Tournaments, Johnson Graphs and NC-Teaching

Hans Ulrich Simon   HSIMON@MPI-INF.MPG.DE

Max Planck Institute for Informatics, Saarbrücken,
and Ruhr-University Bochum

Editors: Shipra Agrawal and Francesco Orabona

Abstract

Some years ago a teaching model, called "No-Clash Teaching" or simply "NC-Teaching", had been suggested that is provably optimal in the following strong sense. First, it satisfies Goldman and Matthias' collusion-freeness condition. Second, the NC-teaching dimension (= NCTD) is smaller than or equal to the teaching dimension with respect to any other collusion-free teaching model. Specifically the NCTD is upper-bounded by the recursive teaching dimension (= RTD). This raised the question about the largest possible gap between the NCTD and the RTD. The main results in this paper are as follows. First, we show that there exists a family $(\mathcal{C})_{n \geq 1}$ of concept classes such that the RTD of $\mathcal{C}$ grows logarithmically in $n = |\mathcal{C}_n|$ while, for every $n \geq 1$, the NCTD of $\mathcal{C}$ equals 1. Since the RTD of a finite concept class $\mathcal{C}$ is generally bounded by $\log |\mathcal{C}|$, the family $(\mathcal{C}_n)_{n \geq 1}$ separates RTD from NCTD in the most striking way. Our first proof of existence of the family $(\mathcal{C}_n)_{n \geq 1}$ makes use of the probabilistic method and random tournaments. But we also present a concrete family of concept classes (leading to a slightly smaller lower bound on $\mathrm{RTD}(\mathcal{C}_n)$) which makes use of so-called quadratic-residue tournaments. Second, we characterize the maximum concept classes of NCTD 1 as classes which are induced by tournaments in a very natural way. Third, we improve the previously best upper bound on the size of a maximum class of NCTD d by a factor of order $\sqrt{d}$. The verification of the new upper bound makes use of Johnson graphs and maximum subgraphs not containing large narrow cliques. The connections between tournaments, Johnson graphs and NC-Teaching revealed here were not known before and might be considered interesting in their own right.

Keywords: no-clash teaching, recursive teaching, tournaments, maximum classes, Johnson graphs

1. Introduction

Learning from examples that were carefully chosen by a teacher (e.g. a human expert) presents an alternative to the commonly used model of learning from randomly chosen examples. A model of teaching should be sufficiently restrictive to rule out collusion between the learner and the teacher. For instance, the teacher should not be able to encode a direct representation of the target concept (such as a Boolean formula or a neural network) within the chosen sequence of examples. Goldman and Mathias (1996) suggested to consider a learner-teacher pair as collusion-free if it satisfies the following condition: if the learner is in favor of concept C after having seen the labeled teaching set $\mathcal{T}$ chosen by the teacher, the learner should again be in favor of C after having seen a superset $\mathcal{S}$ of $\mathcal{T}$ as long as the label assignment in $\mathcal{S}$ still coincides with the label assignment induced by C. In other words: the learners guess C for the target concept should not be altered when the data give even more support to C than the original labeled teaching set $\mathcal{T}$ is giving. Most existing abstract models of teaching are collusion-free in this sense. Kirkpatrick et al. (2019a) introduced a new model, called no-clash teaching or simply NC-teaching, that is collusion-free and furthermore optimal in the following strong sense:

For any model M, let M-TD$(\mathcal{C})$ denote the corresponding teaching dimension of concept class $\mathcal{C}$ (= smallest number that upper-bounds the size of any of the employed teaching sets provided that learner and teacher interact as prescribed by model M). Then $\text{NCTD}(\mathcal{C}) \leq \text{M-TD}(\mathcal{C})$ holds for any model M that satisfies Goldman and Mathias' collusion-freeness criterion. Specifically the NCTD is upper-bounded by the RTD. This raised the question about the largest possible gap between the NCTD and the RTD. The largest gap that was known prior to this work results from the class of parity functions in n Boolean variables: as shown by Falat et al. (2023), the RTD of the parity class equals n while the NCTD of this class is bounded by $n/4$.

The main results in this paper are as follows:

- We show non-constructively (making use of the probabilistic method and random tournaments) that there exists a family $(\mathcal{C}_n)_{n \geq 1}$ of concept classes such that the RTD of $\mathcal{C}_n$ grows logarithmically in $n = |\mathcal{C}_n|$ while, for every $n \geq 1$, the NCTD of $\mathcal{C}_n$ equals 1. Since the RTD of a finite concept class $\mathcal{C}$ is generally bounded by $\log |\mathcal{C}|$, the family $(\mathcal{C}_n)_{n \geq 1}$ separates RTD from NCTD in the most striking way. See Section 3 for details.

- We present a concrete family of concept classes that leads to a slightly smaller lower bound on $\text{RTD}(\mathcal{C}_n)$ while the NCTD of $\mathcal{C}_n$ still equals 1. The construction of this family is based on so-called quadratic-residue tournaments. See Section 4 for details.

- We characterize the maximum concept classes of NCTD 1 as classes which are induced by tournaments in a very natural way. See Section 5.1 for details.

- We improve the best previously known upper bound on the size of maximum classes of NCTD d by a factor of order $\sqrt{d}$. The verification of the new upper bound makes use of Johnson graphs and maximum subgraphs not containing large narrow cliques. See Section 5.2 and the appendix for details.

Section 2 calls to mind the definition of the teaching models and teaching dimensions that we will later deal with. All other formal definitions are given on the way right before they are actually needed.

2. Teaching Models

As usual a *concept over domain* $\mathcal{X}$ is a function from $\mathcal{X}$ to $\{0, 1\}$ or, equivalently, a subset of $\mathcal{X}$. The powerset of $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. The set of all subsets of size d of $\mathcal{X}$ is denoted by $\mathcal{P}_d(\mathcal{X})$. A set $\mathcal{C} \subseteq \mathcal{P}(\mathcal{X})$ whose elements are concepts over domain $\mathcal{X}$ is referred to as a *concept class over* $\mathcal{X}$. The elements of $\mathcal{X}$ are called *instances*. A *(labeled) sample* is a pair (S, b) such that $S \subseteq \mathcal{X}$ and $b : S \to \{0, 1\}$. The sample (S, b) is *realizable by* $\mathcal{C}$ if there exists a concept $C \in \mathcal{C}$ which satisfies

$$\forall s \in S : C(s) = b(s) \ . \tag{1}$$

We will refer to a sample (S, b) with $|S| = k$ as a *k-sample*.

General Assumption: Throughout the paper, the domain $\mathcal{X}$ is assumed to be finite.

By the following definitions, we call to mind some of the most popular teaching models:

Definition 1 (Basic Teaching Model (Shinohara and Miyano, 1991; Goldman and Kearns, 1995))
Let $\mathcal{C}$ be a concept class over $\mathcal{X}$. A teaching set for $C \in \mathcal{C}$ is a subset $D \subseteq \mathcal{X}$ which distinguishes C from any other concept in $\mathcal{C}$, i.e., for every $C' \in \mathcal{C} \setminus \{C\}$, there exists some $x \in D$ such that $C(x) \neq C'(x)$. The size of the smallest teaching set for $C \in \mathcal{C}$ is denoted by $\mathrm{TD}(C, \mathcal{C})$. The teaching dimension of $\mathcal{C}$ in the basic model of teaching is then given by

$$\mathrm{TD}(\mathcal{C}) = \max_{C \in \mathcal{C}} |T(C, \mathcal{C})| .$$

A related quantity is

$$\mathrm{TD}_{min}(\mathcal{C}) = \min_{C \in \mathcal{C}} |T(C, \mathcal{C})| .$$

Note that the set $D \subseteq \mathcal{X}$ in this definition is an *unlabeled* teaching set. In the teacher-learner interaction, the learner would get to know the corresponding *labeled* teaching set. However we do not need to bother the reader with the teacher-learner interaction in order to define $\mathrm{TD}(C, \mathcal{C})$ and $\mathrm{TD}(\mathcal{C})$. An analogous remark applies to the two teaching models that we describe next.

