VC-DIMENSIONS FOR GRAPHS

Evangelos Kranakis[*†] (kranakis@scs.carleton.ca)

Danny Krizanc[*†] (krizanc@scs.carleton.ca)

Berthold Ruf[*§] (ruf@scs.carleton.ca)

Jorge Urrutia[‡†] (jorge@csi.uottawa.ca)

Gerhard J. Woeginger[¶] (woe@fmatbds01.tu-graz.ac.at)

Abstract

We study set systems over the vertex set (or edge set) of some graph that are induced by special graph properties like clique, connectedness, path, star, tree, etc. We derive a variety of combinatorial and computational results on the VC (Vapnik-Chervonenkis) dimension of these set systems.

For most of these set systems (e.g. for the systems induced by trees, connected sets, or paths), computing the VC-dimension is an NP-hard problem. Moreover, determining the VC-dimension for set systems induced by neighborhoods of single vertices is complete for the class LogNP. In contrast to these intractability results, we show that the VC-dimension for set systems induced by stars is computable in polynomial time. For set systems induced by paths or cycles, we determine the extremal graphs G with the minimum number of edges such that $\mathrm{VC}_{\mathcal{P}}(G) \geq k$. Finally, we show a close relation between the VC-dimension of set systems induced by connected sets of vertices and the VC dimension of set systems induced by connected sets of edges; the argument is done via the line graph of the corresponding graph.

[*]Carleton University, School of Computer Science, Ottawa, ON, K1S 5B6, Canada.

[†]Research supported in part by NSERC (National Science and Engineering Research Council of Canada) grant.

[‡]University of Ottawa, Department of Computer Science, Ottawa, ON, K1N 9B4, Canada.

[§]TU Graz, Institute of Theoretical Computer Science Klosterwiesgasse 32/2, A-8010 Graz, Austria. Research supported in part by a grant from the DAAD (German Academic Exchange Service).

[¶]TU Graz, Department of Mathematics, Kopernikusgasse 24, A-8010 Graz, Austria. Research supported by the Spezialforschungsbereich F 003 “Optimierung und Kontrolle”, Projektbereich Diskrete Optimierung.

1980 Mathematics Subject Classification: 68Q25
CR Categories: F.2.0
Key Words and Phrases: Graph properties, Set system, Vapnik-Chervonenkis dimension.
Carleton University, School of Computer Science: SCS-TR-255
Note: This paper will be submitted for publication elsewhere.

1 Introduction

The *Vapnik-Chervonenkis-dimension* of a set system dates back to a seminal paper by Vapnik and Chervonenkis [12] in 1971 on the uniform convergence of relative frequencies of events to their probabilities. It is defined as follows. For $\mathcal{F}$ a family of subsets of a finite set X and $D \subseteq X$, set D is said to be *shattered* by $\mathcal{F}$ iff any subset of D is of the form $D \cap F$ for some $F \in \mathcal{F}$. The Vapnik-Chervonenkis (or VC, for short) *dimension* of $\mathcal{F}$ is the maximum size of a subset of X that is shattered by $\mathcal{F}$. In the meantime, the VC-dimension has proved useful in many areas as in probability theory, in learnability theory (PAC-learnable concept classes can be characterized via the VC-dimension, cf. Blumer, Ehrenfeucht, Haussler, and Warmuth [2]) and in computational geometry (geometric range spaces allow linear sized data structures with sublinear query time iff their VC dimension is finite, cf. Chazelle and Welzl [3]).

Papadimitriou and Yannakakis [8] investigated the computational complexity of computing the VC-dimension. Since $\mathrm{VC}(\mathcal{F}) \leq \log(|\mathcal{F}|)$ holds, the VC-dimension can be computed in $O(|X|^{\log(\mathcal{F})})$ time by simply checking all subsets of X of cardinality $\leq \log(\mathcal{F})$. This indicates that the problem is not NP-complete. To provide stronger evidence against NP-completeness, Papadimitriou and Yannakakis introduced the complexity class LogNP and proved that the following problem is LogNP-complete: Given a family $\mathcal{C}$ of c sets over a set X (by explicit enumeration of all sets in the family) and an integer k, is the VC-dimension of $\mathcal{C}$ at least k ? The class LogNP is sandwiched between P and NP, $P \subseteq$ LogNP $\subseteq$ NP, and the general belief is that both inclusions are proper. Hence, with high probability LogNP-complete problems are neither NP-complete nor solvable in polynomial time.

