Recursive Teaching Dimension, Learning Complexity, and Maximum Classes

Farnam Mansouri,
Advisor: Dr. Adish Singla

Sharif University of Technology

fmansouri@ce.sharif.edu

November 22, 2019

Overview

1. Definitions
   - Basic Definitions
   - Learning Complexity
   - Teaching complexity
2. Recursive Teaching and Query Learning
3. Recursive Teaching Dimension and VC-Dimension
   - Classes with RTD Exceeding VCD
   - Recursive Teaching and Intersection-Closed Classes
   - Recursive Teaching Dimension and Maximum Classes

Basic Definitions

Definition

For $X' \subseteq X$, $C_{|X'} = \{c \cap X' | c \in C\}$

Definition

S is a set of labeled examples, $X(S) = \{x \in X | (x, 0) \in S \text{ or } (x, 1) \in S\}$

Definition (monotonic function)

A function on concepts classes is called monotonic if, $\forall C' \subseteq C, K(C') \leq K(C)$, and is called twofold monotonic if K is monotonic and, $\forall X' \subseteq X, K(C_{|X'}) \leq K(C)$.

Partial equivalence query

- In each query learner represents an function.
- If the target concept is the function the oracle responds "yes", else it will represent a x which target concept and h contradict.
- LC-PARTIAL(C) is the smallest number of queries needed to find any $c* \in C$

Self-Directed model

- Learner selects an instance and predicts the label of it.
- number of wrong predictions according to target concept is number of queries.
- Self-Directed learning complexity is defined as $SDC(C) = min_L max_{c_t \in C}(M_{sd}(L, c_t))$, which $M_{sd}(L, c_t)$ is number of wrong predictions algorithm L made.
- SDC is monotonic

Partial equivalence query

- In each query learner represents an function.
- If the target concept is the function the oracle responds "yes", else it will represent a x which target concept and h contradict.
- LC-PARTIAL(C) is the smallest number of queries needed to find any $c* \in C$
- LC-PARTIAL is monotonic.

Mistake bound

- Algorithm A makes prediction for all instances, mistake bound of A, denoted by $M_A(C)$, is, number of wrong predictions
- $M_{opt}(C) = min_A M_A(C)$
- M_{opt} is twofold monotonic.
- the following chain is well-known:
  $SDC(C) \leq LC - PARTIAL(C) \leq M_{opt}(C)$.

Teaching Complexity

- A teaching set for C with respect to c is $\mathcal{TS}(C,c)$, is a all sets that are only consistent with c and no other concept, and $TS(C,c)$ is $min_{|z|}|C_{z|t}| = 1$.
- we define:
  - $TS_{min}(C) := min_{c \in C} TS(C,c)$
  - $TD(C) = TS_{max}(C) := max_{c \in C} TS(C,c)$ (the classic teaching dimension), note that TD is monotonic.

Definition (teaching plan)

P is sequence $((c_1, S_1), ..., (c_N, S_N))$, with following properties:

- $C = \{c_1, ..., c_n\}$
- $\forall 1 \leq t \leq N : S_t \in \mathcal{TS}(c, \{c_t, ..., c_n\})$

Teaching Complexity (cont.)

- $ord(P) := max_{t=1,...,N-1}|S_t|$ is called order of teaching plan.
- $RTD(C) := min\{ord(P)|P$ is a teaching plan for $C\}$
- $RTD^*(C) := max_{X' \subseteq X} RTD(C_{|X'})$
- a teaching plan is repetition-free if $X(S_1), ..., X(S_N)$ are pairwise distinct (note that $S_1, ..., S_N$ are always pairwise distinct).
- $rfRTD(C) := min\{ord(P)|P$ is a repetition-free teaching plan for $C\}$
- Every concept class has a repetition-free plan:
  Proof: we'll prove by induction on $|X|$, then $C_{x|1}$ by induction has repetition-free plan. We'll define $s = |C_{x|1}|$, if we consider $c_1, ..., c_s$, the repetition-free plan of $C_{x|1}$, and add (x,1) to all $S_i, i \leq s$, and then by induction we know again that $C_{x|0}$ also has repetition-free teaching plan, so we know that this new plan is repetition-free for C, since by induction we know two parts of plan can't have equal S_i, and also $i \leq s < j : S_i \neq S_j$, because the first parts all have x in their $X(S_l)$ but second part don't.

Teaching Complexity (cont.)