Definition 2 (NC-Teaching Model (Kirkpatrick et al., 2019a)) *Let $T : \mathcal{C} \to \mathcal{P}(\mathcal{X})$ be a mapping that assigns to every concept in $\mathcal{C}$ a set of instances. T is called* admissible *for $\mathcal{C}$ in the NC-model[1] of teaching, or simply an* NC-teacher *for $\mathcal{C}$ if, for every $C \neq C' \in \mathcal{C}$, there exists $x \in T(C) \cup T(C')$ such that $C(x) \neq C'(x)$. The teaching dimension of $\mathcal{C}$ in the NC-model of teaching is given by*

$$\mathrm{NCTD}(\mathcal{C}) = \min\{\max_{C \in \mathcal{C}} |T(C)| : T \text{ is an NC-teacher for } \mathcal{C}\} .$$

We say that two concepts C and C' *clash* (with respect to $T : \mathcal{C} \to \mathcal{P}(\mathcal{X})$) if they agree on $T(C) \cup T(C')$, i.e, if they assign the same $0, 1$-label to each instance in $T(C) \cup T(C')$. NC-teachers for $\mathcal{C}$ are teachers who avoid clashes between any pair of distinct concepts from $\mathcal{C}$.

Definition 3 (R-Teaching Model (Zilles et al., 2011)) *Let $\mathcal{C}_{min} \subseteq \mathcal{C}$ be the easiest-to-teach concepts in $\mathcal{C}$, i.e.,*

$$\mathcal{C}_{min} = \{C \in \mathcal{C} : \mathrm{TD}(C, \mathcal{C}) = \mathrm{TD}_{min}(\mathcal{C})\} .$$

The recursive teaching dimension *of $\mathcal{C}$ is then given by*

$$\mathrm{RTD}(\mathcal{C}) = \begin{cases} \mathrm{TD}_{min}(\mathcal{C}) & \textit{if } \mathcal{C} = \mathcal{C}_{min} \\ \max\{\mathrm{TD}_{min}(\mathcal{C}), \mathrm{RTD}(\mathcal{C} \setminus \mathcal{C}_{min})\} & \textit{otherwise} \end{cases} .$$

It was shown by Doliwa et al. (2014) that

$$\mathrm{RTD}(\mathcal{C}) = \max_{\mathcal{C}' \subseteq \mathcal{C}} \mathrm{TD}_{min}(\mathcal{C}') \geq \mathrm{TD}_{min}(\mathcal{C}) .$$

We will refer to $\frac{\mathrm{RTD}(\mathcal{C})}{\mathrm{NCTD}(\mathcal{C})}$ as the *RTD-NCTD ratio* of the concept class $\mathcal{C}$. Because the RTD is lower-bounded by TD_{min}, any lower bound on $\frac{\mathrm{TD}_{min}(\mathcal{C})}{\mathrm{NCTD}(\mathcal{C})}$ is also a lower bound on the RTD-NCTD ratio of $\mathcal{C}$.
Some remarks are in order here:

---

1. NC = No-Clash.

- Modelling a teacher T as a mapping $C \mapsto T(C) \subseteq \mathcal{X}$ reflects that we think of T as to assigning helpful examples to a given concept. This does not necessarily imply that T knows a representation of C in a specific class of representations (like, for instance, knowing the architecture and the weight parameters of a neural network which represents C).

- The NC-model of teaching is of obvious importance because, as explained earlier already, $\mathrm{NCTD}(\mathcal{C})$ lower-bounds $\mathrm{M}\text{-}\mathrm{TD}(\mathcal{C})$ for any model M that satisfies Goldman and Mathias' collusion-freeness condition.

- The R-model of teaching can be considered important for various reasons, including the following ones. First, the parameter $\mathrm{RTD}(\mathcal{C})$ is related to the size of some specific compression schemes for $\mathcal{C}$. Second, teaching concepts of a finite class $\mathcal{C}$ in the R-model is equivalent to teaching concepts of $\mathcal{C}$ in the so-called *preference-based model* (where the teacher and the learner share a preference relation and the learner's guess for the concept C is the most preferred concept in $\mathcal{C}$ among all concepts that are consistent with $T(C)$) We cannot go into further detail here, but the interested reader is referred to (Doliwa et al., 2014; Darnstädt et al., 2013; Gao et al., 2017).

3. Existence of Concept Classes with Large RTD-NCTD Ratio

In this section, we use the probabilistic method and random tournaments for showing that the NC-Teaching Model is much more powerful than the R-Teaching Model. While the application of the probabilistic method will be simple and (more or less) straightforward, the main innovation should be seen in bringing concept classes induced by tournaments into play. It turns out that these classes are trivial-to teach in the NC-Teaching Model and, up to an asymptotically vanishing fraction, hard-to-teach in the R-Teaching Model.

Recall from graph theory that a *tournament or order* n is a complete oriented graph with n vertices. The following notion will play a central role in this and the subsequent section:

Definition 4 (Concept Class Induced by a Tournament) *Let* $G = ([n], E)$ *be a* tournament of order n. *We define* $\mathcal{C}(G) = \{C_1, \ldots, C_n\}$ *where, for* $i = 1, \ldots, n$, *we set*

$$C_i = \{j \in [n] : (i,j) \in E\} \ . \tag{2}$$

We will refer to $\mathcal{C}(G)$ *as the* the concept class induced by G.

Intuitively, we may think of C_i as consisting of all players j who have lost against player i in the tournament G.

Example 1 *Consider the tournament* G_n *with vertices* $1, \ldots, n$ *and with edges* (i, j) *for all* $1 \leq i < j \leq n$ *(i.e., edges are always directed from smaller to larger numbers). Then the concepts in* $\mathcal{C}(G_n)$ *are* $C_i = \{i + 1, \ldots, n\}$ *for* $i = 1, \ldots, n$. *In other words,* $\mathcal{C}(G_n)$ *contains all right half-intervals over domain* $[n]$ *(including* $C_n = \emptyset$ *but excluding* $[n]$*).*

If G is a tournament of order 1, then $\mathcal{C}(G) = \{\emptyset\}$ and $\mathrm{NCTD}(\mathcal{C}(G)) = 0$. For tournaments of order at least 2, we have the following result:

Lemma 5 *Suppose that G is a tournament of order at least 2. Then $\mathrm{NCTD}(\mathcal{C}(G)) = 1$.*

Proof A concept class $\mathcal{C}(G)$ induced by a tournament $G = ([n], E)$ with $n \geq 2$ contains at least two distinct concepts. Hence $\mathrm{NCTD}(\mathcal{C}(G)) \geq 1$.
Consider the mapping $T : \mathcal{C}(G) \to \mathcal{P}_1([n])$ given by $T(C_i) = \{i\}$ for $i = 1, \ldots, n$. Let $C_i \neq C_j$ be two distinct concepts from $\mathcal{C}(G)$. The either $(i, j) \in E$ or $(j, i) \in E$. In the former case, $C_i(j) = 1$ and $C_j(j) = 0$ while, in the latter case, $C_j(i) = 1$ and $C_i(i) = 0$. Thus T avoids clashes between any pair of distinct concepts. It follows that $\mathrm{NCTD}(\mathcal{C}(G)) \leq 1$. ∎

