Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)

The Complexity of Learning Acyclic CP-Nets

Eisa Alanazi, Malek Mouhoub and Sandra Zilles
Department of Computer Science
University of Regina
Regina, SK, Canada
{alanazie,mouhoubm,zilles}@cs.uregina.ca

Abstract

Learning of user preferences has become a core issue in AI research. For example, recent studies investigate learning of Conditional Preference Networks (CP-nets) from partial information. To assess the optimality of learning algorithms as well as to better understand the combinatorial structure of CP-net classes, it is helpful to calculate certain learning-theoretic information complexity parameters. This paper provides theoretical justification for exact values (or in some cases bounds) of some of the most central information complexity parameters, namely the VC dimension, the (recursive) teaching dimension, the self-directed learning complexity, and the optimal mistake bound, for classes of acyclic CP-nets. We further provide an algorithm that learns tree-structured CP-nets from membership queries. Using our results on complexity parameters, we can assess the optimality of our algorithm as well as that of another query learning algorithm for acyclic CP-nets presented in the literature.

1 Introduction

Since preference learning is important in many AI applications, there is a need for a strong theoretical underpinning of research on this topic. In recent years, substantial advances have been made in this field, for example in the design of Conditional Preference Networks (CP-nets) [Boutilier *et al.*, 2004] and the study of their learnability. A CP-net is a compact preference representation for multi-attribute domains where the preference of one attribute may depend on the values of other attributes.

Koriche and Zanuttini (2010) investigated query learning of bounded acyclic CP-nets (i.e., with a bound on the number of attributes on which the preferences for any attribute may depend). Their successful algorithms used both membership and equivalence queries, cf. [Angluin, 1988], while they proved that equivalence queries alone are not sufficient for efficient learnability. CP-nets have also been studied in models of passive learning from examples, both for batch learning [Dimopoulos *et al.*, 2009; Lang and Mengin, 2009; Liu *et al.*, 2014] and for online learning [Guerin *et al.*, 2013].

A fundamental question in assessing the proposed algorithms is how many queries/examples would be needed by the best possible learning algorithm in the given learning model. For several models, lower bounds can be derived from the Vapnik Chervonenkis dimension (VCD, [Vapnik and Chervonenkis, 1971]). This central parameter is one of several that, in addition to yielding bounds on the performance of learning algorithms, provide deep insights into the combinatorial structure of the studied concept class. Such insights can in turn help to design new learning algorithms.

Our main contributions are the following:

(a) We provide the first study that exactly calculates the VCD for the class of unbounded acyclic CP-nets, and give a lower bound for any bound k. So far, the only existing studies present a lower bound [Koriche and Zanuttini, 2010], which we prove incorrect for large values of k, and asymptotic complexities [Chevaleyre *et al.*, 2010]. The latter show that VCD=$\Theta(2^n)$ for $k = n-1$ and $\tilde{\Theta}(n2^k)$ when $k \in o(n)$, in agreement with our result that VCD equals $2^n - 1$ for $k = n-1$, and at least $2^k - 1 + (n-k)2^k$ for general values of k. It should be noted that both previous studies assume that CP-nets can be incomplete, i.e., for some variables no preference relations may be given. In our study, we make the (not uncommon) assumption that CP-nets are complete, but our result on VCD also applies to the more general case that includes incomplete CP-nets. Further, our results are more general than existing ones in that they apply also to CP-nets with multi-valued variables (as opposed to binary variables).

As a byproduct of our study, we obtain that the VCD of the class of *all* consistent CP-nets (whether acyclic or cyclic)[1] equals that of the class of all acyclic CP-nets. Hence, the class of acyclic CP-nets is less expressive than that of all consistent CP-nets, but may (at least in some models) be as hard to learn.

(b) We further provide exact values (or, in some cases, non-trivial bounds) for other important information complexity parameters, namely the teaching dimension [Goldman and Kearns, 1995], the recursive teaching dimension [Zilles *et al.*, 2011], the self-directed learning complexity [Goldman *et al.*, 1993], and the optimal mistake bound [Littlestone, 1988].

[1] A consistent CP-net is one that does not prefer an outcome o over another outcome $\hat{o}$ while at the same time preferring $\hat{o}$ over o. Acyclic CP-nets are always consistent, but cyclic ones are not necessarily so.

(c) We present a new algorithm that learns tree-structured CP-nets from membership queries and use our results on the teaching dimension to show that our algorithm is close to optimal.

(d) We re-assess the degree of optimality of Koriche and Zanuttini's algorithm for learning bounded acyclic CP-nets, using our result on the VCD.

2 Background

A concept is any subset of a given universe $\mathcal{X}$ (called instance space) and a concept class is a set of concepts. In this paper, we assume that $\mathcal{X}$ (and hence every concept class) is finite. In concept learning, the learner receives information on a target concept c^* contained in a fixed concept class and is requested to identify or approximate c^* from the given information. Information consists of labelled examples (x, ℓ) where $x \in \mathcal{X}$, $\ell \in \{0, 1\}$, and $\ell = 1$ iff x is contained in the target concept. We also write $c(x) = 1$ if $x \in c$, and $c(x) = 0$ otherwise. The concept classes we study below are subclasses of the class of so-called conditional preference networks (CP-nets).

2.1 Conditional Preference Networks

We largely follow the notation introduced by Boutilier et al. (2004) in their seminal work on CP-nets.

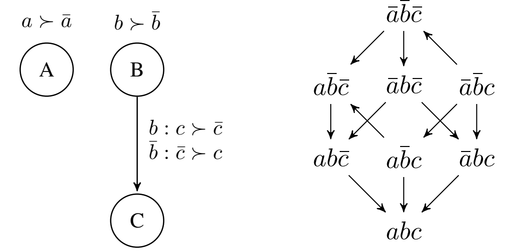
Let $V = \{v_1, \ldots, v_n\}$ be a set of variables. Each variable v_i has a set of possible values (its domain) $D_{v_i} = \{v_1^i, \ldots, v_m^i\}$. We assume that every domain D_{v_i} is of size m, independent of i. An assignment x to a set of variables $X \subseteq V$ is a mapping for every variable $v_i \in X$ to a value from D_{v_i}. We denote the set of all assignments of $X \subseteq V$ by $\mathcal{O}_X$ and remove the subscript when $X = V$. A *preference* is an irreflexive, transitive binary relation $\succ$. For any $o, \hat{o} \in \mathcal{O}$, we write $o \succ \hat{o}$ to denote the fact that o is strictly preferred to $\hat{o}$.

