexity analysis to certain normalized concept classes.

Proposition 29 *Let $\mathcal{C}$ be any non-trivial concept class over a finite domain, with at least two non-empty concepts. Then,* $\mathrm{NCTD}^+(\mathcal{C}) = \mathrm{NCTD}^+(\mathcal{C} \setminus \{\emptyset\})$.

Proof Let $\mathcal{C}'$ denote $\mathcal{C} \setminus \{\emptyset\}$. If $\mathcal{C}' = \mathcal{C}$ there is nothing to show. So, suppose that $\mathcal{C}$ contains the empty concept. If $\mathrm{NCTD}^+(\mathcal{C}) = k$ then trivially $\mathrm{NCTD}^+(\mathcal{C}') \leq k$.

For the converse, suppose that $\mathrm{NCTD}^+(\mathcal{C}') = k$, as witnessed by a mapping T.
Case 1. [T does not assign the empty set to any concept.] In this case one can obviously extend T to assign the empty set to the empty concept and thus teach all of $\mathcal{C}$ without clashing using no negative examples and with teaching sets of size at most k. (There are no clashes, because the empty concept cannot be consistent with any of the teaching sets that use at least one positive example.)
Case 2. [T assigns the empty set to some concept C in $\mathcal{C}'$.] Then let x be an element of C. Such x exists because C is not empty. Define T' to be the same as T, except that T' assigns $\{(x, 1)\}$ to C. The mapping T' is non-clashing since it is an extension of T. Thus Case 2 can be reduced to Case 1. ■

Remark 30 It follows immediately from the proof of Proposition 29 that, for any concept class $\mathcal{C}'$ that does not contain the empty concept, $\mathrm{NCTD}^+(\mathcal{C}') = k \geq 1$ if and only if $\mathrm{NCTD}^+(\mathcal{C}') = k$ is witnessed by a positive non-clashing teacher mapping in which every concept is taught with at least one positive instance (i.e. the empty set is not used for teaching). Hereafter, in our consideration of NCTD^+ decision problems, we will assume that concept classes do not contain the empty set and that positive teacher mappings are restricted to those that use at least one positive instance for each concept.

Our goal in the remainder of this section is to set out hardness results for testing $\mathrm{NCTD} = k?$ and $\mathrm{NCTD}^+ = k?$, for fixed $k \geq 1$. We begin by establishing that testing $\mathrm{NCTD}^+ = 1?$, for a given concept class $\mathcal{C}$ is NP-hard. Other results follow by reduction from the $\mathrm{NCTD}^+ = 1?$ decision problem. (It is straightforward to confirm that all of the decision problems $\mathrm{NCTD} \leq k?$ and $\mathrm{NCTD}^+ \leq k?$ are in NP.)

7.1. Testing if $\mathrm{NCTD}^+ = 1$ is NP-hard

We start by observing that a concept class $\mathcal{C}$ over a finite domain $\mathcal{X}$ can be viewed as a bipartite graph $B_{\mathcal{C},\mathcal{X}}$, with vertex classes $\mathcal{C}$ (black vertices) and $\mathcal{X}$ (white vertices) and an edge from $C_i \in \mathcal{C}$ to $x_j \in \mathcal{X}$ whenever $x_j \in C_i$. Under this interpretation, it follows from Remark 30 that deciding if $\mathcal{C}$ has $\mathrm{NCTD}^+ = 1$ is equivalent to deciding if $B_{\mathcal{C},\mathcal{X}}$ admits a matching M such that (i) M saturates all of the black vertices, and (ii) no two edges of M are part of a 4-cycle in $B_{\mathcal{C},\mathcal{X}}$. (Condition (i) ensures that each concept in $\mathcal{C}$ has an associated positive teaching set of size 1, and condition (ii) ensures that the resulting teacher mapping is non-clashing.)