Theorem 6 *For all sufficiently large n, there exists a tournament G of order n such that*

$$\mathrm{TD}_{min}(\mathcal{C}(G)) \geq \lfloor \log(n) - 2 \log \log(2n) \rfloor - 2 \ .$$

Proof We make use of the probabilistic method. Suppose that $\hat{G} = ([n], \hat{E})$ is a *random tournament*, i.e., for every $1 \leq i < j \leq n$, we decide by means of a fair coin whether (i, j) or (j, i) is included into $\hat{E}$. Consider the class $\mathcal{C} := \{C_1, \ldots, C_n\}$ with C_i as defined in (2) but with $\hat{E}$ at the place of E. Let $k \geq 1$ be a parameter whose precise definition (as a function in n) is postponed to a later stage. For every k-sample (S, b), let $E_{S,b}$ be the event that (S, b) is not realizable by $\mathcal{C}$, i.e., none of the concepts $C \in \mathcal{C}$ satisfies condition (1). Each concept C_i with $i \in [n] \setminus S$ satisfies (1) with a probability of exactly 2^{-k} (and these $n - k$ events are statistically independent). Therefore

$$\Pr(E_{S,b}) \leq (1 - 2^{-k})^{n-k} < \exp(-2^{-k}(n - k)) \leq \exp(-2^{-(k+1)}n) \ .$$

Here the last inequality is valid because we may assume that $n - k \geq n/2$. As there are $\binom{n}{k}2^k$ possible choices for (S, b), an application of the union bound yields

$$\Pr\left(\bigcup_{S,b} E_{S,b}\right) < \binom{n}{k}2^k \exp(-2^{-(k+1)}n) \ .$$

We now set

$$k' = \log(n) - 2 \log \log(2n) - 1 \ \text{ and } \ k = \lfloor k' \rfloor \ .$$

Claim 1: For all sufficiently large n, we have

$$\binom{n}{k}2^k \cdot \exp(-2^{-(k+1)}n)) < 1 \ .$$

Proof of the Claim: After replacing $\binom{n}{k}$ by n^k and taking the logarithm on both hand-sides, we obtain the following sufficient condition:

$$k \log(2n) - 2^{-(k+1)}n < 0 \ .$$

By the above choice of k, we have

$$k \log(2n) \leq k' \log(2n) < \log^2(2n) \ \text{ and } \ 2^{-(k+1)}n \geq 2^{-(k'+1)}n \geq \log^2(2n) \ ,$$

which completes the proof of the claim.

It follows that there is a strictly positive probability for the event that every k-sample (S, b) can be realized by at least one concept from $\mathcal{C}$. This implies that every $(k-1)$-sample can be realized by at least two concepts from $\mathcal{C}$. We may therefore draw the following conclusion: there exists a tournament G of order n such that none of the concepts in $\mathcal{C}$ can be uniquely specified by $k-1$ (or less) labeled examples. This clearly implies that $\mathrm{TD}_{min}(\mathcal{C}) \geq k-1$. ∎

Here is an immediate consequence of Lemma 5 and Theorem 6:

Corollary 7 *There exists a family $(\mathcal{C}_n)_{n \geq 1}$ of concept classes such that, for all sufficiently large n, the RTD-NCTD ratio of $\mathcal{C}_n$ is lower bounded by $\lfloor \log(n) - 2 \log \log(2n) \rfloor - 2$.*

With little extra-effort, one can show that only an asymptotically vanishing fraction of tournaments induces concept classes whose TD_{min} is upper bounded by $\lfloor \log(n) - 2 \log \log(2n) \rfloor - 3$. Hence almost all of these classes are hard to teach in the RTD-model. More precisely, the following holds:

Corollary 8 *Let τ_n denote the fraction of tournaments $G = ([n], E)$ such that $\mathrm{TD}_{min}(\mathcal{C}(G)) \leq \lfloor \log(n) - 2 \log \log(2n) \rfloor - 3$. Then, for all sufficiently large n, we have that*

$$\tau_n \leq \frac{1}{(2n)^{\log(2n)}} \quad . \tag{3}$$

Proof We use the probabilistic method thereby proceeding almost as in the proof of Theorem 6. In the sequel, we stress the differences to that proof:

- We set $k' = \log(n) - 2 \log \log(2n) - 2$ and $k = \lfloor k' \rfloor$.

- We have to show that[2]

$$\binom{n}{k} 2^k \cdot \exp\left(-2^{-(k+1)} n\right) < \frac{1}{(2n)^{\log(2n)}} \tag{4}$$

  holds for all sufficiently large n.

It suffices[3] to show that

$$k \log(2n) - 2^{-(k+1)} n < \log\left(\frac{1}{(2n)^{\log(2n)}}\right) = -\log^2(2n) \quad .$$

By the above choice of k, we have

$$k \log(2n) \leq k' \log(2n) < \log^2(2n) \text{ and } 2^{-(k+1)} n \geq 2^{-(k'+1)} n \geq 2 \log^2(2n) \quad ,$$

which completes the proof of (4). From these findings, it is easy to deduce (3).[4] ∎

---

2. Compare with Claim 1.
3. Compare with the proof of Claim 1.
4. Compare with the end of the proof of Theorem 6.

4. Construction of Concept Classes with a Large RTD-NCTD Ratio

In this section, we construct a concrete family of concept classes $\mathcal{C}_n$ such that $\mathrm{TD}_{min}(\mathcal{C}_n)$ grows logarithmically in n whereas $\mathrm{NCTD}(\mathcal{C}_n) = 1$ for every $n \geq 1$. It turns out that a suitable choice for $\mathcal{C}_n$ is the concept class induced by a tournament of order n, provided that the latter has the "strong S_k-property". The definition of the (weak and strong) S_k-property will be given in Section 4.1. In Section 4.2, we show that so-called QR-tournaments (of sufficiently large order) have the strong S_k-property. In Section 4.3, we argue that the concept classes induced by QR-tournaments have the desired properties.

4.1. Tournaments with the S_k-Property

The S_k-property of a tournament will be considered here in a weak and in a strong form:

Definition 9 (Weak S_k-Property) *A tournament $G = (V, E)$ is said to have the* weak S_k-property *if the following holds: for any choice of k distinct vertices $a_1, \ldots, a_k \in V$, there exists a vertex $x \in V$ such that $(x, a_j) \in E$ for $j = 1, \ldots, k$.*

Definition 10 (Strong S_k-Property) *A tournament $G = (V, E)$ is said to have the* strong S_k- property *if the following holds: for any choice of k distinct vertices $a_1, \ldots, a_k \in V$ and any choice of $b_1, \ldots, b_k \in \{\pm 1\}^k$, there exists a vertex $x \in V$ such that the following holds:*

$$\forall j = 1, \ldots, k : \left\{ \begin{array}{ll} (x, a_j) \in E & \text{if } b_j = +1 \\ (a_j, x) \in E & \text{if } b_j = -1 \end{array} \right. . \tag{5}$$

Let $f(k)$ (resp. $F(k)$) be the smallest number $n \geq k + 1$ such that there exists a tournament of order n which has the weak (resp. the strong) S_k-property. The following is known about the function $f(k)$:

$$2^{k-1}(k + 2) - 1 \leq f(k) \leq \min \left\{ n : \binom{n}{k}(1 - 2^{-k})^{n-k} < 1 \right\} \leq (1 + o(1)) \ln(2) k^2 2^k .$$