- RTD is monotonic.
- a teaching plan in canonical form is choosing the easiest to learn concept every time, i.e., $|S_t| = TS_{min}(c_t, \{c_t, ..., c_N\})$.
- RTD is equal to any teaching plan in canonical form.

Teaching Complexity (cont.)

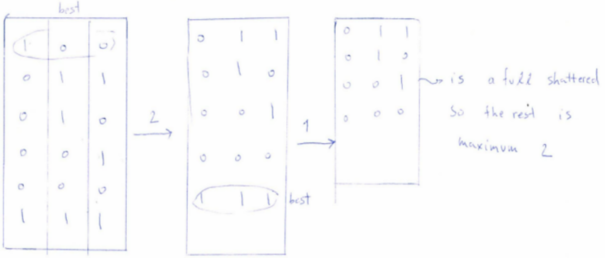


Figure: An example of teaching plan in canonical form

Teaching Complexity (cont.)

- **lemma1:**
  1. if K is monotonic and $\forall C : TS_{min}(C) \leq K(C)$, then $\forall C : RTD(C) \leq K(C)$
  2. if K is twofold monotonic and $\forall C : TS_{min}(C) \leq K(C)$, then $\forall C : RTD^{*}(C) \leq K(C)$
- **lemma2:** $RTD(C) = max_{C' \subseteq C} TS_{min}(C')$.
  proof: let c_1 be the first concept in canonical teaching plan, then, $RTD(C) = max\{TS(c_1, C), RTD(C \backslash \{c_1\}\}$, by induction we'll determine $RTD(C) \leq max_{C' \subseteq C} TS_{min}(C')$ [why is it less or equal?]. Now we'll find C'_0 which $TS_{min}(C'_0) = max_{C' \subseteq C} TS_{min}(C')$, since RTD is monotonic,
  $RTD(C) \geq RTD(C'_0) \geq TS_{min}(C'_0) = max_{C' \subseteq C} TS_{min}(C')$.

Corollary (Corollary 3)

All the following are the same:

1. $\exists k, \forall C : RTD^*(C) \leq k.VCD(C)$
2. $\exists k, \forall C : RTD(C) \leq k.VCD(C)$
3. $\exists k, \forall C : TS_{min}(C) \leq k.VCD(C)$

Proof: we know that $RTD^*(C) \geq RTD(C) \geq TS_{min}(C)$, so if 1 is true, 2 and 3 will be true. The other way around is true by applying lemma 1 and 2, since VCD is two-folded monotonic

Unlabeled Compression Schemes

lemma: we know that iff $C_{|X'} = \phi_d(|X'|)$

Definition (unlabeled compression scheme for a maximum class of VC-dimension d)

r is an injective mapping that assigns to every concept c a set of size at most d such that has non-clashing property, i.e.,
$\forall c \neq c' \in C, \exists x \in r(C) \cup r(C') : C(x) \neq C'(x)$.

Definition (acyclic non-clashing property)

C have acyclic non-clashing property if there is order $C_1, ..., C_N$ of C such that $\forall i < j, \exists x \in r(C_i) : C_i(x) \neq C_j(x)$.

Unlabeled Compression Schemes (cont.)

We know that for every sample S [size of S can be everything?], labeled according to a concept there is exactly one concept in C that is consistent with S and satisfies $r(C) \subseteq X(S)$, so we can encode S to r(C), and we can decode r(c) again to C to find S labels, we call that unlabeled compression scheme.

Recursive Teaching and Query Learning

Corollary (corollary 3)

1. *If $VCD(C) = 1$, then $TRD(C) = SDC(C) = 1$.*
2. *$RTD(\text{Monotone Monomials}) = SDC(\text{Monotone Monomials}) = 1$.*
3. $RTD(\mathit{Monomials}) = SDC(\mathit{Monomials}) = 2$.
4. $RTD(BOX_n^d) = SDC(BOX_n^d) = 2$.
5. *$RTD(m\text{-Term Monotone DNF}) \leq SDC(m\text{-Term Monotone DNF}) \leq m$.*
6. *$SDC(m\text{-Term Monotone DNF}) \geq RTD(m\text{-Term Monotone DNF}) \geq m$ provided that the number of Boolean variables is at least $m^2 + 1$.*

[I couldn't understand this proof]

lemma3: if $RTD(C) \geq 2$, then $RTD(C) \geq \frac{log|C|}{1+log|X|}$, and repetition-free teaching plans for C are of order at least $\frac{log|C|}{log|X|}$.
proof: P has —C— pairwise different teaching sets, and every teaching set is a subset of X of size at most k. Thus,
$|C| \leq \sum_{i=1}^{k} \left(\begin{array}{c} i \\ |X| \end{array} \right) 2^i \leq 2^k \phi_k(|X|) \leq (2|X|)^k.$
similarly we can proof for repetition-free, just the 2^i will be missing.