CP-nets provide a compact representation of preferences over $\mathcal{O}$. For every $v_i \in V$, the decision maker chooses a set $Pa(v_i) \subseteq V \backslash \{v_i\}$ of parent variables that influence the preference order of v_i. For every $u \in \mathcal{O}_{Pa(v_i)}$, the decision maker specifies a total ordering $\succ_u^{v_i}$ over D_{v_i}. We refer to $\succ_u^{v_i}$ as the conditional preference statement of v_i in the context of u. The set of preference statements $\{\succ_{u_1}^{v_i}, \ldots, \succ_{u_k}^{v_i}\}$, where the u_j are all contexts that yield non-empty conditional preference statements, is the *Conditional Preference Table* $\mathrm{CPT}(v_i)$. Such statements expressed in a CPT are under the ceteris paribus interpretation.

Definition 1. *[Boutilier* et al.*, 2004] Given,* V*,* $Pa(v)$*, and* $\mathrm{CPT}(v)$ *for* $v \in V$*, a CP-net is a directed graph* (V, E)*, where, for any* $v_i, v_j \in V$*,* $(v_i, v_j) \in E$ *iff* $v_i \in Pa(v_j)$*.*

Example 1. *Figure 1a shows a CP-net over* $V = \{A, B, C\}$ *with* $D_A = \{a, \bar{a}\}$*,* $D_B = \{b, \bar{b}\}$*,* $D_C = \{c, \bar{c}\}$*. Each variable is annotated with its* CPT*. For variable A, the user prefers* a *to* $\bar{a}$ *unconditionally. For C, the preference depends on the values of B, i.e.,* $Pa(C) = \{B\}$*. For instance, in the context of* $\bar{b}$*,* $\bar{c}$ *is preferred over* c*.*

Two outcomes $o, \hat{o} \in \mathcal{O}$ are swap outcomes ('swaps' for short) if they differ in the value of exactly one variable v_i; then v_i is called the swapped variable [Boutilier *et al.*, 2004].



(a) The CP-net (b) The induced graph

Figure 1: An acyclic CP-net and its induced graph

We use $o[X]$ to denote the projection of o onto $X \subset V$ and write $o[v_i]$ instead of $o[\{v_i\}]$. The size of a preference table for a variable v_i, denoted by $size(\mathrm{CPT}(v_i))$, is the number of preference statements it holds. The size of a CP-net is the sum of its tables' sizes.

Example 2. *In Figure 1a,* $abc, \bar{a}bc$ *are swaps over the swapped variable A. The CP-net size is* $1 + 1 + 2 = 4$*.*

The *induced graph* of a CP-net N is defined as follows. Each vertex in the induced graph represents an outcome $o \in \mathcal{O}$. An edge from $\hat{o}$ to o exists iff $(o, \hat{o}) \in \mathcal{O} \times \mathcal{O}$ is a swap w.r.t. some $v_i \in V$ and $o[v_i]$ precedes $\hat{o}[v_i]$ in $\succ_{o[Pa(v_i)]}^{v_i}$ [Boutilier *et al.*, 2004]. N defines a partial order $\succ$ over $\mathcal{O}$ that is given by the transitive closure of its induced graph. If $o \succ \hat{o}$ we say N entails $(o, \hat{o})$. N is consistent if there is no $o \in \mathcal{O}$ with $o \succ o$, i.e., if its induced graph is acyclic. Since acyclic CP-nets are always consistent [Boutilier *et al.*, 2004], their class is particularly interesting for learnability studies, as is its subclass of separable CP-nets, which have no edges.

Example 3. *In the network in Figure 1a,* $abc \succ \bar{a}\bar{b}c$*, as witnessed by the path* $\bar{a}\bar{b}c \to a\bar{b}c \to abc$ *in the induced graph shown in Figure 1b.*

As CP-net semantics are completely determined by the preference relation over swaps, one may consider the set $\mathcal{X}_{swap} = \{(o_1, o_2) \in \mathcal{O} \times \mathcal{O} \mid (o_1, o_2)$ is a swap and o_1 is lexicographically smaller than $o_2\}$ as the instance space; a quick calculation (which is omitted due to space constraints) shows that its size is $|\mathcal{X}_{swap}| = nm^{n-1}\binom{m}{2} = \frac{m^n n(m-1)}{2}$. (We assume a fixed lexicographic order.) For $x = (o_1, o_2) \in \mathcal{X}_{swap}$, let $V(x)$ denote the swapped variable of x. We refer to the first and second outcomes of an example x as $x.1$ and $x.2$, respectively. We use $x[\Gamma]$ to denote the assignments (in both $x.1$ and $x.2$) of $\Gamma \subseteq V \backslash \{V(x)\}$.

In what follows, we fix $k \in \{0, \ldots, n-1\}$ and consider the class $\mathcal{C}_{ac}$ of all acyclic CP-nets whose nodes have an indegree of at most k. We use $\mathcal{C}_{unb}$, $\mathcal{C}_{tree}$ and $\mathcal{C}_{sep}$ for the classes of acyclic CP-nets with indegree at most $n-1, 1$, and 0, respectively. A concept c contains a swap pair x iff the corresponding CP-net entails $(x.1, x.2)$. By $size(c)$, we refer to the size of the CP-net represented by c. Lastly, $\mathcal{M} = \max\{size(c) \mid c \in \mathcal{C}_{ac}\}$ is the maximum number of statements in any concept in $\mathcal{C}_{ac}$. It can be verified that $\mathcal{M} = (n-k)m^k + \sum_{i=0}^{k-1} m^i$.

2.2 Information Complexity Parameters

The combinatorial structure of a concept class $\mathcal{C}$ has implications on the complexity of learning $\mathcal{C}$, in particular on the sample complexity (sometimes called information complexity), which refers to the number of labelled examples the learner needs in order to identify any target concept in the class under the constraints of a given learning model. One of the most important complexity parameters studied in machine learning is the Vapnik-Chervonenkis dimension (VCD). In what follows, let $\mathcal{C}$ be a concept class over the (finite) instance space $\mathcal{X}$.