The lower bound was shown by Szekeres and Szekeres (1965) and the upper bound was shown by Erdös (1963). The latter is an easy application of the probabilistic method. Note that the gap between the lower and the upper bound is of order k. Clearly $f(k) \leq F(k)$ so that each lower bound on $f(k)$ is a lower bound on $F(k)$ too. Moreover, an obvious application of the probabilistic method yields an upper bound on $F(k)$ that differs from the above upper bound on $f(k)$ only by inserting an additional factor 2^k in front of $\binom{n}{k}$.[5] Hence we get

$$2^{k-1}(k + 2) - 1 \leq F(k) \leq \min \left\{ n : 2^k \binom{n}{k}(1 - 2^{-k})^{n-k} < 1 \right\} \leq (1 + o(1)) \ln(2) k^2 2^k .$$

4.2. Construction of Tournaments with the S_k-Property

As outlined in Section 4.1, the probabilistic method yields good upper bounds on $f(k)$ or $F(k)$, however without providing us with a concrete tournament which satisfies this bound. As far as the function $f(k)$ is concerned, Graham and Spencer (1971) have filled this gap. They defined and analyzed a tournament that is is based on the quadratic residues and non-residues in a prime field $\mathbb{F}_p$. It became known under the name quadratic-residue tournament (or briefly QR-tournament):

---

5. This factor accounts for the possible choices of $b_1, \ldots, b_k$.

Definition 11 (QR-Tournament) *Let p be a prime such that $p \equiv 3 \pmod 4$. The* QR-tournament *of order p is the tournament (V, E) given by*

$$V = \{0, 1, \ldots, p-1\} \ \text{and} \ E = \{(x, y) \in V \times V : x - y \text{ is a quadratic residue modulo } p\}$$

Let $\chi : \mathbb{F}_p \to \{-1, 0, 1\}$ be the function

$$\chi(x) = \begin{cases} +1 & \text{if } x \neq 0 \text{ is a quadratic residue modulo } p \\ -1 & \text{if } x \neq 0 \text{ is a quadratic non-residue modulo } p \\ 0 & \text{if } x = 0 \end{cases} \ .$$

In the sequel, p always denotes a prime that is congruent to 3 modulo 4 (so that -1 is a quadratic non-residue). Note that the graph (V, E) in Definition 11 is indeed a tournament because $\chi(y - x) = -\chi(x - y)$ so that exactly one of the edges (x, y) and (y, x) is included in E. Graham and Spencer have shown the following result:

Theorem 12 ((Graham and Spencer, 1971)) *The QR-tournament of order p has the weak S_k-property provided that $p > k^2 2^{2k-2}$.*

As we show now, the same construction works for the strong S_k-property:

Theorem 13 *The QR-tournament of order p has the strong S_k-property provided that $p > k^2 2^{2k-2}$.*

Proof The following proof builds on Graham and Spencer's proof for the (weaker) statement with the weak (instead of the strong) S_k-property. But we will have to deal with the variables $b_1, \ldots, b_k \in \{\pm 1\}$ that occur only in the definition of the strong S_k-property.
Consider first the case $k = 1$. A tournament has the strong S_1-property iff no vertex has in- or outdegree $p - 1$. Every QR-tournament has this property because every vertex has in- and outdegree $\frac{p-1}{2} < p - 1$.
The remainder of the proof is devoted to the case $k \geq 2$. Let $G = (V, E)$ be the QR-tournament of order p. Let $a_1, \ldots, a_k \in V$ be k distinct vertices and let $b_1, \ldots, b_k \in \{\pm 1\}$. Set $A = \{a_1, \ldots, a_k\}$, $a = (a_1, \ldots, a_k)$, $b = (b_1, \ldots, b_k)$ and consider the auxiliary functions

$$g(a, b) = \sum_{x \in V \setminus A} \prod_{j=1}^{k} [1 + b_j \chi(x - a_j)] \ \text{and} \ h(a, b) = \sum_{x=0}^{p-1} \prod_{j=1}^{k} [1 + b_j \chi(x - a_j)] \ .$$

An inspection of $g(a, b)$ reveals that there exists an $x \in V$ which satisfies (5) if and only if $g(a, b) > 0$. It suffices therefore to show that $g(a, b) > 0$. To this end, we decompose $g(a, b)$ according to

$$g(a, b) = p + (h(a, b) - p) - (h(a, b) - g(a, b)) \ .$$

In order to show that $g(a, b) > 0$, it suffices to show that

$$|h(a, b) - p| \leq \sqrt{p} \cdot ((k-2)2^{k-1} + 1) \ \text{and} \ h(a, b) - g(a, b) \leq 2^k \tag{6}$$

because $p - \sqrt{p}((k-2)2^{k-1} + 1) - 2^k > 0$ provided that $p > k^2 2^{2k-2}$, as an easy calculation shows.[6]

---

6. This calculation makes use of the case-assumption $k \geq 2$.

We still have to verify (6). In order to get $|h(a,b) - p| \leq \sqrt{p}((k-2)2^{k-1} + 1)$, we apply the distributive law and rewrite $h(a,b)$ as follows:

$$h(a,b) = \sum_{x=0}^{p-1} 1 + \sum_{x=0}^{p-1}\sum_{j=1}^{k} b_j \chi(x - a_j) + \sum_{r=2}^{k} S_r \tag{7}$$

where

$$S_r = \sum_{x=0}^{p-1} \sum_{1 \leq j_1 < ... < j_r \leq k} \prod_{i=1}^{r} b_{j_i} \chi(x - a_{j_i}) = \sum_{1 \leq j_1 < ... < j_r \leq k} \left(\prod_{i=1}^{r} b_{j_i} \right) \sum_{x=0}^{p-1} \prod_{i=1}^{r} \chi(x - a_{j_i}) \ . \tag{8}$$

Since $\sum_{x=0}^{p-1} 1 = p$ and

$$\sum_{x=0}^{p-1}\sum_{j=1}^{k} b_j \chi(x - a_j) = \sum_{j=1}^{k} b_j \underbrace{\sum_{x=0}^{p-1} \chi(x - a_j)}_{=0} = 0,$$

we can bring (7) in the form

$$h(a,b) - p = \sum_{r=2}^{k} S_r \ .$$

Burgess (1962) has shown that

$$\left| \sum_{x=0}^{p-1} \prod_{i=1}^{r} \chi(x - a_{j_i}) \right| \leq (r-1)\sqrt{p}$$

holds for every fixed choice of $1 \leq j_1 < ... < j_r \leq k$. In combination with (8), it follows that

$$|h(a,b) - p| = \left| \sum_{r=2}^{k} S_r \right| \leq \sqrt{p} \cdot \sum_{r=2}^{k} \binom{k}{r}(r-1) \ .$$

A straightforward calculation shows that $\sum_{r=2}^{k} \binom{k}{r}(r-1) = (k-2)2^{k-1} + 1$. We may therefore conclude that the first inequality in (6) is valid. We finally have to show that $h(a,b) - g(a,b) \leq 2^k$. Note first that

$$h(a,b) - g(a,b) = \sum_{i=1}^{k} \prod_{j=1}^{k} [1 + b_j \chi(a_i - a_j)] \ .$$

We call $\prod_{j=1}^{k}[1 + b_j \chi(a_i - a_j)]$ the *contribution* of i to $h(a,b) - g(a,b)$. Set $I_b = \{i \in \{0,1,\ldots,p-1\} : b_i = b\}$ for $b = \pm 1$. The following observations are rather obvious:

- Every i makes a contribution of either 0 or 2^{k-1}.

- If i makes a non-zero contribution, then $\chi(a_i - a_j) = b_j$ for every $j \neq i$.