We refer to the problem of deciding if a given bipartite graph B with vertex partition (V_b, V_w) admits a matching M such that (i) M saturates all of the vertices in V_b, and (ii) no two edges of M

are part of a 4-cycle in B, as the *Non-Clashing Bipartite Matching Problem.* The NP-hardness of deciding $\mathrm{NCTD} = 1?$ is thus an immediate consequence of the following:

Theorem 31 *The Non-Clashing Bipartite Matching Problem is NP-hard.*

The proof of Theorem 31 is by reduction from the familiar NP-hard problem 3-SAT. The details are given in Appendix A.

Remark 32 The reduction produces a bipartite graph whose vertices have degree bounded by five. One can conclude then that testing $\mathrm{NCTD}^+ = 1?$ is NP-hard even if concepts contain at most five instances, and instances are contained in at most five concepts. It is natural to ask to what extent this can be tightened. In Appendix B.1, we describe a modification of the reduction that produces a bipartite graph whose vertices have degree bounded by three, from which it follows that testing $\mathrm{NCTD}^+ = 1?$ is NP-hard even if concepts contain at most three instances, and instances are contained in at most three concepts. On the other hand, if either (i) all concepts have at most two instances, or (ii) all instances are contained in at most two concepts, the bipartite graph $B_{\mathcal{C},\mathcal{X}}$ has the property that the degree of all vertices in one of its two parts bounded by at most two. In this case, it follows immediately from the algorithm in Appendix B.2 that testing $\mathrm{NCTD}^+ = 1?$ can be done in polynomial time.

7.2. Testing if $\mathrm{NCTD} = 1$ is NP-hard

We reduce the $\mathrm{NCTD}^+ = 1$ decision problem to the $\mathrm{NCTD} = 1$ decision problem. Given an instance of the $\mathrm{NCTD}^+ = 1$ decision problem, specifically a pair $(\mathcal{C}, \mathcal{X})$, where $\mathcal{C}$ is a concept class over the finite domain $\mathcal{X}$, we make four disjoint copies $(\mathcal{C}^i, \mathcal{X}^i)$, $i \in \{1, 2, 3, 4\}$, and take their union to be an instance of the $\mathrm{NCTD} = 1$ decision problem. We will argue that any $\mathrm{NCTD} = 1$ solution of this composite concept class must use only positive examples for teaching concepts in at least one of the four component concept classes; in this sense it must include a $\mathrm{NCTD}^+ = 1$ solution of the instance $(\mathcal{C}, \mathcal{X})$.

Suppose that some $\mathrm{NCTD} = 1$ solution of the composite concept class uses a negative example for at least one concept in each of the four component concept classes, and consider any four such concepts $C^i \in \mathcal{C}^i$, $i \in \{1, 2, 3, 4\}$. Note that there cannot exist concepts C^i and C^j, with $i \neq j$ that are taught using negative examples drawn from $\mathcal{X}^{i'}$ and $\mathcal{X}^{j'}$, respectively, where $i' \neq j$ and $j' \neq i$, since these would necessarily clash. It follows immediately that no concept C^i is taught with a negative example drawn from its own domain $\mathcal{X}^i$. Furthermore, every domain $\mathcal{X}^i$ must be the source of a negative example for some concept C^j, where $j \neq i$. But this leaves only the possibility that, for some (possibly different) indexing of these four concepts, C^1 is taught with a negative example from $\mathcal{X}^2$ and C^3 is taught with a negative example from $\mathcal{X}^4$, which once again violates the non-clashing property.

7.3. Testing if $\mathrm{NCTD}^+ = k$ is NP-hard, for $k > 1$.

Again we describe a reduction from the $\mathrm{NCTD}^+ = 1$ decision problem. Given an instance of the $\mathrm{NCTD}^+ = 1$ decision problem, specifically a pair $(\mathcal{C}, \mathcal{X})$, where $\mathcal{C}$ is a concept class over the finite domain $\mathcal{X}$ disjoint from $\{x_1, \ldots, x_{k-1}\}$, we construct the concept class $\mathcal{P}_{k-1} \sqcup \mathcal{C}$, where $\mathcal{P}_{k-1}$ denotes the power set on $\{x_1, \ldots, x_{k-1}\}$. By Lemma 20, we know that $\mathrm{NCTD}^+(\mathcal{P}_{k-1} \sqcup \mathcal{C}) = k - 1 + \mathrm{NCTD}^+(\mathcal{C})$, so $\mathrm{NCTD}^+(\mathcal{C}) = 1$ iff $\mathrm{NCTD}^+(\mathcal{P}_{k-1} \sqcup \mathcal{C}) = k$.