Definition 2. *[Vapnik and Chervonenkis, 1971] A subset $Y \subseteq \mathcal{X}$ is shattered by $\mathcal{C}$ if the projection of $\mathcal{C}$ onto Y has $2^{|Y|}$ concepts. The VC dimension of $\mathcal{C}$, denoted by $\mathrm{VCD}(\mathcal{C})$, is the size of the largest subset of $\mathcal{X}$ that is shattered by $\mathcal{C}$.*

The number of randomly chosen examples needed to identify concepts from $\mathcal{C}$ in the PAC-learning model is linear in $\mathrm{VCD}(\mathcal{C})$ [Blumer *et al.*, 1989]. By contrast to learning from random examples, in teaching models, the learner is provided with well-chosen labelled examples.

Definition 3. *[Goldman and Kearns, 1995; Shinohara and Miyano, 1991] A teaching set for a concept $c^* \in \mathcal{C}$ with respect to $\mathcal{C}$ is a set $S = \{(x_1, \ell_1), \ldots, (x_z, \ell_z)\}$ of labelled examples such that c^* is the only concept $c \in \mathcal{C}$ that satisfies $c(x_i) = \ell_i$ for all $i \in \{1, \ldots, z\}$. The teaching dimension of c with respect to $\mathcal{C}$, 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 $\mathcal{C}$, denoted by $\mathrm{TD}(\mathcal{C})$, is given by $\mathrm{TD}(\mathcal{C}) = \max\{\mathrm{TD}(c, \mathcal{C}) \mid c \in \mathcal{C}\}$.*

$\mathrm{TD}_{min}(\mathcal{C}) = \min\{\mathrm{TD}(c, \mathcal{C}) \mid c \in \mathcal{C}\}$ denotes the smallest TD of any $c \in \mathcal{C}$. A well-studied variation of teaching is called recursive teaching. Its complexity parameter, the recursive teaching dimension, is defined by recursively removing from $\mathcal{C}$ all the concepts with the smallest TD and then taking the maximum over the smallest TDs encountered in that process. For the corresponding definition of teachers, see [Zilles *et al.*, 2011].

Definition 4. *[Zilles* et al.*, 2011] Let $\mathcal{C}_0 = \mathcal{C}$ and, for all i such that $\mathcal{C}_i \neq \emptyset$, define $\mathcal{C}_{i+1} = \mathcal{C}_i \setminus \{c \in \mathcal{C}_i \mid \mathrm{TD}(c, \mathcal{C}_i) = \mathrm{TD}_{min}(\mathcal{C}_i)\}$. The recursive teaching dimension of $\mathcal{C}$, denoted by $\mathrm{RTD}(\mathcal{C})$, is defined by $\mathrm{RTD}(\mathcal{C}) = \max\{\mathrm{TD}_{min}(\mathcal{C}_i) \mid i \geq 0\}$.*

As opposed to the TD, the RTD exhibits interesting relationships to the VCD. For example, every maximum class, i.e., a class $\mathcal{C}$ whose size $|\mathcal{C}|$ meets Sauer's upper bound $\binom{|\mathcal{X}|}{0} + \binom{|\mathcal{X}|}{1} + \ldots + \binom{|\mathcal{X}|}{\mathrm{VCD}(\mathcal{C})}$ [Sauer, 1972], fulfills $\mathrm{RTD}(\mathcal{C}) = \mathrm{VCD}(\mathcal{C})$; the same is true for classes of VCD 1 and for intersection-closed classes [Doliwa *et al.*, 2014].

We will further determine complexity parameters for online prediction, namely the self-directed learning complexity and the optimal mistake bound. A self-directed learner passes a prediction $(x, \ell) \in \mathcal{X} \times \{0, 1\}$ to an oracle, which responds with the information whether or not the target concept c^* fulfills $c^*(x) = \ell$. In case $c^*(x) \neq \ell$, the learner has made a mistake. The self-directed learning complexity $\mathrm{SDC}(\mathcal{C})$ is the smallest number z for which some self-directed learner exists that makes no more than z mistakes on any concept in $\mathcal{C}$ [Goldman *et al.*, 1993]. In classical online learning [Littlestone, 1988], the sequence of instances x for which the learner makes label predictions is determined by an adversary. The best worst-case number of mistakes achievable in this model, where again the worst case is taken over all concepts in $\mathcal{C}$, is called the optimal mistake bound of $\mathcal{C}$, denoted by $\mathrm{OPT}(\mathcal{C})$.

3 Complexity Results

Table 1 summarizes our complexity results for acyclic CP-nets whose nodes have indegrees bounded by k. The two extreme cases are unbounded acyclic CP-nets ($k = n - 1$, $\mathcal{C}_{unb}$) and separable CP-nets ($k = 0, \mathcal{C}_{sep}$).

The most striking observation from our results is that VCD, RTD, and SDC are equal for all values of m in $\mathcal{C}_{unb}$. Further, when $m = 2$ (the most studied case in the literature), we have that TD equals the instance space size $n2^{n-1}$. A close inspection of the case $m = 2$ shows that $\mathcal{X}_{swap}$ has only n instances that are relevant for $\mathcal{C}_{sep}$, and $\mathcal{C}_{sep}$ corresponds to the class of all concepts over these n instances. Thus the values of VCD, TD, RTD, SDC, and OPT are trivially equal to n in this special case. As mentioned earlier, maximum classes and intersection-closed classes fulfill VCD = RTD, but the class of binary separable CP-nets is the only class to which this result applies, since $\mathcal{C}_{ac}$ is neither maximum nor intersection-closed as soon as $k > 0$ or $m > 2$ (we omit the proofs due to space constraints). The classes $\mathcal{C}_{unb}$ are thus the first natural ones in the literature for which VCD = RTD is known to hold without the fulfillment of any of Doliwa *et al.*'s sufficient conditions like being maximum or intersection-closed. This suggests that the combinatorial structure of CP-nets is interesting from a computational learning theory point of view.

In the case of online prediction, for $m \leq 11$, $\lceil log(m!) \rceil$ is known to be the minimum number of comparisons needed to sort m elements [Sloane, 2016]. However, for most practical applications, $m \leq 11$ is sufficient and thus our results are still useful for judging the optimality of online learning algorithms. The fact that SDC is asymptotically strictly smaller than OPT shows that actively selecting examples strictly decreases the number of mistakes when m is large.

The remainder of this section is dedicated to proving the statements from Table 1. Due to space constraints, some parts of proofs are sketched only. Our first theorem substantially improves on (and corrects) a result by Koriche and Zanuttini (2010), who present a lower bound on $\mathrm{VCD}(\mathcal{C}_{ac})$; their bound is in fact incorrect unless $k \ll n$.