- For each $b \in \{\pm 1\}$, at most one $i \in I_b$ makes a non-zero contribution.[7]

These observations imply that $h(a,b) - g(a,b) \leq 2^k$, which concludes the proof of the theorem. ∎

---

7. If two distinct $i, i' \in I_b$ made a non-zero contribution, then we would get $\chi(a_i - a_{i'}) = b = \chi(a_{i'} - a_i)$, which is in contradiction to $\chi(a_{i'} - a_i) = -\chi(a_i - a_{i'})$.

4.3. Implications on Teaching

We first observe the following:

Corollary 14 *For every tournament G with the strong S_k-property:* $\mathrm{TD}_{min}(\mathcal{C}(G)) \geq k$.

Proof An inspection of Definition 10 reveals that $\mathcal{C}(G)$ has the strong S_k-property iff every set $A \subseteq V$ of size k is shattered by $\mathcal{C}(G)$. This clearly implies that $\mathrm{TD}_{min}(\mathcal{C}(G)) \geq k$. ∎

Our main observation is that the quadratic-residue tournament of order p induces a concept class whose TD_{min} grows logarithmically with p:

Theorem 15 *Let p be a prime that is congruent to 3 modulo 4. Let G_p be the quadratic-residue tournament of order p. Then*

$$\mathrm{TD}_{min}(\mathcal{C}(G_p)) \geq \left\lfloor \frac{1}{2}\log(p) - \log\log(p) \right\rfloor \ . \tag{9}$$

Proof According to Theorem 13, the tournament G_p has the strong S_k-property provided that

$$k^2 2^{2k-2} < p \ , \tag{10}$$

In this case, we may conclude that $\mathrm{TD}_{min}(G_p) \geq k$. A straightforward calculation shows that (10) holds whenever $k \leq \frac{1}{2}\log(p) - \log\log(p)$. Setting k equal to the right hand-side in (9), the corollary follows. ∎

The following result is an immediate consequence of Theorem 15:

Corollary 16 *There exists a family $(\mathcal{C}_p)$ of concept classes, with p ranging over all primes being congruent to 3 modulo 4, such that, for all sufficiently large p, the RTD-NCTD ratio of $\mathcal{C}_p$ is lower bounded by $\left\lfloor \frac{1}{2}\log(p) - \log\log(p) \right\rfloor$.*

5. NC-Maximum Classes

The following bound of the Sauer-Shelah type is well known:

Theorem 17 ((Kirkpatrick et al., 2019a)) *Any concept class which has NCTD d and is defined over a domain of size n contains at most $2^d \binom{n}{d}$ concepts.*

An *NC-maximum class (with respect to parameters n and d)* is a concept class of largest size among all classes having NCTD d and being defined over a domain $\mathcal{X}$ of size n. The size of such a class will be denoted by $M_{NC}(n, d)$ in what follows. From Theorem 17, we conclude that $M_{NC}(n, d) \leq 2^d \binom{n}{d}$. We will see in the course of this section that this upper bound on $M_{NC}(n, d)$ is tight only for $d = 1$ and, for $d \geq 2$, it can be improved (at least) by a factor of order $\sqrt{d}$. Moreover, we will characterize the NC-maximum classes with respect to parameters n and $d = 1$.

5.1. NC-Maximum Classes of NCTD 1

Let $G = ([n], E)$ be a tournament of order n and let $\mathcal{C}(G) = \{C_1, \ldots, C_n\}$, with C_i as defined in (2), be the concept class induced by G. We now extend the class $\mathcal{C}(G)$ by setting $\mathcal{C}'(G) = \{C_1, \ldots, C_n, \bar{C}_1, \ldots, \bar{C}_n\}$ where $\bar{C}_i = [n] \setminus C_i$ for $i = 1, \ldots, n$. Lemma 5 can now be strengthened as follows:

Theorem 18 *Suppose that G is a tournament. Then $\mathrm{NCTD}(\mathcal{C}'(G)) = 1$.*

Proof Let $G = ([n], E)$ be a tournament of order $n \geq 1$. The concept class $\mathcal{C}'(G)$ contains at least two distinct concepts (e.g., the concepts $\emptyset$ and $\{1\}$ if $n = 1$). Hence $\mathrm{NCTD}(\mathcal{C}'(G)) \geq 1$.
Consider the mapping $T : \mathcal{C}'(G) \to \mathcal{P}_1([n])$ given by $T(C_i) = T(\bar{C}_i) = \{i\}$ for $i = 1, \ldots, n$. It was shown in the proof of Lemma 5 that T avoids clashes between C_i and C_j for all $i \neq j$ in the range from 1 to n. For reasons of symmetry, T avoids also the clashes between $\bar{C}_i$ and $\bar{C}_j$. Since $C_i(i) = 0$ and $\bar{C}_i(i) = 1$, T avoids clashes between C_i and $\bar{C}_i$ for all $i = 1, \ldots, n$. Consider now two indices $i \neq j$ in the range from 1 to n and the concept C_i and $\bar{C}_j$. Then either $(i, j) \in E$ or $(j, i) \in E$. In the former case, $C_i(i) = 0$ and $\bar{C}_j(i) = 1$ while, in the latter case $C_i(j) = 0$ and $\bar{C}_j(j) = 1$. This brief discussion shows that T avoids clashes between any pair of distinct concepts. It follows that $\mathrm{NCTD}(\mathcal{C}'(G)) \leq 1$. ∎

Theorem 19 *A concept class $\mathcal{C}$ over domain $[n]$ is an NC-maximum class of NCTD 1 if and only if $\mathcal{C} = \mathcal{C}'[G]$ for some tournament G of order n.*

Proof The if-direction is an immediate consequence of Theorems 17 and 18, and the fact that $|\mathcal{C}'(G)| = 2n = 2^1 \cdot \binom{n}{1}$. We may therefore focus on the only-if direction.
Suppose that $\mathcal{C}$ is an NC-maximum class of NCTD 1 over $[n]$. The first part of the proof in combination with Theorem 17 implies that $|\mathcal{C}| = 2n$. Let $T : \mathcal{C} \to \mathcal{P}_1([n])$ be an NC-teacher for $\mathcal{C}$. It follows that each set $\{i\} \in \mathcal{P}_1([n])$ is assigned to exactly two concepts.[8] Moreover these two concepts must disagree on $\{i\}$. We denote the concept with NC-teaching set $\{i\}$ that contains i (resp. does not contain i) by $\bar{C}_i$ (resp. by C_i). Fix two indices $i \neq j$ and consider the following assertions:

1. C_i disagrees with C_j on $\{j\}$.

2. C_i agrees with $\bar{C}_j$ on $\{j\}$.

3. $\bar{C}_j$ disagrees with C_i on $\{i\}$.

4. $\bar{C}_j$ agrees with $\bar{C}_i$ on $\{i\}$.

5. $\bar{C}_i$ disagrees with $\bar{C}_j$ on $\{j\}$.

6. $\bar{C}_i$ agrees with C_j on $\{j\}$.

7. C_j disagrees with $\bar{C}_i$ on $\{i\}$.

8. C_j agrees with C_i on $\{i\}$.

---

8. Assigning $\{i\}$ to three or more concepts would result in a clash.

11

Since the assignment of NC-teaching sets to concepts avoids clashes, it is easily seen that the following holds:

- Any assertion is an immediate logical consequence of the preceding one.

- The first assertion is an immediate logical consequence of the last one.