A special class of set systems arises in connection with graphs. Haussler and Welzl [5] introduced the VC-dimension of a graph as an example in their study of simplex-range queries with epsilon nets. Their definition is as follows. For $G = (V, E)$ a simple, loopless, undirected graph with vertex set

V and edge set E, the closed neighborhood $N(v)$ of a vertex $v \in V$ is the set consisting of the vertex v together with all vertices adjacent to it. A set $D \subseteq V$ of vertices is called shattered if it is shattered by the family $\mathcal{F}_{nbd} = \{N(v) : v \in V\}$ of neighborhoods of G (in the sense of the above definition of shatteredness). Since a graph has as many neighborhoods as it has vertices, its VC-dimension clearly is at most $\log |V|$. Anthony et al [1] study this VC-dimension in more detail and show that the threshold probability for a random graph to have VC-dimension $\geq d$ is about $p = n^{-1/d}$, for d sufficiently large, where n is the number of vertices of the graph.

1.1 Results of the paper

The VC-dimension of a graph as defined by Haussler and Welzl is defined via subsets of V that are neighborhhoods of single vertices. It is natural to investigate a more general concept where the VC dimension results from set systems induced by other properties on sets of vertices as e.g. cliques, connected sets, paths, stars, trees, cycles, etc. In this paper we will introduce and study the VC-dimensions for all these properties.

Connectedness. We will study in detail the VC-dimension for set systems induced by *connected sets* and show that for a given graph, the maximum size of a shattered set for the connectedness property differs by at most one from the number of leaves in a maximum leaf spanning tree. Hence, we can approximate this VC-dimension by applying the approximation algorithms for maximum leaf spanning trees derived by Lu and Ravi [9]. Moreover, we prove that computing the VC-dimension for set systems induced by *connected sets* is NP-complete.

The reader should note that the LogNP-completeness complexity result derived by Papadimitriou and Yannakakis [8] is *not* in contradiction to our NP-completeness result: [8] considered a problem where the input is given by explicit enumeration of all sets, whereas in our case the input is implicitely described via the graph and hence a potentially exponential number of sets (all connected subsets of the graph) is encoded by a structure of polynomial size (edge and vertex set of the graph).

Paths. Computing the VC-dimension of set systems induced by *paths* will also be proved to be NP-hard. Moreover, we give a complete combinatorial characterization of the graphs for which this VC-dimension equals three, and we provide upper and lower bounds on the number of edges in terms of the number of vertices and the VC-dimension.

Neighborhoods and stars. In contrast to the two NP-hardness results above, we show that computing the VC-dimension for set systems induced by *neighborhoods* is LogNP-complete and that the computation of the VC-dimension for set systems induced by *stars* can be done even in polynomial time.

Connected sets of edges. Finally, we will study shattering principles for families of edge-sets and the corresponding *edge*-VC *(or* EVC *for short) dimension* of graphs. We will show that the EVC-dimension for set systems over edges induced by connected edge sets in some graph G is related in a specific way to the VC-dimension for set systems over vertices induced by connected vertex sets in the corresponding line graph of G. Moreover the problem of computing the EVC dimension for connected sets is shown to be NP-complete.

1.2 Organization of the paper

The paper is organized into sections as follows. Section 2 introduces several general concepts and gives all basic definitions. Section 3 deals with the VC-dimension resulting from neighborhoods and stars, and Section 4 with the VC dimension resulting from connected sets and trees. Section 5 treats the VC-dimension for paths and Section 6 the VC-dimension for cycles. Section 7 states the results on the VC-dimension for edges, and Section 8 finishes the paper with the conclusion.

2 VC-dimensions for Vertices

In this section we give precise definitions for the notion of the VC-dimension of a graph G with respect to certain graph properties. Let $G = (V, E)$ be a graph with vertex set V and edge set E. Let $\mathcal{P}$ be a family of subgraphs of G. subsets of V. Typical choices for $\mathcal{P}$ include families of subgraphs which are cliques, connected, neighborhoods, paths, trees, etc.

Definition 2.1 *Let $\mathcal{P}$ be a family of subgraphs of G. We say that a subset $A \subseteq V$ is $\mathcal{P}$-*shattered *if and only if for all $B \subseteq A$ there exists a subgraph in $\mathcal{P}$ on a set of vertices $C \subseteq V$ such that $B = C \cap A$. Then the* VC-dimension with respect to $\mathcal{P}$ *of G is defined by*

$$\mathrm{VC}_{\mathcal{P}}(G) = \max\{|A| : A \text{ is } \mathcal{P} - \text{shattered}\} \tag{1}$$

Thus, depending on the family $\mathcal{P}$, we get the following notions of VC-*dimensions:* VC_{con}, VC_{path}, VC_{star}, VC_{tree}, VC_{cycle}, VC_{nbd}, *for the properties connected, path, star, tree, cycle, neighborhood, respectively.*

Note that the VC-dimensions as defined in Anthony et al. [1] is the same as our VC_{nbd}. The following examples might be helpful for a better understanding of these definitions.