Using a slightly move involved reduction from the $\mathrm{NCTD}^{+} = 1$ decision problem, we can show that the $\mathrm{NCTD} = k$ decision problem is also NP-hard. By comparison, the corresponding decision problems for PBTD (equivalently, for RTD) can be solved efficiently. See (Kirkpatrick et al., 2019) for details.

8. Conclusions

No-clash teaching represents the limit of data efficiency that can be achieved in teaching settings obeying Goldman and Mathias's notion of collusion-freeness. Therefore, it is the sole most promising collusion-free teaching model to shed light on two open problems in computational learning theory, namely (i) to find a teaching complexity parameter that is upper-bounded by a function linear in VCD, and (ii) to establish an upper bound on the size of smallest sample compression schemes that is linear in VCD. If *any* collusion-free teaching model yields a complexity upper-bounded by (a function linear in) VCD, then no-clash teaching does. Likewise, if *any* collusion-free model is powerful enough to compress concepts as efficiently as sample compression schemes do, then no-clash teaching is.

The most fundamental open question resulting from our paper is probably whether NCTD is upper-bounded by VCD in general.

Furthermore, our results introduce some intriguing connections between NCTD and the well-studied field of constrained matching in bipartite graphs that may open up a line of study that relates teaching complexity, as well as sample compression and VCD, to fundamental issues in matching theory.

References

Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135, 1980.

Brenna Argall, Sonia Chernova, Manuela M. Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 2009.

Frank Balbach. Measuring teachability using variants of the teaching dimension. *Theoretical Computer Science*, 397(1–3):94–113, 2008.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML*, pages 41–48, 2009.

Jérémie Chalopin, Victor Chepoi, Shay Moran, and Manfred K. Warmuth. Unlabeled sample compression schemes and corner peelings for ample and maximum classes. *ArXiv*, abs/1812.02099, 2018. URL http://arxiv.org/abs/1812.02099.

Malte Darnstädt, Thorsten Kiss, Hans Ulrich Simon, and Sandra Zilles. Order compression schemes. *Theor. Comput. Sci.*, 620:73–90, 2016.

François Denis. Learning regular languages from simple positive examples. *Machine Learning*, 44 (1/2):37–66, 2001.

Thorsten Doliwa, Gaojian Fan, Hans Ulrich Simon, and Sandra Zilles. Recursive teaching dimension, VC-dimension and sample compression. *J. Mach. Learn. Res.*, 15:3107–3131, 2014.

Sally Floyd and Manfred Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning*, 21(3):1–36, 1995.

Ziyuan Gao, Christoph Ries, Hans Ulrich Simon, and Sandra Zilles. Preference-based teaching. *J. Mach. Learn. Res.*, 18(31):1–32, 2017.

Sally A. Goldman and Michael J. Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50(1):20–31, 1995.

Sally A. Goldman and H. David Mathias. Teaching a smarter learner. *J. Comput. Syst. Sci.*, 52(2): 255–267, 1996.

Mark K Ho, Michael Littman, James MacGlashan, Fiery Cushman, and Joseph L Austerweil. Showing versus doing: Teaching by demonstration. In *Advances in Neural Information Processing Systems 29 (NIPS)*, pages 3027–3035. 2016.

Lunjia Hu, Ruihan Wu, Tianhong Li, and Liwei Wang. Quadratic upper bound for recursive teaching dimension of finite VC classes. In *Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017*, pages 1147–1156, 2017.