Recursive Teaching and Query Learning (cont.)

Definition

$X' \subseteq X$ is C-distinguishing if, all $c \in C$ are different on it.

Definition

Two matrices are equivalent if their incident matrices are equal up to permutation of rows or columns, and up to flipping all bits of a subset of the rows.

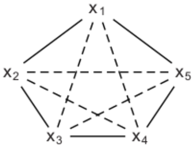
[I took out one theorem because it was trivial]

Classes with RTD Exceeding VCD

Kulmann paper represented a concept class, which $TS_{min}(C) = 3$, but $VCD(C) = 2$. now Warmuth represented a concept class with same property but it's instance size is 5 and number of concepts is 10.

x_1	x_2	x_3	x_4	x_5
1	1	0	0	0
0	1	1	0	0
0	0	1	1	0
0	0	0	1	1
1	0	0	0	1
0	1	0	1	1
0	1	1	0	1
1	0	1	1	0
1	0	1	0	1
1	1	0	1	0

(a)



(b)

Figure 1: The smallest concept class $\mathcal{C}_W$ with $\mathrm{RTD}(\mathcal{C}_W) > \mathrm{VCD}(\mathcal{C}_W)$. The function table to the left can be extracted from the graph to the right by picking concept $\{x_i, x_j\}$ for every solid line and $\mathcal{X} \setminus \{x_i, x_j\}$ for every dashed line.

Classes with RTD Exceeding VCD (cont.)

- **lemma4:** $RTD(C) \leq |X| - 1$ unless $C = 2^X$.
  proof: If $C \neq 2^X$, then C must contain a concept c such that $C\delta x \in C$ for some instance $x \in X$. Then, c can be uniquely identified within C using the instances from X x and the corresponding labels. Use this argument again each step, $RTD(C) \leq |X| - 1$ (We can extend this argument to rfRTD).
- We'll define $C[x, \lambda] := \{c \in C : c(x) = \lambda\}$, and we'll call an x redundant if $C[x, \lambda] = \varnothing$ for some λ
- **lemma5:** If $TS_{min}(C) \geq 2$, and x does not contain redundant instances. Then, $\forall x, \lambda : VCD(C[x, \lambda]) \geq 2$. **Proof:** If $VCD(C[x, \lambda]) \leq 1$, then by corollary 3, $TS_{min}(C[x, \lambda]) \leq RTD(C[x, \lambda]) \leq 1$, now choose (x, λ) as first teaching element, and $\mathcal{TS}(c^*, C[x, \lambda])$ as second one.

- **lemma6:** If $VCD(C) = 2$, $TS_{min}(C) = 3$, and X doesn't have redundant instances. Then $|X| \geq 5$ and, $\forall x, \lambda, |C_{x,\lambda}| \geq 5$.
  Outline of the proof: By previous lemma, $VCD(C[x, \lambda]) \geq 2$, and $VCD(C) = 2$, so $VCD(C[x, \lambda]) = 2$, Let $c_1, c_2, c_3, c_4 \in C_{x,\lambda}$ be concepts that shatter x', x'' in $X \backslash \{x\}$. For one of them, say c_1, $c_1 \delta \{x\}_{|\{x,x',x''\}}$ isn't in C, unless, $VCD(C) \geq 3$. If those 4 concepts are the only concepts in C, then $(x', c_1(x'))$, $(x'', c_1(x'')$ will be teaching sequence for c_1, so $|C_{x,\lambda}| \geq 5$.
  $|C_{x,\lambda} \geq 5|$, so $|X \ \{x\}| \geq 3$, now imagine $X = \{x, y, z, u\}$, we know that two instances which shatter, $C[x, 0]$ and $C[x, 1]$ are different, imagine they are $\{y, z\}$ and $\{z, u\}$ resp. since they can't be teaching set for any c, this will result that $VCD(C) = 3$.

Classes with RTD Exceeding VCD (cont.)