Theorem 1. $\mathrm{VCD}(\mathcal{C}_{unb}) = m^n - 1$, $\mathrm{VCD}(\mathcal{C}_{sep}) = (m-1)n$ *and* $\mathrm{VCD}(\mathcal{C}_{ac}) \geq (m-1)\mathcal{M}$.

As any consistent CP-net (whether acyclic or cyclic) defines an irreflexive, transitive relation, a result from [Booth *et al.*, 2010] implies that the VCD of the class of all consistent unbounded CP-nets is at most $m^n - 1$. By our Theorem 1, this VCD is *equal to* $m^n - 1$. Sauer's Lemma then bounds the number of consistent CP-nets from above by $\sum_{i=0}^{m^n-1} \binom{|\mathcal{X}_{swap}|}{i}$. This also means though, that acyclic CP-

class	VCD	TD	RTD	SDC	OPT
$\mathcal{C}_{ac}$	$\geq (m-1)\mathcal{M}$	$n(m-1)\mathcal{U}$	$(m-1)\mathcal{M}$	$(m-1)\mathcal{M}$	$\geq \lceil log(m!) \rceil \mathcal{M}$
$\mathcal{C}_{unb}$	$m^n - 1$	$n(m-1)m^{n-1}$	$m^n - 1$	$m^n - 1$	$\geq \lceil log(m!) \rceil \frac{m^n-1}{m-1}$
$\mathcal{C}_{sep}$	$(m-1)n$	$(m-1)n$	$(m-1)n$	$(m-1)n$	$\geq \lceil log(m!) \rceil n$

Table 1: Summary of complexity results. $\mathcal{M} = (n-k)m^k + \sum_{i=0}^{k-1} m^i$; $\mathcal{U}$ is defined after Theorem 2.

nets, while less expressive, are in some sense as hard to learn as all consistent CP-nets.

The proof of Theorem 1 relies on decomposing $\mathcal{C}_{ac}$ as a direct product of concept classes over subsets of $\mathcal{X}_{swap}$.

Definition 5. *Let $\mathcal{C}_i \subseteq 2^{\mathcal{X}_i}$ and $\mathcal{C}_j \subseteq 2^{\mathcal{X}_j}$ be concept classes with $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$. The concept class $\mathcal{C}_i \times \mathcal{C}_j \subseteq 2^{\mathcal{X}_i \cup \mathcal{X}_j}$ is defined by $\mathcal{C}_i \times \mathcal{C}_j = \{c_i \cup c_j \mid c_i \in \mathcal{C}_i \text{ and } c_j \in \mathcal{C}_j\}$. For concept classes $\mathcal{C}_1, \ldots, \mathcal{C}_r$, we define $\prod_{i=1}^r \mathcal{C}_i = \mathcal{C}_1 \times \cdots \times \mathcal{C}_r = (\cdots((\mathcal{C}_1 \times \mathcal{C}_2) \times \mathcal{C}_3) \times \cdots \times \mathcal{C}_r)$.*

It is well-known that $\mathrm{VCD}(\prod_{i=1}^{t} \mathcal{C}_i) = \sum_{i=1}^{t} \mathrm{VCD}(\mathcal{C}_i)$.

For any $v_i \in V$ and any $\Gamma \subseteq V \setminus \{v_i\}$, we define $\mathcal{C}^{\Gamma}_{\mathrm{CPT}(v_i)}$ to be the concept class consisting of all preference relations corresponding to some $\mathrm{CPT}(v_i)$ where $Pa(v_i) = \Gamma$ and $|\Gamma| \leq k$; here the instance space is the set of all swap pairs x with $V(x) = v_i$. Now, if we fix the context of v_i by fixing an assignment $\gamma \in \mathcal{O}_\Gamma$ of all variables in Γ, we obtain a concept class $\mathcal{C}^{\Gamma}_{\succ^{v_i}_{\gamma}}$, which corresponds to the set of all preference statements concerning the variable v_i conditioned on the context γ. Its instance space is the set of all swaps x with $V(x) = v_i$ and $x[\Gamma] = \gamma$.

Recall that $V = \{v_1, \ldots, v_n\}$. By S_n we denote the class of all permutations of $\{1, \ldots, n\}$.

Proof of Theorem 1. Lemma 1 (below) states that $\mathcal{C}_{ac}$ equals $\bigcup_{\sigma \in S_n} \prod_{i=1}^{n} \bigcup_{\Gamma \subseteq \{v_{\sigma(1)}, \ldots, v_{\sigma(i-1)}\}, |\Gamma| \leq k} \prod_{\gamma \in \mathcal{O}_\Gamma} \mathcal{C}^{\Gamma}_{\succ^{v_{\sigma(i)}}_{\gamma}}$, which yields the bound $\mathrm{VCD}(\mathcal{C}_{ac}) \geq \max_{\sigma \in S_n} \sum_{i=1}^{n} \max_{\Gamma \subseteq \{v_{\sigma(1)}, \ldots, v_{\sigma(i-1)}\}, |\Gamma| \leq k} \sum_{\gamma \in \mathcal{O}_\Gamma} \mathrm{VCD}(\mathcal{C}^{\Gamma}_{\succ^{v_{\sigma(i)}}_{\gamma}})$. Using $\mathrm{VCD}(\mathcal{C}^{\Gamma}_{\succ^{v_{\sigma(i)}}_{\gamma}}) = m-1$ (see Lemma 2), independent of Γ and γ, one obtains, for any $\sigma \in S_n$,

$$\begin{aligned} \mathrm{VCD}(\mathcal{C}_{ac}) &\geq (m-1) \sum_{i=1}^{n} \max_{\Gamma \subseteq \{v_{\sigma(1)}, \ldots, v_{\sigma(i-1)}\}, |\Gamma| \leq k} |\mathcal{O}_\Gamma| \\ &= (m-1) \sum_{i=1}^{n} \max_{\Gamma \subseteq \{v_{\sigma(1)}, \ldots, v_{\sigma(i-1)}\}, |\Gamma| \leq k} m^{|\Gamma|} \\ &= (m-1)\mathcal{M}\,. \end{aligned}$$

For $k \in \{0, n-1\}$, to verify $\mathrm{VCD}(\mathcal{C}_{ac}) \leq (m-1)\mathcal{M}$, consider any shattered set Y of size $(m-1)\mathcal{M}$. One can show that there is some labeling of Y that agrees with exactly one concept in $\mathcal{C}_{ac}$, so that no set $Y' \supset Y$ is shattered (and thus no set of size $> (m-1)\mathcal{M}$ is shattered) by $\mathcal{C}_{ac}$. ■

