# TD8 - VC-dimension

*Indication of hardness: from (*) to (****).*

## 1 Hitting Set problems

**Exercise 1 - $d$-intervals**
An *(d)-interval graph $G$* is an intersection graph of ($d$) intervals in the line. In other words, vertices are represented via ($d$) intervals in the line. And there is an edge between two vertices if at least one of their corresponding intervals intersect.

For instance, the vertices of 1-interval graphs are intervals $[a, b]$ and there is an edge between the vertices represented by $[a, b]$ and $[c, d]$ if and only if the intervals $[a, b]$ and $[c, d]$ intersect.

A point $x$ of the line intersect a vertex $v$ of a $d$-interval graph if $x$ is in one of the $d$-intervals of $v$. We denote by $\tau$ the minimum number of points of the line intersecting all the vertices of a $d$-interval graph. We denote by $\nu$ the maximum number of vertices $X$ of $G$ whose intervals are pairwise disjoint (i.e. for every $x, x' \in X$, the union of the intervals of $x$ and the union of the intervals of $x'$ are disjoint). In other words, $\nu$ is the maximum size of an independent set of $G$.

(a) Formulate the problem of finding $\tau$ as a Hitting Set problem in a hypergraph.
(b) Prove that the dual of this problem is the MIS in $G$. In other words, the optimal of the dual is $\nu$.
(c) Prove that $\tau = \nu$ for 1-interval graphs algorithmically.
(d) Another proof. Prove that for unit interval graphs, the constraint matrix is TU. Conclude that $\tau = \nu$.
(e) Propose a lower bound on the gap between $\tau$ and $\nu$ for $d$-interval graphs.

## 2 VC-dimension

**Exercise 2 - VC-dimension of graph classes (**)**
Prove that the neighborhood hypergraph of the following classes have VC-dimension at most:

(a) 2 for interval graphs (see Exercise 1 for a definition).
(b) 2 for graphs of girth at least 5. The girth of a cycle is the size of a minimum cycle.
(c) 3 for Unit Disks graphs. A unit disk graph is a graph where vertices can be represented as unit disks in the plane. And there is an edge between two vertices if the corresponding disks intersect.
(d) Prove that these bounds are tight.

**Exercise 3 - Dual hypergraph and dual VC-dimension (**)**
A hypergraph $H = (V, E)$ can be seen as a bipartite graph $B$ with vertex set $V \cup E$ and where there is an edge between $v$ and $e$ in $B$ if $v \in e$ in $H$.

(a) Interpret the VC-dimension in this bipartite graph.

(b) The dual VC-dimension is the VC-dimension of the bipartite graph where we permute the roles of hyperedges and vertices. Prove that

$$DVC \leq 2^{VC+1} - 1$$

where DVC is the dual VC-dimension.

**Exercise 4 - VC-dimension and bipartite graphs (***)**
A graph $G$ contains all the bipartite graphs of size $k$ if, for every bipartite graph $H$ with both side of size, there exist two disjoint subsets $A, B$ of size $k$ such that the edges between $A$ and $B$ correspond to the graph $H$. Note that we make no asumption on the possible edges inside $A$ and $B$. In this exercise we define the VC-dimension of a graph $G$ as the VC-dimension of the *neighborhood hypergraph* (vertices correspond to vertices of $G$ and we create a hyperedge for each vertex $v$ where $e_v$ is the closed neighorhood of $v$ in $G$).

(a) Prove that if $G$ contains all the bipartite graphs of size $k$, then its VC-dimension is at least $\lfloor \log(k) \rfloor$.

(b) Prove that if $G$ has VC-dimension at least $k$, then it contains all the bipartite graphs of size $(k - \lceil \log(k) \rceil - 1) \times (k - \lceil \log(k) \rceil - 1)$.

**Exercise 5 - Proof of Sauer's Lemma (**)** The goal of this exercise is to prove the Sauer's Lemma we have seen during the lectures: *Let $H = (V, E)$ be a simple hypergraph of VC-dimension $d$. For every set $X \subseteq V$, the number of (distinct) traces of $E$ on $X$ is at most $\sum_{i=0}^{d} \binom{|X|}{i}$.*

1. Prove that when $d = 0$ or $n = 1$ the conclusion holds.

2. Assume now that $d \geq 1$ and $n \geq 1$. Let $v$ be a vertex and let $E_1, E_2$ be a partition of $E$ defined as follows:

$$E_1 = \{e \text{ such that } v \in e \text{ and } e \setminus v \in E\}$$

$$E_2 = \{e \text{ such that } \exists e' \in E \setminus E_1, e = e' \setminus v\}$$

   Let us define $H_1$ (resp. $H_2$) as the hypergraph on vertex set $V \setminus v$ and with edge set $E_1$ (resp. $E_2$).

   Prove that both hypergraphs are simple.