David Kirkpatrick, Hans U. Simon, and Sandra Zilles. Optimal collusion-free teaching. *ArXiv*, 2019.

Dima Kuzmin and Manfred K. Warmuth. Unlabeled compression schemes for maximum classes. *J. Mach. Learn. Res.*, 8:2047–2081, 2007.

Nick Littlestone and Manfred K. Warmuth. Relating data compression and learnability. Technical Report, UC Santa Cruz, 1986.

Shay Moran and Amir Yehudayoff. Sample compression schemes for VC classes. *J. ACM*, 63(3): 21:1–21:10, 2016.

Shay Moran, Amir Shpilka, Avi Wigderson, and Amir Yehudayoff. Teaching and compressing for low VC-dimension. *CoRR*, abs/1502.06187, 2015.

Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13 (1):145–147, 1972.

Ingo Schwab, Wolfgang Pohl, and Ivan Koychev. Learning to recommend from positive evidence. In *Proceedings of the 5th International Conference on Intelligent User Interfaces, IUI*, pages 241–247, 2000.

Patrick Shafto, Noah D. Goodman, and Thomas L. Griffiths. A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive psychology*, 71:55–89, 2014.

Ayumi Shinohara and Satoru Miyano. Teachability in computational learning. *New Gen. Comput.*, 8:337–348, 1991.

Vladimir N. Vapnik and Alexey Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Probability and Appl.*, 16(2):264–280, 1971.

Chunlin Wang, Chris H. Q. Ding, Richard F. Meraz, and Stephen R. Holbrook. Psol: a positive sample only learning algorithm for finding non-coding RNA genes. *Bioinformatics*, 22(21):2590–2596, 2006.

Kenneth Wexler and Peter W. Culicover. *Formal principles of language acquisition*. MIT Press, 1980.

Xiaojin Zhu, Adish Singla, Sandra Zilles, and Anna N. Rafferty. An overview of machine teaching. *CoRR*, abs/1801.05927, 2018.

Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Models of cooperative teaching and learning. *J. Mach. Learn. Res.*, 12:349–384, 2011.

Appendix A. Proof of Theorem 31

Proof We describe a parsimonious reduction from the familiar NP-hard problem 3-SAT, an instance of which is a set $\mathcal{D} = \{D^1, \ldots, D^m\}$ of clauses, each of which is a disjunction of three literals drawn from an underlying set $\mathcal{V} = \{V^1, \ldots, V^n\}$ of variables. Specifically, given an instance $\mathcal{D}$ of 3-SAT, we construct a bipartite graph $B_{\mathcal{D}}$ (vertices are either black or white, and all edges join a black vertex to a white vertex) that admits a matching M such that (i) M saturates all of the black vertices, and (ii) no two edges of M are part of a 4-cycle in B, if and only if the instance $\mathcal{D}$ is satisfiable.

To this end, we first associate with each variable V^i a *variable gadget*: a ring of $4m$ vertices, with alternating subscripted labels v^i and w^i, emphasizing its bipartite nature (cf. Figure 1(a)). A matching that saturates all of the v^i-vertices (black) of this gadget is of one of two types, illustrated in Figure 1(b) and (c)), which we associate with the two possible truth assignments to V^i.

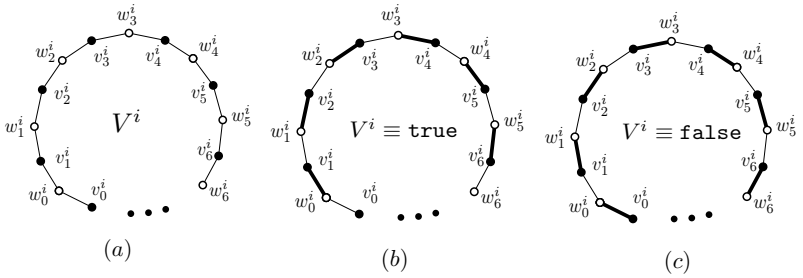

Figure 1: VariableGadget

We associate with each clause D^j a *clause gadget* consisting of 10 vertices, with subscripted labels p^j, q^j, r^j and s^j (cf. Figure 2(a)). It is straightforward to confirm that any matching that saturates all of the r^j and q^j-vertices (black) must use exactly one of the three $p^j q^j$-edges, illustrated in Figure 2(b) (c) and (d)). We refer to the $p^j q^j$-edges as *portals* of the clause gadget, since their endpoints are the only points of connection with other parts of the full construction.