Lemma 1.

$$\mathcal{C}_{ac} = \bigcup_{\sigma \in S_n} \prod_{i=1}^{n} \bigcup_{\Gamma \subseteq \{v_{\sigma(1)}, \ldots, v_{\sigma(i-1)}\}, |\Gamma| \leq k} \prod_{\gamma \in \mathcal{O}_\Gamma} \mathcal{C}^{\Gamma}_{\succ^{v_{\sigma(i)}}_{\gamma}}\,.$$

Proof. By definition, for $v \in V$ and $\Gamma \subseteq V \setminus \{v\}$, the class $\mathcal{C}^{\Gamma}_{\mathrm{CPT}(v)}$ equals $\prod_{\gamma \in \mathcal{O}_\Gamma} \mathcal{C}^{\Gamma}_{\succ^{v}_{\gamma}}$. (Any concept representing a preference table for v with $Pa(v) = \Gamma$ corresponds to a union of concepts each of which represents a preference statement over D_v conditioned on some context $\gamma \in \mathcal{O}_\Gamma$.)

Any concept corresponds to choosing a set Γ_v of parent variables of size at most k for each variable v, which means $\mathcal{C}_{ac} \subseteq \prod_{i=1}^{n} \bigcup_{\Gamma \subseteq V \setminus \{v_i\}, |\Gamma| \leq k} \mathcal{C}^{\Gamma}_{\mathrm{CPT}(v_i)}$. By acyclicity, $v_j \in Pa(v_i)$ implies $v_i \notin Pa(v_j)$, so that for each concept $c \in \mathcal{C}_{ac}$ some $\sigma \in S_n$ fulfills $c \in \prod_{i=1}^{n} \bigcup_{\Gamma \subseteq \{v_{\sigma(1)}, \ldots, v_{\sigma(i-1)}\}, |\Gamma| \leq k} \mathcal{C}^{\Gamma}_{\mathrm{CPT}(v_{\sigma(i)})}$. Thus, $\mathcal{C}_{ac} \subseteq \bigcup_{\sigma \in S_n} \prod_{i=1}^{n} \bigcup_{\Gamma \subseteq \{v_{\sigma(1)}, \ldots, v_{\sigma(i-1)}\}, |\Gamma| \leq k} \prod_{\gamma \in \mathcal{O}_\Gamma} \mathcal{C}^{\Gamma}_{\succ^{v_{\sigma(i)}}_{\gamma}}$.

Similarly, one can argue that every concept in the class on the right hand side represents an acyclic CP-net with parent sets of size at most k. With $\mathcal{C}^{\Gamma}_{\mathrm{CPT}(v_i)} = \prod_{\gamma \in \mathcal{O}_\Gamma} \mathcal{C}^{\Gamma}_{\succ^{v_i}_{\gamma}}$, the statement of the lemma follows. ■

Lemma 2. $\mathrm{VCD}(\mathcal{C}^{\Gamma}_{\succ^{v_i}_{\gamma}}) = m - 1$ *for any $v_i \in V$, $\Gamma \subseteq V \setminus \{v_i\}$, and $\gamma \in \mathcal{O}_\Gamma$.*

Proof. Let $v_i \in V$, $\Gamma \subseteq V \setminus \{v_i\}$, and $\gamma \in \mathcal{O}_\Gamma$. We show that $\mathcal{C}^{\Gamma}_{\succ^{v_i}_{\gamma}}$ shatters some set of size $m-1$, but no set of size m. Note first that, by definition, $\mathcal{C}^{\Gamma}_{\succ^{v_i}_{\gamma}}$ is simply the class of all total orders over the domain D_{v_i} of v_i.

To show $\mathrm{VCD}(\mathcal{C}^{\Gamma}_{\succ^{v_i}_{\gamma}}) \geq m-1$, choose any set of $m-1$ swaps over $\Gamma \cup \{v_i\}$ with fixed context γ, in which the pairs of swapped values in v_i are $(v^i_1, v^i_2), \ldots, (v^i_{m-1}, v^i_m)$.

Fix any set $S \subseteq \mathcal{X}_{swap}$ of m swaps over $\Gamma \cup \{v_i\}$ with fixed context γ. To show that S is not shattered, consider the undirected graph G with vertex set D_{v_i} in which an edge between v^i_r and v^i_s exists iff S contains a swap pair flipping v^i_r to v^i_s or vice versa. G has m vertices and m edges and thus contains a cycle. The directed versions of G correspond to the labellings of S; therefore some labelling ℓ of S corresponds to a cyclic directed version of G, which does not induce a total order over D_{v_i}. Hence the labelling ℓ is not realized by $\mathcal{C}^{\Gamma}_{\succ^{v_i}_{\gamma}}$, so that S is not shattered by $\mathcal{C}^{\Gamma}_{\succ^{v_i}_{\gamma}}$. ■

For studying teaching complexity, it is useful to identify concepts that are "easy to teach." To this end, we use the notion of subsumption [Koriche and Zanuttini, 2010]: given CP-nets N, N', we say N subsumes N' if for all $v_i \in V$ the following holds: If $y_1 \succ y_2$ is specified in $\mathrm{CPT}(v_i)$ in N' for some context γ', then $y_1 \succ y_2$ is specified in $\mathrm{CPT}(v_i)$ in N for some context containing γ'. If in addition $N \neq N'$, we say that N strictly subsumes N'.

Now let $\mathcal{C} \subseteq \mathcal{C}_{ac}$. A concept $c \in \mathcal{C}$ is maximal in $\mathcal{C}$ if no $c' \in \mathcal{C}$ strictly subsumes c. Note that maximal concepts in $\mathcal{C}_{ac}$

are of size $\mathcal{M}$. The following lemma formalizes the intuition that maximal concepts are "easy to teach."

Lemma 3. *For any maximal concept c in a concept class $\mathcal{C} \subseteq \mathcal{C}_{ac}$, we have* $\mathrm{TD}(c, \mathcal{C}) \leq (m-1)size(c)$.

Proof. Every statement in the CP-net N represented by c corresponds to an order of m values for some variable v_i under a fixed context γ. For every such order $y_1 \succ_\gamma^{v_i} \ldots \succ_\gamma^{v_i} y_m$, we include $m-1$ positively labelled swap examples in a set T. For $1 \leq j \leq m-1$, the jth such example labels a pair $x = (x.1, x.2)$ of swap outcomes with $V(x) = v_i$, the projection of x onto $\{v_i\}$ is (y_j, y_{j+1}), and the projection of x onto the remaining variables contains γ. The set T then has cardinality $(m-1)size(c)$ and is obviously consistent with N.