3. Prove that both hypergraphs have VC-dimension $d - 1$.

4. Conclude.

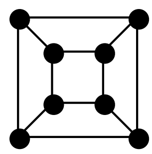
**Exercise 6 - Proof of Haussler-Welzl result (****)** A *measure* of a hypergraph is a weight (*i.e.*non negative) function on the vertex set such that the sum of the weights equals one. Let $H$ be a hypergraph and $\mu$ be a measure on the vertex set of $H$. An *$\epsilon$-net* is a subset of vertices $X$ such that every hyperedge of weight at least $\epsilon$ intersects $X$. Our goal is to prove the following theorem of the lectures: *Every hypergraph of VC-dimension $d$ (and weight function on it) has an $\epsilon$-net of size $\mathcal{O}(\frac{d\ln(d/\epsilon)}{\epsilon})$.*

We want to prove it for the uniform measure. Let $X$ be a subset selected uniforly at random of size $s := C \cdot (d/\epsilon)\ln(d/\epsilon)$.

1. Using Tchebychev inequality prove that $\mathbb{P}(|e \cap X| \geq s\epsilon/2) \geq 1/2$.
2. Let us call $E_0$ the event "there exists a hyperedge which is not intersected by $X$". Prove that if $\mathbb{P}(E_0) < 1$, then the conclusion holds.
3. Let $Y$ be another set of size $|X|$ drawn independently uniformly at random. A hyperedge $e$ *heavily intersects* $Y$ if $|e \cap Y| \geq s\epsilon/2$.
   Call $E_1$ the event "there exists a hyperedge which is not intersected by $X$ and which heavily intersects $Y$".
4. Prove that $\mathbb{P}(E_1) \geq 1/2\mathbb{P}(E_0)$ and $\mathbb{P}(E_1) \leq \mathbb{P}(E_0)$ .
5. Prove that $\mathbb{P}(E_1) < 1/2$.
   *Hint:* Let $Z = X \cup Y$. Let $e$ be a hyperedge heavily intersecting $Y$ and which do not intersect $X$. How many partitions of the set $Z$ permits to reach this objective if we assume that $X$ and $Y$ is a bipartition of $Z$?
6. Using Sauer's Lemma, what can you say about the number of traces on $Z$? Conclude.

# 3 Application of VC-dimension

**Exercise 7 - Graph coloring.**
The goal of this exercice consists in showing the following result:
*Every (triangle,cube)-free graph on $n$ vertices and of minimum degree $c \cdot n$ can be colored using at most $\mathcal{O}(\frac{\log(1/c)}{c})$ colors.*



A *cube* is the graph represented at the right. A graph is *triangle-free* if it does not contain any clique of size at least 3. A graph is cube-free if the restriction of the graph to 8 vertices is never the cube. A *coloring* of $G$ consists in coloring the vertices of $G$ in such a way adjacent vertices receive distinct colors.
In what follows, we say that $G$ is in $\mathcal{C}$ if $G$ is a (triangle,cube)-free graph of minimum degree $c\cdot n$. Recall that for $v \in V$, $N(v)$ is the set of neighbors of $v$ and $N[v] = N(v) \cup \{v\}$. A *dominating* set of $G$ is a subset $X$ of vertices such that $N[X] = V(G)$, i.e. every vertex is in $X$ or in the neighborhood of a vertex of $X$.

(a) Formulate the minimum dominating set problem as an ILP.

(b) We consider the closed-neighborhood hypergraph of $G$ denoted by $H_G$. The vertex set of $H_G$ is the vertex set of $G$. A set $S$ is a hyperedge of $H$ if there exists $v \in V(G)$ with $S = N[v]$.

Prove that $X \subseteq V$ is a Dominating Set of $G$ if and only if $X$ is a Hitting Set of $H_G$.

(c) Prove that if $G \in \mathcal{C}$ has a dominating set of size at most $k$, then $G$ can be colored with $2k$ colors.

(d) Let $G \in \mathcal{C}$. Give an upper bound on the optimal value of the fractional relaxation of the Hitting Set LP of $H_G$?

(e) Let $G \in \mathcal{C}$. Give an upper bound on the VC-dimension of $H_G$.

(f) Conclude.

(g) Are your bounds of questions (d) and (e) tight? Provide lower bounds.

**Exercise 8 - Balls of planar graphs (****)**

Let $G$ be a connected graph. The distance between $x$ and $y$ is the length of a minimum path between $x$ and $y$. A dominating set at distance $d$ is a subset of vertices $X$ such that every vertex is at distance at most $d$ from a vertex of $X$. The diameter of a graph $G$ is the maximum of the minimum distance between the vertices of $G$.

(a) Define the minimum dominating set at distance $d$ as a Hitting Set problem.

(b) Interpret the dual of this problem.

(c) Assume that $G$ has diameter at most $2d$, show that the (integral) optimal value of the dual is at most one.

(d) A planar graph does not admit $K_5$ as a topological minor. In particular it means that there do not exist 5 vertices $v_1, \ldots, v_5$ and a collection of paths $P_{ij}$ for $i < j \leq 5$ such that $P_{ij}$ and $P_{kl}$ do not intersect if $\{i, j\}$ disjoint from $\{k, l\}$.
Using this lemma, prove that the 2VC-dimension of the hypergraph of the balls of radius $d$ is bounded.

(e) Conclude that the dominating set is bounded.