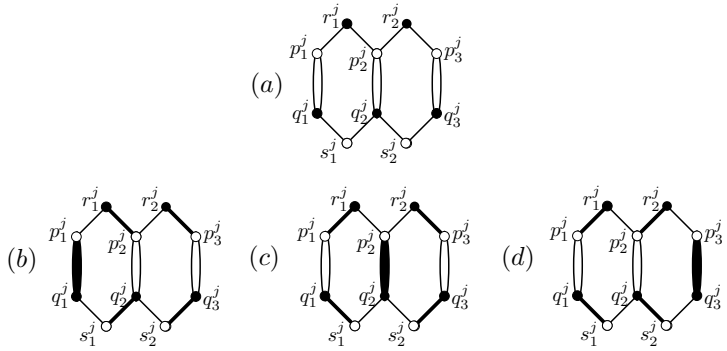

Figure 2: ClauseGadget

We complete the construction by adding edges from vertex gadgets to appropriate clause gadget portals. Specifically, (i) if the k-th literal in clause D^j is V^i, then we add edges from v^i_{2j} to p^j_k and q^j_k to w^i_{2j} (cf. Figure 3(a)) and (ii) if the k-th literal in clause D^j is $\overline{V^i}$, then we add edges from v^i_{2j} to p^j_k and q^j_k to w^i_{2j-1} (cf. Figure 3(b)). These *connector* edges, shown dashed in Figures 3(a) and (b), are forbidden in any matching satisfying the constraints set out above, by the inclusion, for each such edge, of a pair of additional vertices and associated edges, as illustrated in Figure 3(c). (Observe that since the graph has the same number of black and white vertices, a matching that saturates all of the black vertices must also saturate all of the white vertices. Thus, for each connector edge, the middle edge of its bridging path is forced to belong to the matching; otherwise, the end edges of the bridging path must both be chosen, resulting in a clash.)

It follows that if the k-th literal in clause D^j is V^i, and the edge $p^j_k q^j_k$ belongs to the constrained matching then edge $v^i_{2j} w^i_{2j}$ cannot belong. Similarly, if the k-th literal in clause D^j is $\overline{V^i}$, and the edge $p^j_k q^j_k$ belongs to the constrained matching then edge $v^i_{2j} w^i_{2j-1}$ cannot belong.

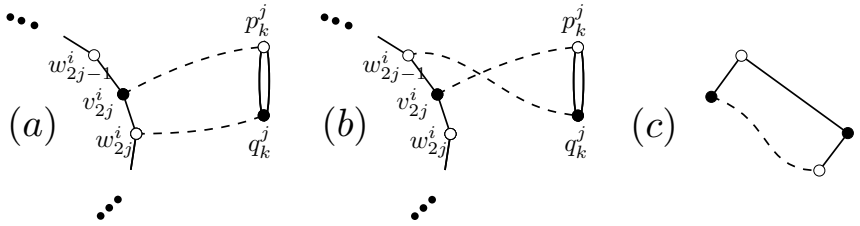

Figure 3: ConnectorGadgets

To complete the proof it remains to argue that the resulting graph $B_{\mathcal{D}}$ admits a matching M such that (i) M saturates all of the black vertices, and (ii) no two edges of M are part of a 4-cycle in $B_{\mathcal{D}}$, if and only if the instance $\mathcal{D}$ is satisfiable. Suppose first that $B_{\mathcal{D}}$ admits such a matching M. Since none of the connector edges are included in M, it follows (as argued above) that in every vertex gadget the black vertices are saturated in one of the two ways illustrated in Figure 1(b) and 1(c)). Similarly, in every clause gadget, the black vertices are saturated in one of the three ways