It remains to show that no other CP-net in $\mathcal{C}$ is consistent with T. Suppose some $c' \neq c$ in $\mathcal{C}$ is consistent with T. Since $c' \neq c$, there is some $v_i \in V$, $\gamma \in \mathcal{O}_{V \setminus \{v_i\}}$, and $y, y' \in D_{v_i}$ such that $y \succ_\gamma^{v_i} y'$ holds in c' while $y' \succ_\gamma^{v_i} y$ holds in c. Thus (i) c' disagrees with some statement in a preference table of c, or (ii) c' has a statement in one of its preference tables that is not contained in c. (i) is impossible since c' is consistent with T, and (ii) is impossible since c is maximal in $\mathcal{C}$. ∎

Lemma 4. *Each non-maximal $c' \in \mathcal{C}_{ac}$ is strictly subsumed by some $c \in \mathcal{C}_{ac}$ s.t.* $\mathrm{TD}(c', \mathcal{C}_{ac}) \geq \mathrm{TD}(c, \mathcal{C}_{ac})$.

Proof. From the graph G' for c', we build a graph G by adding the maximum possible number of edges to a single variable v. As c' is not maximal, it is possible to add at least one edge. The CP-nets corresponding to G and G' differ only in $\mathrm{CPT}(v)$. Let c be the concept representing G and z be the size of its CPT for v. A smallest teaching set T' for c' can be modified to a teaching set for c by replacing only those examples that refer to the swapped variable v; $(m-1)z$ examples suffice. To distinguish c' from c, T' must contain at least $(m-1)z$ examples referring to the swapped variable v ($m-1$ for each context in $\mathrm{CPT}(v)$ in c). Hence $\mathrm{TD}(c', \mathcal{C}_{ac}) \geq \mathrm{TD}(c, \mathcal{C}_{ac})$. ∎

Using these lemmas, one can show that $\mathrm{TD}(\mathcal{C}_{ac})$ equals the TD of separable CP-nets within $\mathcal{C}_{ac}$ and $\mathrm{RTD}(\mathcal{C}_{ac})$ is the TD of maximal concepts within $\mathcal{C}_{ac}$. The latter is at most $(m-1)\mathcal{M}$ by Lemma 3, and can be verified to be at least $(m-1)\mathcal{M}$ when arguing that a teaching set for a maximal concept must contain $m-1$ examples for each statement in its CPTs, so as to determine the preferences for each context. We thus obtain the following theorem.

Theorem 2. $\mathrm{RTD}(\mathcal{C}_{ac}) = (m-1)\mathcal{M}$.

For the TD of separables in $\mathcal{C}_{ac}$, consider any unconditional $\mathrm{CPT}(v_i) = \{y_1 \succ \cdots \succ y_m\}$. For every $R \subseteq V \setminus \{v_i\}$, $|R| = k$, we create the dummy $\mathrm{CPT}(v_i)$ where $Pa(v_i) = R$ with the same statement $y_1 \succ \cdots \succ y_m$ in every context of $\mathcal{O}_R$. Any teaching set must show that under any context, we have the same statement $y_1 \succ \cdots \succ y_m$.

Thus, a minimal teaching set restricted to $\mathrm{CPT}(v_i)$ is a smallest set of examples $\mathcal{U}_i$ such that if projected to any subset R of size k, $\mathcal{U}_i$ contains m^k contexts. Each of the statements of the form $y_1 \succ \cdots \succ y_m$ can be taught by $(m-1)$ labelled examples, and one does so for each element of $\mathcal{U}_i$. For each variable v_i, a respective set of examples is included in the teaching set. One can show that fewer examples are not sufficient for teaching a separable CP-net. We denote the cardinality of $\mathcal{U}_i$, which is independent of i, by $\mathcal{U}$. In the binary case, $\mathcal{U}_i$ is known as a "$(n-1, k)$-universal set of minimum size" [Damaschke, 2000]. In combination with some obvious bounds, we obtain the following theorem.

Theorem 3. *For $0 \leq k \leq n-1$, we have $n(m-1)m^k \leq \mathrm{TD}(\mathcal{C}_{ac}) = n(m-1)\mathcal{U} \leq n(m-1)\binom{n-1}{k}m^k$. If $k=0$, then $\mathcal{U} = 1$, so that $\mathrm{TD}(\mathcal{C}_{sep}) = (m-1)n$. If $k=1$, then $\mathcal{U} = m$, so that $\mathrm{TD}(\mathcal{C}_{tree}) = (m-1)mn$. If $k = n-1$, then $\mathcal{U} = m^{n-1}$, so that $\mathrm{TD}(\mathcal{C}_{unb}) = (m-1)nm^{n-1}$.*

Theorem 3 implies that, for $\mathcal{C}_{unb}$, the ratio of TD over instance space size $|\mathcal{X}_{swap}|$ is $\frac{2}{m}$. In particular, in the case of binary CP-nets (i.e., when $m = 2$), which is the focus of most of the literature on learning CP-nets, the TD equals the instance space size. However, maximal concepts have a TD far below the worst-case TD.

Theorem 4. $\mathrm{SDC}(\mathcal{C}_{ac}) = (m-1)\mathcal{M}$.

Proof. From [Doliwa *et al.*, 2014] and Theorem 2 we get $\mathrm{SDC}(\mathcal{C}_{ac}) \geq \mathrm{RTD}(\mathcal{C}_{ac}) = (m-1)\mathcal{M}$. For the upper bound, note that any concept in $\mathcal{C}_{ac}$, when fixing a variable v and a context $\gamma \in \mathcal{O}_{V \setminus \{v\}}$, induces a total order on D_v. Goldman *et al.* (1993) discussed a prediction strategy (basically an insertion sort) to learn a total order over m items while making at most $m-1$ mistakes. Separately for each variable v, a self-directed learner fixes an arbitrary context $\gamma \in \mathcal{O}_{V \setminus \{v\}}$ and learns a preference over D_v with Goldman *et al.*' s strategy. For other contexts on v, the learner will assume the same preference relation unless it makes a mistake (which will cause it to learn a new preference over D_v). For each preference to be learned ($\leq \mathcal{M}$ in total), the learner makes at most $(m-1)$ mistakes, for a total of $\leq (m-1)\mathcal{M}$ mistakes. ∎

Theorem 5. $\mathrm{OPT}(\mathcal{C}_{ac}) \geq \lceil log(m!) \rceil \mathcal{M}$.