Theorem (Theorem 1)

If $RTD(C) > VCD(C)$, then $|C| \geq 10$, and $|X| \geq 5$.

Proof: If $VCD(C) \leq 1$, then $RTD(C) = VCD(C)$, then $VCD(C) \geq 2$, so $RTD(C) \geq 3$. We'll assume $RTD(C) = TS_{min}(C)$, because else there is a $C' \in C$ which $RTD(C) = TS_{min}(C')$ and $VCD(C) \geq VCD(C')$ so we can apply the same argument to C', also we assume no x is redundant. If $RTD(C) = 3$, then $VCD(C) = 2$, then by lemma 6, we know $|X| \geq 5$, and $|C| = |C[x,0]| + C[x,1] \geq 10|$. If $RTD(C) \geq 4$, then we know $RTD(C) \leq |X| - 1$, so $|X| \geq 5$, and we know $RTD(C) \leq log|C|$, so $|C| \geq 16 > 10$.
This means that Warmuth class is the minimal class that $RTD(C) > VCD(C)$.

Classes with RTD Exceeding VCD (cont.)

Definition

$C_1 \uplus C_2 := \{A \cup B | A \in C_1, B \in C_2\}$

Lemma: $\forall K \in \{VCD, TS_{min}, RTD\} : K(C_1 \uplus C_2) = K(C_1) + K(C_2)$. If $A_1, ..., A_M$ and $B_1, ..., B_N$ are C_1, and C_2 teaching plan resp. then $A_i \cup B_1, ..., A_i \cup B_N$ in round $i \in [M]$, so $RTD(C_1 \uplus C_2) \leq RTD(C_1) + RTD(C_2)$. For reverse side now we'll choose $C_1' \subseteq C_1$ and $C_2' \subseteq C_2$ in a way which $RTD(C_1) = TS_{min}(C_1')$ and $RTD(C_1) = TS_{min}(C_1')$. Then, $RTD(C_1 \uplus c_2) \geq TS_{min}(C_1' \uplus C_2') = TS_{min}(C_1') + TS_{min}(C_2') = RTD(C_1) + RTD(C_2)$.

Corollary

If we define $C_W^n := C_W \uplus ... \uplus C_W$. Then $VCD(C_W^n) = 2n$, and $RTD(C_W^n) = 3n$ (same result cannot be applied for rfRTD).

Recursive Teaching and Intersection-Closed Classes

Definition (sional boxes)

$BOX_n^d := \{[a_1 : b_1] * ... * [a_d : b_d] | \forall i = 1, ..., d : 1 \leq a_i, b_i \leq n\}$

Note that $BOX_n^d \in IC$.

Definition (samllest concept in C containing T)

$\langle T \rangle_C := \cap_{T \subseteq c \in C} c$

Definition (spanning set w.r.t C)

$\langle S \rangle_C = \langle T \rangle_C$, and we call it a minimal spanning set if $\forall S' \subset S : \langle S' \rangle_C \neq \langle S \rangle_C$, also I(C) is the largest minimal spanning set w.r.t C.

note that $\forall c' \subset C : I(C_{|c'}) \leq I(C)$ since each spanning set w.r.t to C is also a spanning set for $T \subseteq c'$ is also spanning set for T with respect to $C_{|c'}$.

lemma7: for $C \in IC$, $RTD(C) \leq I(C)$.
Proof: Let $c_1, ..., c_N$, be in topological order such that if $C_i \supset C_j$, then $i < j$, now if we consider s_i a member of spanning set of c_i, then we can recognize c_i in $c_i, ..., c_N$ with members of s_i since c_i isn't in any c_j so $|s_i| \leq I(c_i) \leq I(C) \leq VCD(C)$.

Corollary

for $C \in IC$, $RTD*(C) \leq VCD(C)$, since for every $X' in X$, $C_{|X'}$ is still intersection-closed, and $VCD(C_{|X'}) \leq VCD(C)$.

We know that there are a group of concept classes which $VCD(C) = d, SDC(C) \geq m$, by the previous corollary and this we'll conclude that the gap between RTD and SDC can be arbitrary high.

Recursive Teaching and Intersection-Closed Classes (cont.)