It follows that these eight assertions are equivalent. An inspection of the second and the fifth assertion reveals that $\bar{C}_i = [n] \setminus C_i$. Consider now the directed graph $G = ([n], E)$ with

$$E = \{(i, j) : C_i \text{ agrees with } \bar{C}_j \text{ on } \{j\}\} \ .$$

An inspection of the second and the seventh assertion reveals that exactly one of the edges (i, j) and (j, i) belongs to E. It follows that G is a tournament. Moreover, the above definition of E makes sure that, for every $i \in [n]$, $C_i = \{j \in [n] : (i, j) \in E\}$. We may therefore conclude that $\mathcal{C} = \mathcal{C}'[G]$. ∎

5.2. NC-Maximum Classes of Higher NCTD

Here we only state the main results and give a rough idea. The proofs, based on a new relation between NC-teaching and Johnson graphs, will be given in the appendix.

Theorem 20 *For $2 \le d \le n$, the following holds:*

$$M_{NC}(n, d) \le \left(2\sqrt{\frac{2}{d+1}} - \frac{2}{d+1} \right) \cdot 2^d \binom{n}{d} < 2^d \binom{n}{d} \ .$$

Note that the upper bound on $M_{NC}(n, d)$ in this theorem improves on the previously best bound, $2^d \binom{n}{d}$, by a factor of order $\sqrt{d}$.

Let $\mathcal{C}$ be a concept class that has NCTD d and is defined over a domain of size n. The proof in the appendix will make use of the following:

- An NC-teacher T assigning sets of size at most d to each concept in $\mathcal{C}$ can be normalized so as to assign only sets of size precisely d. Then T is formally a mapping from $\mathcal{C}$ to $\mathcal{P}_d(\mathcal{X})$.

- For any $F \in \mathcal{P}_d(\mathcal{X})$, let $m(F)$ denote the number of concepts $C \in \mathcal{C}$ such that $T(C) = F$. Then $|\mathcal{C}| = \sum_{F \in \mathcal{P}_d(\mathcal{X})} m(F)$ and $0 \le m(F) \le 2^d$ for every $F \in \mathcal{P}_d(\mathcal{X})$.[9]

- The upper bound $2^d \binom{n}{d}$ on $|\mathcal{C}|$ could be tight only, if we had $m(F) = 2^d$ for every $F \in \mathcal{P}_d(\mathcal{X})$.

- The theory of Johnson graphs can be used for showing the following "key result": for any t in the range from 2 to d, the number of $F \in \mathcal{P}_d(\mathcal{X})$ such that $m(F) > \frac{2^{d+1}}{t+1}$ is at most $\frac{t}{d+1} \binom{n}{d}$. This result will be the main building block in the proof of Theorem 20.

For $d = 2$, the upper bound on $M_{NC}(n, d)$ from Theorem 20 can be slightly improved:

Theorem 21 *For every $n \ge 2$, the following holds:*

$$M_{NC}(n, 2) \le \frac{(5n - 4)n}{3} \approx \frac{5n^2}{3} \ .$$

Similar slight improvements are possible for other small values of d.

---

9. Assigning F to more than 2^d concepts would result in a clash.

6. Open Problems

Theorem 19 characterizes NC-maximum classes of NCTD 1 as classes of the form $\mathcal{C}'[G]$ for some tournament G. Hence the structure of NC-maximum classes of dimension 1 is now perfectly known, whereas the structure of NC-maximum classes of higher dimension is still unknown.

Open Problem 1: Find structural properties which are shared by all NC-maximum classes of NCTD $d \geq 2$.

An obstacle for solving the first open problem is that we do not even know the size $M_{NC}(n, d)$ of NC-maximum classes having NCTD $d \geq 2$ and being defined over a domain of size n. While we can conclude from Theorem 19 that $M_{NC}(n, 1) = 2n$, the quantity $M_{NC}(n, d)$ with $d \geq 2$ is still unknown to us (although Theorem 20 makes a first step towards finding non-trivial bounds on $M_{NC}(n, d)$ for $d \geq 2$).

Open Problem 2: Find better (upper and lower) bounds on $M_{NC}(n, d)$ respectively, if possible, determine $M_{NC}(n, d)$ exactly.

As usual, $\mathrm{VCD}(\mathcal{C})$ denotes the VC-dimension (Vapnik and Chervonenkis, 1971) of $\mathcal{C}$. A problem that has been open for quite a while (Simon and Zilles, 2015) is the following one:

Open Problem 3: Is there a universal constant $c > 0$ such that, for each finite concept class $\mathcal{C}$, we have that $\mathrm{RTD}(\mathcal{C}) \leq c \cdot \mathrm{VCD}(\mathcal{C})$?

The corresponding problem with NCTD as the place of RTD is open as well. Only the following is known so far:

- As shown by Hu et al. (2017), $\mathrm{RTD}(\mathcal{C}) = O\left(\mathrm{VCD}(\mathcal{C})^2\right)$.

- As shown by Chen et al. (2016), there exists a class $\mathcal{C}$ such that $\mathrm{RTD}(\mathcal{C}) \geq \frac{5}{3} \cdot \mathrm{VCD}(\mathcal{C})$.

- It has been shown quite recently by Chalopin et al. (2022) that each maximum[10] class $\mathcal{C}$ has an unlabelled compression scheme of order $\mathrm{VCD}(\mathcal{C})$. As already observed by Kirkpatrick et al. (2019b); Falat et al. (2023), this implies that $\mathrm{NCTD}(\mathcal{C}) \leq \mathrm{VCD}(\mathcal{C})$ is true for every maximum class $\mathcal{C}$.

As yet, no concept class is known whose NCTD exceeds its VCD.

Acknowledgments

---

10. Maximum class is meant here with respect to VCD (not with respect to NCTD).

References

D. A. Burgess. On character sums and primitive roots. *Proc. London Math. Soc.*, 12(3):179–192, 1962.

Jérémie Chalopin, Victor Chepoi, Shay Moran, and Manfred K. Warmuth. Unlabeled sample compression schemes and corner peelings for ample and maximum classes. *Journal of Computer and System Sciences*, 127:1–28, 2022.

Xi Chen, Yu Cheng, and Bo Tang. On the recursive teaching dimension of VC classes. In *Proceedings of NeurIPS 2016*, pages 2164–2171, 2016.

Malte Darnstädt, Thorsten Doliwa, Hans U. Simon, and Sandra Zilles. Order compression schemes. In *Proceedings of the 24th International Conference on Algorithmic Learning Theory (ALT 2013)*, pages 173–187, 2013.

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.

Paul Erdös. On a problem of graph theory. *Math. Gaz.*, 47(361):200–223, 1963.

Shaun Falat, David Kirkpatrick, Hans U. Simon, Abolghasem Soltani, and Sandra Zilles. On batch teaching without collusion, 2023. To appear in JMLR.

Ziyuan Gao, Christoph Ries, Hans U. Simon, and Sandra Zilles. Preference-based teaching. *Journal of Machine Learning Research*, 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. *Journal of Computer and System Sciences*, 52(2):255–267, 1996.

Ronald L. Graham and Joel H. Spencer. A constructive solution to a tournament problem. *Canad. Math. Bull.*, 14(1):45–48, 1971.

Lunija Hu, Ruihan Wu, Tianhong Li, and Liwei Wang. Quadratic upper bound for recursive teaching dimension of finite VC classes. In *Proceedings of Machine Learning Research*, volume 65, pages 1147–1156, 2017.

Gyula Katona, Tibor Nemetz, and Miklós Simonovits. On a problem of Turán in the theory of graphs. *Mat. Lapok*, 15:228–238, 1964.