Proof. Any learner must identify up to $\mathcal{M}$ preference statements (for a fixed variable and context) separately. Each such statement is a permutation of m elements, and its identification requires at least $\lceil log(m!) \rceil$ comparisons, in the worst case. The adversary can force the learner to make as many mistakes as comparisons are needed, yielding the lower bound. ∎

4 Near-Optimal Query Learning

In what follows, we will demonstrate in two scenarios how our complexity results can be used to assess the optimality of learning algorithms, in particular, algorithms learning a target CP-net N^* from membership or equivalence queries. A membership (equivalence) query is a swap x (a CP-net N', resp.) that the learner passes to an oracle; the oracle responds 'yes' or 'no', depending on whether N^* entails x (whether N^* equals N', resp.) If an equivalence query for N' is answered 'no', the oracle also presents a swap that is entailed by either N^* or N', but not by both. An algorithm learns a class $\mathcal{C}$ from a given type of queries if it identifies any target concept $c^* \in \mathcal{C}$ from a number of queries that is polynomial in the size of the underlying representation of c [Angluin, 1988].

4.1 Learning Tree-Structured CP-nets from Membership Queries

Koriche and Zanuttini (2010) present an algorithm for learning a binary tree-structured CP-net N^* that may be *incomplete* in that it may have empty CPTs (i.e., it learns a superclass of $\mathcal{C}_{tree}$ for $m = 2$.) Their learner uses at most $n_{N^*} + 1$ equivalence and $4n_{N^*} + e_{N^*}\log(n)$ membership queries, where n_{N^*} is the number of relevant variables and e_{N^*} the number of edges in N^*. We present a method learning any CP-net in $\mathcal{C}_{tree}$ (i.e., *complete* tree CP-nets) for any m, using only membership queries.

For a CP-net N, a conflict pair w.r.t. v_i is a pair (x, x') of swaps such that (i) $V(x) = V(x') = v_i$, (ii) $x.1$ and $x'.1$ agree on v_i, (iii) $x.2$ and $x'.2$ agree on v_i, and (iv) N entails one of the swaps x, x', but not the other. If v_i has a conflict pair, then v_i has a parent variable v_j whose values in x and x' are different. Such a variable v_j can be found with $\log(n)$ membership queries by binary search (each query halves the number of candidate variables with different values in x and x') [Damaschke, 2000].

We use this binary search to learn tree-structured CP-nets from membership queries, by exploiting the following fact: if a variable v_i in a tree CP-net has a parent, then a conflict pair w.r.t. v_i exists and can be detected by asking membership queries to sort m "test sets" for v_i. Let $(v_1^i, \dots, v_m^i)$ an arbitrary but fixed permutation of D_{v_i}. Then, for all $j \in \{1, \dots, m\}$, a test set $I_{i,j}$ for v_i is defined by $I_{i,j} = \{(v_j^1, \dots, v_j^{i-1}, v_r^i, v_j^{i+1}, \dots, v_j^n) \mid 1 \leq r \leq m\}$. Since v_i has no more than one parent, determining preference orders over m such test sets of size m is sufficient for revealing conflict pairs, rather than having to test all possible contexts in $\mathcal{O}_{V \setminus \{v_i\}}$.

Now we have a simple strategy for learning tree CP-nets with membership queries:

1. For every variable v_i, ask $O(m\log(m))$ membership queries from the $\binom{m}{2}$ swaps over $I_{i,j}$ to obtain orders over $I_{i,j}$ for all j.
2. If at least two of the orders differ, i.e., there is a conflict pair, find the only parent of v_i by $log(n)$ further queries, else $Pa(v_i) = \emptyset$.
3. $\mathrm{CPT}(v_i)$ is determined by the queries above.

It is not hard to see that this algorithm learns a target CP-net $N^* \in \mathcal{C}_{tree}$ with at most $nmO(m\log(m)) + e_{N^*}\log(n)$ membership queries, where e_{N^*} is the number of edges in N^*. In particular, for the binary case we need $2n + e_{N^*}\log(n)$ queries at most, i.e., compared to Koriche and Zanuttini's method, when focusing only on tree CP-nets with non-empty CPTs, we reduce the number of membership queries by a factor of 2, and drop the requirement for equivalence queries. As TD lower-bounds the required number of membership queries, our method uses at most $\log(m) + e_{N^*}\log(n)$ queries more than an optimal one, see Theorem 3.

4.2 Learning Incomplete and Complete Acyclic CP-nets from Queries

Koriche and Zanuttini (2010) provide an algorithm that learns the class $\mathcal{C}_{ac}^*$ of complete and incomplete (binary) acyclic CP-nets with nodes of bounded indegree from equivalence and membership queries. To evaluate their algorithm, they compare its query complexity to $\log(4/3)\mathrm{VCD}(\mathcal{C}_{ac}^*)$, which lower-bounds the required number of membership and equivalence queries [Auer and Long, 1999]. They calculate a lower bound on $\mathrm{VCD}(\mathcal{C}_{ac}^*)$, in lieu of the exact value, cf. their Theorem 6.

$\mathcal{C}_{ac}^* \supseteq \mathcal{C}_{ac}$ implies $\mathrm{VCD}(\mathcal{C}_{ac}^*) \geq \mathrm{VCD}(\mathcal{C}_{ac})$. It is not hard to see that in fact $\mathrm{VCD}(\mathcal{C}_{ac}^*) = \mathrm{VCD}(\mathcal{C}_{ac}) \geq \mathcal{M}$ (for $m = 2$), since the largest shattering capacity is obtained when specifying CPTs for all variables, i.e., when using complete CP-nets. However, Koriche and Zanuttini's lower bound on $\mathrm{VCD}(\mathcal{C}_{ac}^*)$ exceeds $\mathrm{VCD}(\mathcal{C}_{unb}^*) = 2^n - 1$ in some cases. In particular, in case $k = n - 1$, their lower bound (on $\mathrm{VCD}(\mathcal{C}_{unb}^*)$) exceeds the exact value of $2^n - 1$, cf. our Theorem 1. It appears as though their bound is correct only for $k \ll n$ as it is evaluated to $(n+(c-1))2^{n-(c+1)} - n^2 - (c-1)n + \frac{(2c-1)(n+(c-1))}{2}$ which exceeds $\mathrm{VCD}(\mathcal{C}_{unb}^*)$ for small values of c where $k = n - c$.