illustrated in Figure 2(b), 2(c) and 2(d)). Suppose that the portal edge $p_k^j q_k^j$ of the gadget associated with clause D^j belongs to the matching M. Then, by our choice of connector edges, if the k-th literal in clause D_j is V^i, it must be that edge $v_{2j}^i w_{2j}^i$ does not belong to M, that is the matching on the variable gadget associated with V^i has the associated truth assignment `true`. Similarly, if the k-th literal in clause D_j is $\overline{V^i}$, it must be that edge $v_{2j}^i w_{2j-1}^i$ does not belong to M, that is the matching on the variable gadget associated with V^i has the associated truth assignment `false`. It follows that the truth assignment to the variables in $\mathcal{V}$, associated with the matchings induced on the vertex gadgets, satisfies all of the clauses in $\mathcal{D}$.

On the other hand, suppose that $\mathcal{D}$ is satisfiable, that is there is an assignment of truth values to the variables in $\mathcal{V}$ that satisfies all of the clauses in $\mathcal{D}$. Then, if we (i) choose the matching on the vertex gadget associated with V^i to be the one corresponding to its truth assignment, and (ii) choose any matching on the clause gadget associated with clause D^j including a portal edge associated with one of the satisfied literals in D^j, and (iii) choose all of the edges added to prevent the choice of connector edges, it is straightforward to confirm that the chosen edges form a matching M in $B_{\mathcal{D}}$ such that (i) M saturates all of the black vertices, and (ii) no two edges of M are part of a 4-cycle in $B_{\mathcal{D}}$. ■

Appendix B. Complexity of Degree-bounded Instances of Non-clashing Bipartite Matching

The reduction described in the proof of Theorem 31 produces a bipartite graph whose vertices have degree bounded by five. (This occurs for the vertices p_2^j and q_2^j of the clause gadgets, both of which have three incident edges within the gadget and two from a bridged connector.) It is natural to ask if the hardness result continues to hold if this degree bound is reduced. In the next subsection we describe a fairly simple modification of both our clause and connector structures allow us to reduce the maximum degree to three. Following that, we show that if the maximum degree among vertices in either part of a given bipartite graph is reduced to two there is a polynomial time algorithm to decide if it admits a non-clashing matching.

B.1. A modified reduction with maximum degree three

We begin by describing a new clause gadget, illustrated in Figure 4(a), with the same p-q portal structure as before but with the additional property that all p and q vertices have degree two. It is straightforward to confirm that, up to symmetry, the matching illustrated in Figure 4(b) is the only matching that saturates all of the vertices using only edges internal to the gadget.

Next we describe a somewhat more complicated connector structure that is used to link vertices in the variable gadgets with portal vertices of the new clause gadget. Schematically, as illustrated in Figures 5(a) and (b), the connector structure plays exactly the same role as its counterpart (pair of bridged edges) in the earlier construction. The new connector structure, illustrated in Figures 5(c), also contains edges, dashed as before, that cannot be part of any perfect non-clashing matching. Their role, as before, is simply to constrain the choice of other edges (in any perfect non-clashing matching).

It is easiest to argue first that neither of the dashed diagonals can be used. If both are used then edge $r_4 s_4$ must also be used, creating a clash. On the other hand if just one, say $r_3 s_5$ is used, then either $r_4 s_4$ must also be used or both $r_4 r_5$ and $s_3 s_4$ must be used, creating a clash in either case.