Definition (nested difference with depth d)

$DIFF^1(C) := C$,
$DIFF^d(C) := \{c \backslash D | c \in C, D \in DIFF^{d-1}(C)\}$,
$DIFF^{\leq d}(C) := \bigcup_{i=1}^{d} DIFF^d(C)$

Example: $DIFF^4(C)$ has the form $c_1 \backslash (c_2 \backslash (c_3 \backslash c_4))$ which, $c_1, c_2, c_3, c_4 \in C$, we can assume that $c_1 \supseteq c_2 \supseteq ...$ if $C \in IC$.

Theorem (theorem 2)

If $C \in IC$, then $RTD(DIFF^{\leq d}(C)) \leq d.I(C)$

Proof: we write $c \in DIFF^{\leq d}(C)$ in the form:
$c = c_1 \backslash (c_2 \backslash (...(c_{d-1} \backslash c_d)...))$, $c_j \in C \cup \{\varnothing\}$, $c_j \supseteq c_{j+1}$. We define $d_j = c_j \backslash (c_{j+1} \backslash (...(c_{d-1} \backslash c_d)...))$. We'll assume that the representation is minimal, i.e., $\forall j : c_j = \langle c_j \backslash d_j \rangle_C$ (so each c will have only a single representation) We'll define lexicographic ordering inductively as followed for each $c, c' \in C$: $c \sqsupset c'$ if $c_1 \supset c_1'$ or $c_1 = c_1'$ also $d_1 \sqsupset d_1'$. we'll sort $c_1, ..., c_N$ in lexicographical ordering and for each c we represent following teaching sequence for c, we add the minimal spanning set of c_j / d_j in it's minimal representation with label of 1 for odd j, and label of zero otherwise. every spanning set is smaller than I(C) so $|T| \leq d * I(C)$ for every c.

Recursive Teaching and Intersection-Closed Classes (cont.)

Corollary

$C_1, \ldots, C_r \in IC$, then $RTD(DIFF^{\leq d}(C_1 \cup \ldots \cup C_r)) \leq d \sum_{i=1}^{r} I(C_i)$.

Proof: *We'll define $C = \{c_1 \cup \ldots \cup c_r | c_i \in C_i\}$*

1. $C_1 \cup \ldots \cup C_r \subseteq C$.
2. $C \in IC$.
3. *If S_i is spanning set of c_i, then $S_1 \cup \ldots \cup S_r \in C$.*

By 2 and 3 we'll conclude $I(C) \leq I(C_1) + \ldots + I(C_r)$, and By 1 we know that $I(C_1 \cup \ldots \cup C_r) \leq I(C)$.

Maximum Classes Properties

Definition (one-inclusion graph)

$\mathcal{G}(C)$ is a graph defined on C which:

- nodes are concepts
- edges are two concepts who differ only in one coordinate, and we call edge who differ in only coordinate i, color i.

Maximum Classes Properties (cont.)

Definition (corner-peeling plan)

a sequence $P = ((c_1, C'_1), ..., (c_N, C'_N))$, with the following properties:

1. $C = \{c_1, ..., c_N\}$
2. $\forall t : C'_t$ is cube in $\{c_t, ..., c_N\}$ which contains c_t and all of its neighbors in $\mathcal{G}(\{c_t, ..., C_N\})$

Definition

The nodes c_t are called the corners of the cubes C'_t, respectively.

Definition (order of the corner-peeling plan)

Dimension of the largest cube among $C'_1, ..., C'_N$.

Maximum Classes Properties (cont.)

Definition (shortest path closed Concept class)

$\forall c, c' \in C : \mathcal{G}(C)$ contains a path from c to c' with length $d(c, c')$ (hamming distance)

Theorem

1. *If a maximum class C has a corner-peeling plan of order VCD(C), then an unlabeled compression scheme for C is obtained by setting* $r(C_t)$ *equal to the set of colors in cube* C'_t *for* $t = 1, ..., N$.
2. *every maximum class can be VCD(C)-corner-peeled.*

Maximum Classes Properties (cont.)

Corollary (corollary 3)

from Kuhlmann paper we know $TS_{min} \leq SDC(C)$ *since SDC is monotonic and* M_{opt} *is twofold monotonic, by lemma1 we can drive:*

1. $RTD(C) \leq SDC(C) \leq LC - PARTIAL(C) \leq M_{opt}(C)$
2. $RTD^*(C) \leq M_{opt}(C)$

Recursive Teaching Dimension and Maximum Classes

Definition (strong corner-peeling plan)

is a corner peeling plan which the second property is replaced as follows: For all $t \leq N$, C'_t is a cube in $\{C_t, ..., C_N\}$ which contains c_t and whose colors (augmented by their labels according to c_t) form a teaching set for $C_t \in \{C_t, ..., C_N\}$.

we call all colors in C'_t, S_t.

lemma8: A strong corner-peeling plan induces a teaching plan of the same order.

lemma9: Every strong corner-peeling plan is a corner-peeling plan.