For any k, their algorithm learns a superclass $\mathcal{C}_{ac}^*$ of $\mathcal{C}_{ac}$ for $m = 2$ (where the additional concepts are the incomplete ones) using at most $r_{N^*} + e_{N^*}\log(n) + e_{N^*} + 1$ queries in total, for a target CP-net N^* with r_{N^*} statements and e_{N^*} edges. In the worst case $r_{N^*} = \mathcal{M} \leq \mathrm{VCD}(\mathcal{C}_{ac}^*)$ and $e_{N^*} = \binom{k}{2} + (n-k)k$ (i.e., N^* is maximal w.r.t. $\mathcal{C}_{ac}^*$). This yields $\mathcal{M} + e_{N^*}(\log(n) + 1)$ queries for their algorithm, which exceeds the lower bound $\log(4/3)\mathrm{VCD}(\mathcal{C}_{ac}^*)$ by at most $\log(3/2)\mathrm{VCD}(\mathcal{C}_{ac}^*) + e_{N^*}\log(n)$. This is a more refined assessment compared to the reported term $e_{N^*}\log(n)$, and it holds for *any* value of k.

5 Conclusions

We determined exact values or non-trivial bounds on the parameters VCD, TD, RTD, SDC, and OPT for the classes of complete k-bounded acyclic CP-nets for any k. Our VCD values still apply to the class of potentially incomplete k-bounded acyclic CP-nets, and thus correct a mistake in [Koriche and Zanuttini, 2010]. Our TD value shows that our proposed algorithm for learning complete tree CP-nets is close to optimal.

To the best of our knowledge, $\mathcal{C}_{unb}$ is the first known non-maximum (and not intersection-closed) class that is interesting from an application point of view and satisfies RTD = VCD. Thus further studies on the structure of CP-nets may be helpful toward the solution of an open problem concerning the general relationship between RTD and VCD [Simon and Zilles, 2015].

Our results may also have implications on the study of consistent CP-nets. Since the class of acyclic CP-nets is less expressive than that of all consistent CP-nets, while having the same information complexity in terms of VCD, it would be interesting to find out whether learning algorithms for acyclic CP-nets can be easily adapted to consistent CP-nets in general.

Acknowledgements

Eisa Alanazi's research was supported by the Saudi Cultural Bureau in Canada. Malek Mouhoub and Sandra Zilles were supported by the Discovery Grant program of the Natural Sciences and Engineering Research Council of Canada (NSERC). Sandra Zilles was also supported by the NSERC Canada Research Chairs program.

References

[Angluin, 1988] Dana Angluin. Queries and concept learning. *Machine Learning*, 2(4):319–342, 1988.

[Auer and Long, 1999] Peter Auer and Phil Long. Structural results about on-line learning models with and without queries. *Machine Learning*, 36(3):147–181, 1999.

[Blumer *et al.*, 1989] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36(4):929–965, 1989.

[Booth *et al.*, 2010] Richard Booth, Yann Chevaleyre, Jérôme Lang, Jérôme Mengin, and Chattrakul Sombattheera. Learning conditionally lexicographic preference relations. In *ECAI*, pages 269–274, 2010.

[Boutilier *et al.*, 2004] Craig Boutilier, Ronen I. Brafman, Carmel Domshlak, Holger H. Hoos, and David Poole. CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2004.

[Chevaleyre *et al.*, 2010] Yann Chevaleyre, Frédéric Koriche, Jérôme Lang, Jérôme Megine, and Bruno Zanuttini. Learning ordinal preferences on multiattribute domains: the case of CP-nets. In *Preference Learning*, pages 273–296. Springer-Verlag, 2010.

[Damaschke, 2000] Peter Damaschke. Adaptive versus non-adaptive attribute-efficient learning. *Machine Learning*, 41(2):197–215, 2000.

[Dimopoulos *et al.*, 2009] Yannis Dimopoulos, Loizos Michael, and Fani Athienitou. Ceteris paribus preference elicitation with predictive guarantees. In *IJCAI*, pages 1890–1895, 2009.

[Doliwa *et al.*, 2014] Thorsten Doliwa, Gaojian Fan, Hans Ulrich Simon, and Sandra Zilles. Recursive teaching dimension, VC-dimension and sample compression. *Journal of Machine Learning Research*, 15:3107–3131, 2014.

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

[Goldman *et al.*, 1993] Sally A. Goldman, Ronald L. Rivest, and Robert E. Schapire. Learning Binary Relations and Total Orders. *SIAM Journal on Computing*, 22(5):1006–1034, 1993.

[Guerin *et al.*, 2013] Joshua T. Guerin, Thomas E. Allen, and Judy Goldsmith. Learning CP-net preferences online from user queries. In *ADT*, pages 208–220, 2013.

[Koriche and Zanuttini, 2010] Frédéric Koriche and Bruno Zanuttini. Learning conditional preference networks. *Artificial Intelligence*, 174(11):685–703, 2010.

[Lang and Mengin, 2009] Jérôme Lang and Jérôme Mengin. The complexity of learning separable ceteris paribus preferences. In *IJCAI*, pages 848–853, 2009.

[Littlestone, 1988] Nick Littlestone. Learning quickly when irrelevant attributes abound: a new linear threshold algorithm. *Machine Learning*, 2(4):245–318, 1988.

[Liu *et al.*, 2014] Juntao Liu, Yi Xiong, Caihua Wu, Zhijun Yao, and Wenyu Liu. Learning conditional preference networks from inconsistent examples. *IEEE Trans. Knowl. Data Eng.*, 26(2):376–390, 2014.

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

[Shinohara and Miyano, 1991] Ayumi Shinohara and Satoru Miyano. Teachability in computational learning. *New Generation Computing*, 8(4):337–347, 1991.

[Simon and Zilles, 2015] Hans Ulrich Simon and Sandra Zilles. Open problem: Recursive teaching dimension versus VC dimension. In *COLT*, pages 1770–1772, 2015.

[Sloane, 2016] Neil J.A. Sloane. The on-line encyclopedia of integer sequences. Sequence A036604, 2016. Accessed: 10-April-2016.

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

[Zilles *et al.*, 2011] Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Models of cooperative teaching and learning. *Journal of Machine Learning Research*, 12:349–384, 2011.