David G. Kirkpatrick, Hans U. Simon, and Sandra Zilles. Optimal collusion-free teaching. In *Proceedings of ALT 2019*, pages 506–528, 2019a.

David G. Kirkpatrick, Hans Ulrich Simon, and Sandra Zilles. Optimal collusion-free teaching. *CoRR*, abs/1903.04012, 2019b. URL http://arxiv.org/abs/1903.04012.

W. Mantel. Problem 28. *Wiskundige Opgaven*, 10:60–61, 1907.

Pál Turán. On an extremal problem in graph theory. *Math. Fiz. Lapok*, 48:436–452, 1941.

Ayumi Shinohara and Satoru Miyano. Teachability in computational learning. *New Generation Computing*, 8(4):337–348, 1991.

Hans U. Simon and Sandra Zilles. Open problem: Recursive teaching dimension versus VC dimension. In *Proceedings of the 28th Annual Conference on Learning Theory*, pages 1770–1772, 2015.

Esther Szekeres and George Szekeres. On a problem of Schütte and Erdös. *Math. Gaz.*, 49(369): 290–293, 1965.

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

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

Appendix A. No-Clash Teaching Sets and their Relation to Johnson Graphs

In Section A.1, we call to mind the definition of Johnson graphs (and related notions) along with some facts (all of which are well known and also easy to verify). In Section A.2, the tools from Section A.1 are used to improve the upper bound $2^d \binom{n}{d}$ on $M_{NC}(n, d)$ by a factor of order $\sqrt{d}$.

A.1. Johnson Graphs and their Subgraphs

Definition 22 (Johnson graph) *Let $J(n, k)$ denote the graph with vertex set $\mathcal{P}_k([n])$ and an edge between $A, B \in \mathcal{P}_k([n])$ iff $|A \cap B| = k - 1$. The graphs $J(n, k)$ with $1 \leq k \leq n$ are called* Johnson graphs *(named after the former American mathematician Selmer M. Johnson). A clique $\mathcal{K} \subseteq \mathcal{P}_k(n)$ in $J(n, k)$ is said to be* wide *if the sets in $\mathcal{K}$ have a common intersection of size $k - 1$. Analogously, $\mathcal{K}$ is said to be* narrow *if the union of all sets in $\mathcal{K}$ has size $k + 1$.*

Warning: The distinction between wide and narrow cliques would be blurred if we represented the $N = \binom{n}{k}$ vertices simply by numbers $1, \ldots, N$. In what follows, the representation of the N vertices by k-subsets of $[n]$ is essential.

Note that $J(n, 1)$ is isomorphic to the complete graph K_n. The vertices $\{1\}, \ldots, \{n\}$ of $J(n, 1)$ form a wide clique. $J(n, 2)$ is isomorphic to the line graph $L(K_n)$. $J(k, k)$ is a graph with a single vertex $[k]$ (and no edges). $J(k + 1, k)$ is isomorphic to K_{k+1}. The $k + 1$ vertices of $J(k + 1, k)$ form a narrow clique. Cliques of size 2 in $J(n, k)$ are wide and narrow. Cliques of size 3 or more cannot be wide and narrow at the same time. A clique of size 3 is also called *triangle* in the sequel. Here are some more of the known (and easy-to-check) facts concerning Johnson graphs:

Lemma 23

1. *Distinct sets in $\mathcal{P}_k([n])$ with a common intersection of size $k - 1$ (resp. a union of size $k + 1$) must necessarily form a clique.*

2. *Any clique in $J(n, k)$ is wide or narrow.*

3. *The mapping $A \mapsto [n] \setminus A$ is a graph isomorphism between $J(n, k)$ and $J(n, n - k)$. This isomorphism transforms narrow cliques into wide cliques, and vice versa.*

For any $\mathcal{F} \subseteq \mathcal{P}_k(n)$, we denote by $\langle \mathcal{F} \rangle$ the subgraph of $J(n, k)$ induced by $\mathcal{F}$. We denote the subgraph relation by "$\leq$" (e.g., $\langle \mathcal{F} \rangle \leq J(n, k)$). The following observation is rather obvious:

Lemma 24

1. *A graph with edge set $\mathcal{F} \subseteq \mathcal{P}_2([n])$ contains a triangle iff $\langle \mathcal{F} \rangle \leq J(n, 2)$ contains a narrow triangle.*

2. *A graph with edge set $\mathcal{F} \subseteq \mathcal{P}_2([n])$ contains a vertex of degree c or more iff $\langle \mathcal{F} \rangle \leq J(n, 2)$ contains a wide clique of size c.*

We now fix some notation. The size of the largest $\mathcal{F} \subseteq \mathcal{P}_k(n)$ such that $\langle \mathcal{F} \rangle \leq J(n, k)$ does not contain a narrow $(t + 1)$-clique is denoted by $H_t(n, k)$. Moreover, we set $h_t(n, k) = \binom{n}{k}^{-1} \cdot H_t(n, k)$. For any $\mathcal{F} \subseteq \mathcal{P}_k([n])$ and $I \subseteq [n]$, we define

$$\mathcal{F}_{+I} = \{J \in \mathcal{F} | J \subseteq I\} \ .$$

Note that any (narrow or wide) clique in $\mathcal{F}_{+I}$ would be a clique of the same type and size within $\mathcal{F}$. Hence, if $\langle \mathcal{F} \rangle$ does not contain a narrow $(t + 1)$-clique, then the same holds for $\mathcal{F}_{+I}$. Given these notations and observations, the following holds:

Lemma 25 *For all $1 \leq t \leq k \leq n - 2$:*

$$h_t(k, k) = 1 \ and \ h_t(n, k) \leq h_t(n - 1, k) \leq h_t(k + 1, k) = \frac{t}{k + 1} \ .$$

Proof $J(k, k)$ is a graph consisting of a single isolated vertex and $J(k + 1, k)$ is a narrow clique of size $k + 1$. Hence $H_t(k, k) = 1$ and $H_t(k + 1, k) = t$, which implies that $h_t(k, k) = 1$ and $h_t(k+1, k) = \frac{t}{k+1}$. The proof can now be accomplished by showing that $h_t(n, k) \leq h_t(n-1, k)$.[11] Fix a family $\mathcal{F} \subseteq \mathcal{P}_k([n])$ of size $H_t(n, k)$ such that $\langle \mathcal{F} \rangle \leq J(n, k)$ does not contain a narrow $(t + 1)$-clique.. There are $k \cdot H_t(n, k)$ occurrences of elements from $[n]$ within the sets of $\mathcal{F}$. By the pigeon-hole principle, there exists an $i \in [n]$ that occurs in at most $\frac{k}{n} \cdot H_t(n, k)$ sets of $\mathcal{F}$. Set $I = [n] \setminus \{i\}$. It follows that

$$H_t(n - 1, k) \geq |\mathcal{F}_{+I}| \geq \left(1 - \frac{k}{n}\right) H_t(n, k) \ .$$

Hence

$$h_t(n, k) \leq \frac{n}{n - k} \binom{n}{k}^{-1} \binom{n - 1}{k} h_t(n - 1, k) = h_t(n - 1, k) \ .$$

---

11. This inequality is, in principle, known from (Katona et al., 1964). The proof in (Katona et al., 1964) is written in Hungarian and it is formulated for hereditary properties of hypergraphs: if we view $\mathcal{F} \subseteq \mathcal{P}_k(n)$ as a set of k-uniform hyperedges, then not containing $t + 1$ hyperedges whose union is of size $k + 1$ will become a hereditary hypergraph property. In our application of this result, it is however more intuitive to view the elements of $\mathcal{F}$ as vertices of the Johnson graph. In order to make this paper more self-contained, we therefore included the short proof for $h_t(n, k) \leq h_t(n - 1, k)$, which uses a simple averaging argument.