Proof: if this isn't true then there is a c_i neighbor with c_t, which they're difference isn't in teaching set but this is a contradiction because c_t and c_i are consistent in all other x.

Recursive Teaching Dimension and Maximum Classes (cont.)

lemma10: Let C be a shortest-path closed concept class. Then, every corner-peeling plan for C is strong. **proof**: assume that there is a c which is consistent with c_t in S_t, then shortest path between c and c_t cannot be in C'_t, so one of $C - T$ neighbors cannot be in C'_T, which is a contradiction.

Corollary

Every corner-peeling plan for a maximum class is strong, and therefore induces a teaching plan of the same order.

Proof: *Since every maximum class is shortest-path closed all of it's corner-peeling plans are strong, also it has been proved every maximum class C can be VCD(C)-corner-peeled. Thus, we conclude that* $RTD(C) \leq VCD(C)$.

Recursive Teaching Dimension and Maximum Classes (cont.)

lemma: Every $k \leq VCD(C)$ instances for maximum concept classes can be shattered this means $RTD(C) \geq TS_{min}(C) \geq VCD(C)$.

Corollary

For maximum classes, $RTD(C) = VCD(C)$, and since $\forall X' \subseteq XC_{|X'}$ is maximum, $RTD^*(C) = VCD(C)$.

lemma11:

1. Every repetition-free teaching plan of order d for C induces a representation mapping r of order d for C given by $r(C_t) = X(S_t)$ for $t \leq N$. Moreover, r has the acyclic non-clashing property.
2. Every representation mapping r of order d for C that has the acyclic non- clashing property induces a teaching plan given by $S_t = \{(x, C_t(x)) | x \in r(C_t)\}$ for $t \leq N$. Moreover, this plan is repetition-free.

Recursive Teaching Dimension and Maximum Classes (cont.)

Proof:

1. For $t \leq t'$ if c_t and $c_{t'}$ clash this means c_t and $c_{t'}$ are consistent on S_t which is contradiction.
2. if S_t is not a teaching plan for c_t then there is some $c_{t'}, t' > t$ which $c_{t'}$ and c_t are consistent on $r(c_t)$ so c_t and $c_{t'}$ have clash.

Corollary

Let C be maximum of VC-dimension d. Then, there is a one-one mapping between repetition-free teaching plans of order d for C and unlabeled compression schemes with the acyclic non-clashing property.

Shortest-Path Closed Classes

Theorem

If C is shortest-path closed, Then $TS_{avg}(C) < 2VCD(C)$.

By previous lemma we know that $TS_{avg}(C)$ is equal to average vertex degree of $\mathcal{G}(C)$ which is twice the density of the graph, and it has been proven that $dens(\mathcal{G}(C)) < VCD(C)$.

Theorem

If $VCD(C) = 1$, then $TS_{avg}(C) < 2$.

The Kuhlmann proof for this was flawed, we give alternative one, we know that every concept class with VCD(C) can be embedded to a maximum class, and maximum classes are shortest path closed, so by previous lemma we'll drive $TS_{avg}(C) < 2$

Shortest-Path Closed Classes (cont.)

Note that,: Kushilevitz et al. (1996) have Presented C_n which $TS_{avg}(C_n) = \Omega(\sqrt{|C_n|})$, but $VCD(C_n) \leq log|C_n|$.
lemma: If $\forall c \in C : deg_{\mathcal{G}(C)}(c) \geq |X| - 1$ then C is shortest path closed.
proof: pick c, c' concepts with minimum distance d. By contradiction, all neighbors of c with hamming distance of d-1 to c' cannot be in C, so $deg_{\mathcal{G}(C)}(c) \geq |X| - 1$.
Rubinstein et al. presented a graph with property $\forall c \in C : deg_{\mathcal{G}(C)}(c) \geq |X| - 1$ which $TS_{min} > VCD(C)$. Thus, the inequality $TS_{min}(C) \leq VCD(C)$ does not generalize from maximum classes to shortest-path closed classes.

The End