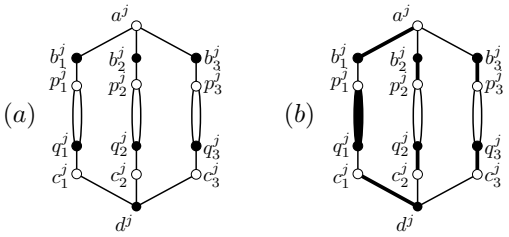

Figure 4: NewClauseGadget

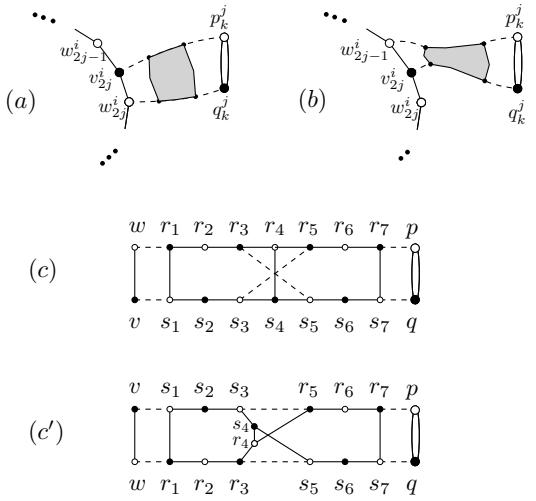

Figure 5: NewConnectorGadgets

By parity, an even number of the horizontal dashed edges are used in any perfect matching. Since it is impossible to choose both wr_1 and vs_1 (or both r_7p and s_7q) in a non-clashing matching, it suffices to rule out the case where exactly one of wr_1 and vs_1 and exactly one of r_7p and s_7q belong to a perfect matching. Suppose r_7p (but not s_7q) is chosen. Then the matching is forced to include r_5r_6 and s_6s_7 (in order to saturate r_6 and s_7). This in turn forces the choice of r_3r_4 and s_4s_5 (in order to saturate r_4 and s_5), creating a clash. By symmetry, it follows that none of the horizontal dashed edges can be used in a perfect non-clashing matching.

It remains to argue that (i) if a non-clashing matching contains edge pq then edge vw cannot belong (and vice versa); (ii) there is a non-clashing matching of the connector gadget that contains edge pq but leaves both v and w exposed (and vice versa); and (iii) there is a non-clashing matching of the connector gadget that leaves all of v, w, p and q exposed. For (i), we observe that, by chained forcing as above, the inclusion of pq forces the inclusion of r_1s_1 (and, by symmetry, the inclusion of vw forces the inclusion of r_7s_7). Properties (ii) and (iii) are illustrated in Figure 6.

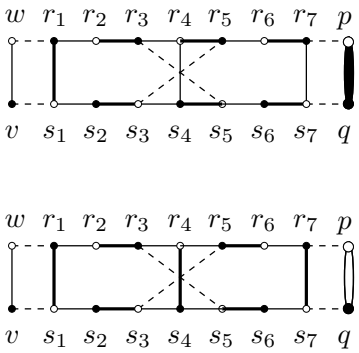

Figure 6: ConnectorMatchings

B.2. An efficient algorithm for Non-Clashing Bipartite Matching, when the maximum degree on either part is at most two

Suppose we are given a bipartite graph B whose vertices are either black or white, and all edges join a black vertex to a white vertex. We want to determine if B admits a matching M such that (i) M saturates all of the black vertices, and (ii) no two edges of M are part of a 4-cycle in B.

Suppose further that the vertices on one of the two parts of B all have degree at most two. We can assume, without loss of generality that they all have degree exactly two, since edges with an endpoint of degree one can be (incrementally) included in a maximum matching M without risk of being part of a 4-cycle in B.

We say that a pair of vertices in this degree-bounded part are *twins* if they have the same adjacent vertices. We can assume that B has no twins since (i) twins cannot both be saturated without producing a forbidden 4-cycle, and therefore (ii) the existence of black twins immediately precludes

a non-clashing matching, and (iii) any pair of white twins can be replaced by a single copy of the twinned vertex.

With this simplification it is easy to confirm that any matching that saturates the black vertices, the existence of which can be determined in polynomial time, must be non-clashing,