We briefly note that the proof of the $h(n, k) \leq h(n - 1, k)$ made use only of the fact that the feature of avoiding a narrow $(t+1)$-clique is inherited from $\mathcal{F}$ to $\mathcal{F}_{+I}$. Hence the same monotonicity is valid whenever this kind of inheritance is granted.

The parameter $h_2(n, 2)$ can be determined exactly:

Remark 26 *For every $n \geq 2$, we have $h_2(n, 2) \leq \frac{n}{2(n-1)}$. Moreover, this holds with equality if n is even.*

Proof According to Mantel's theorem ((Mantel, 1907)) — in a more general form known as Turan's theorem ((Pál Turán, 1941)) — any triangle-free graph with n vertices has at most $n^2/4$ edges. For even n, this bound is tight because the complete bipartite graph $K_{n/2, n/2}$ is triangle-free and has $n^2/4$ edges. In combination with Lemma 24, we may conclude that $H_2(n, 2) \leq n^2/4$, and this holds with equality if n is even. Hence $h_2(n, 2) \leq \binom{n}{2}^{-1} \cdot \frac{n^2}{4} = \frac{n}{2(n-1)}$, again with equality if n is even. ∎

A.2. New Bounds on the Size of NC-Maximum Classes

Let $\mathcal{C}$ be a concept class over $[n]$ such that $\mathrm{NCTD}(\mathcal{C}) = d$, as witnessed by an NC-teacher $T : \mathcal{C} \to \mathcal{P}_d(n)$. Let $\mathcal{F} = \{T(C) : C \in \mathcal{C}\} \subseteq \mathcal{P}_d(n)$ be the family of all teaching sets assigned by T to the concepts of $\mathcal{C}$. For every $F \in \mathcal{F}$, let $1 \leq m(F) \leq 2^d$ denote the number of concepts $C \in \mathcal{C}$ with $T(C) = F$. Clearly $|\mathcal{C}| = \sum_{F \in \mathcal{F}} m(F)$. For every $2 \leq t \leq d$, we define

$$\mathcal{F}_t = \left\{ F \in \mathcal{F} : m(F) > \frac{2^{d+1}}{t+1} \right\} . \tag{11}$$

We view the sets in $\mathcal{F}_t$ as vertices in the Johnson graph $J(n, d)$ so that $\langle \mathcal{F}_t \rangle$ denotes the subgraph of $J(n, d)$ induced by $\mathcal{F}_t$. With these notations, the following holds:

Lemma 27 *The graph $\langle \mathcal{F}_t \rangle \leq J(n, d)$ does not contain a narrow $(t + 1)$-clique.*

Proof Assume for contradiction that $\langle \mathcal{F}_t \rangle$ does contain a narrow $(t + 1)$-clique $\mathcal{K}$, say $\mathcal{K} = \{F_1, \ldots, F_{t+1}\} \subseteq \mathcal{F}_t$. Set $D = F_1 \cup \ldots \cup F_{t+1} \subseteq [n]$. The definition of a narrow clique in $J(n, d)$ implies hat $|D| = d+1$. From the definition of $\mathcal{F}_t$, we may infer that $m(F_1) + \ldots + m(F_{t+1}) > 2^{d+1}$. Thus $\mathcal{C}$ contains more than 2^{d+1} concepts C whose NC-teaching set $T(C)$ belongs to $\mathcal{K}$. By the pigeon-hole principle, there must be two distinct concepts C_1 and C_2 such that $T(C_1), T(C_2) \in \mathcal{K}$ and C_1 and C_2 coincide on D. But, since $T(C_1) \cup T(C_2) \subseteq D$, this means that C_1 and C_2 clash with respect to T. We arrived at a contradiction. ∎

Here is an immediate consequence of Lemmas 25 and 27:

$$|\mathcal{F}_t| \leq H_t(n, d) = \binom{n}{d} h_t(n, d) \leq \frac{t}{d+1} \binom{n}{d} . \tag{12}$$

This upper bound on $|\mathcal{F}_t|$ is what we called the "key result" in Section 5.2.

We are now in the position to verify Theorem 20 from Section 5.2, which is covered by the following result:

Theorem 28 *For $2 \leq t \leq d \leq n$, the following holds:*

$$M_{NC}(n, d) \leq \left(h_t(n, d) + (1 - h_t(n, d)) \frac{2}{t+1} \right) \cdot 2^d \binom{n}{d} \ .$$

Moreover, for $t = \lfloor \sqrt{2(d+1)} \rfloor$, one gets

$$M_{NC}(n, d) \leq \left(2\sqrt{\frac{2}{d+1}} - \frac{2}{d+1} \right) \cdot 2^d \binom{n}{d} < 2^d \binom{n}{d} \ . \tag{13}$$

Proof Let $\mathcal{C}$ be an NC-maximum class for parameters n and d. Then $|\mathcal{C}| = M_{NC}(n, d)$. Consider an NC-teacher $T : \mathcal{C} \to \mathcal{P}_d(n)$ for $\mathcal{C}$. Let $\mathcal{F} = \{T(C) : C \in \mathcal{C}\}$ and let $\mathcal{F}_t \subseteq \mathcal{F}$ be as defined in (11). The size of $\mathcal{C}$ can be bounded as follows:

$$|\mathcal{C}| = \sum_{F \in \mathcal{F}} m(F) = \sum_{F \in \mathcal{F}_t} m(F) + \sum_{F \notin \mathcal{F}_t} m(F) \leq |\mathcal{F}_t| \cdot 2^d + \left(\binom{n}{d} - |\mathcal{F}_t| \right) \cdot \frac{2^{d+1}}{t+1} \ .$$

The right hand-side is made larger when we replace $|\mathcal{F}_t|$ by its upper bound $\binom{n}{d} h_t(n, d)$ from (12). Plugging in this upper bound, we obtain

$$|\mathcal{C}| \leq \left(h_t(n, d) + (1 - h_t(n, d)) \frac{2}{t+1} \right) \cdot 2^d \binom{n}{d} \ .$$

According to Lemma 25, we have $h_t(n, d) \leq \frac{t}{d+1}$. Moreover, we may set $t = \lfloor \sqrt{2(d+1)} \rfloor$ and can then proceed as follows:

$$h_t(n, d) + (1 - h_t(n, d)) \frac{2}{t+1} \leq \frac{t}{d+1} + \left(1 - \frac{t}{d+1} \right) \frac{2}{t+1} \leq 2\sqrt{\frac{2}{d+1}} - \frac{2}{d+1} \ .$$

Finally observe that

$$2\sqrt{\frac{2}{d+1}} - \frac{2}{d+1} \leq 1$$

with equality for $d = 1$ only. Putting everything together, we obtain (13). ■

The following corollary is a reformulation of Theorem 21 in Section 5.2:

Corollary 29 *For every $n \geq 2$, the following holds:*

$$M_{NC}(n, 2) \leq \frac{(5n - 4)n}{3} \approx \frac{5n^2}{3} \ .$$

Proof We know from Theorem 28 that

$$M_{NC}(n, 2) \leq \left(h_2(n, 2) + (1 - h_2(n, 2)) \frac{2}{3} \right) \cdot 4 \binom{n}{2} \ .$$

We know from Remark 26 that $h_2(n, 2) \leq \frac{n}{2(n-1)}$. The assertion of the corollary now follows from a straightforward calculation. ■

Similar slight improvements of the bound in Theorem 28 are possible for other small values of d.