Example 2.1 For the complete graph K_n on n vertices, $\mathrm{VC}_{con}(K_n) = n$ holds since any subset of the vertices is connected. For a path P_n on $n \geq 2$ vertices, we get $\mathrm{VC}_{con}(P_n) = 2$: A set of 3 vertices cannot be shattered since it is impossible to connect the outer two vertices without using the inner vertex. By the same argument we see that $\mathrm{VC}_{tree}(P_n) = \mathrm{VC}_{path}(P_n) = 2$. For a cycle C_n on $n \geq 3$ vertices, it can be checked that $\mathrm{VC}_{con}(C_n) = 3$. ■

Lemma 2.1 *If $\mathcal{P} \subseteq \mathcal{P}'$ then $\mathrm{VC}_{\mathcal{P}}(G) \leq \mathrm{VC}_{\mathcal{P}'}(G)$.* ■

The problem of computing the $\mathrm{VC}_{\mathcal{P}}$-dimension of a graph for a given graph property $\mathcal{P}$ can be formulated as the following decision problem.

> **Problem** $\mathrm{VC}_{\mathcal{P}}$:
>
> **Instance:** A graph $G = (V, E)$, a positive integer $k \leq |V|$.
>
> **Question:** Is there an $A \subseteq V$ with $|A| \geq k$ such that for all $B \subseteq A$ there is a subgraph $G' = (V', E')$ of G having property $\mathcal{P}$ such that $B = V' \cap A$?

Computing the VC-dimension for some graph property $\mathcal{P}$ is sometimes equivalent to well-known problems studied in complexity theory. For example, if $\mathcal{P}$ is the family of *cliques*, or the family of *independent sets*, it is easy to verify that $\mathrm{VC}_{clique}(G)$ (respectively, $\mathrm{VC}_{independent}(G)$) equals the size of the largest clique (respectively, largest independent set) in the graph G. It is well-known that both of these problems are NP-complete (see e.g. Garey and Johnson [4]). A related optimization problem, due to Yannakakis [11] (cf. also Garey and Johnson [4], problems [GT21] and [GT22]) is the following:

> **Instance:** A graph $G = (V, E)$, a positive integer $k \leq |V|$.
>
> **Question:** Is there a $V' \subseteq V$ with $|V'| \geq k$ such that the subgraph $G' = (V', E')$ of G induced by V' fulfills the property $\mathcal{P}$?

This problem was proven to be NP-hard for many graph properties, like clique, independent set, planarity, bipartiteness, etc.

3 Neighborhoods and Stars

In this section, we investigate the VC-dimension for *neighborhoods* and for *stars*. The notion of VC-dimension for neighborhoods was introduced by Haussler and Welzl [5].

Theorem 3.1 *There is an* $O(\min\{n^2 2^d, n^{\log n}\})$ *algorithm for computing a maximum size set of vertices shattered by neighborhoods in a graph* $G = (V, E)$*, where* $|V| = n$ *and* d *is the maximum degree of* G*. Hence, this algorithm is polynomial for maximum degree* $d = O(\log n)$*.*

Proof (Outline) A maximum size shattered set must be a subset of a neighborhood. There are as many neighborhoods as vertices, i.e. n, and each neighborhood has at most 2^d subsets. We can test if a given set of size $\leq d$ is shattered by neighborhoods in time $O(n2^d)$. This gives the $O(n^2 2^d)$ upper bound. The $n^{\log n}$ upper bound is obvious. ∎

Now we turn to the complexity of computing VC_{nbd} in general graphs (without bounded maximum degree). We show that this problem is LogNP-complete (for definitions see page 2 of this paper and [8]).

Theorem 3.2 *It is* LogNP*-complete to decide for a given graph* $G = (V, E)$ *and an integer* k*, whether* $\mathrm{VC}_{nbd}(G) \geq k$ *holds.*

Proof This problem is a subproblem of computing the general VC-dimension: Just set $X = V$ and let $\mathcal{F}$ contain all sets $N(v)$, $v \in V$ (the number of sets is polynomial in the input length). Hence, the problem is in LogNP and it remains to prove LogNP-hardness.

Consider an instance of the problem by Papadimitriou and Yannakakis, i.e. an enumeration of the sets in a family $\mathcal{F}$ of sets over X and an integer k. The question is to decide whether the VC-dimension of $\mathcal{F}$ is at least k. Without loss of generality we may assume that $|\mathcal{F}| = |X|$ (otherwise introduce new elements that do not occur in any set or introduce new sets that are empty). Define $n = |X|$ and $\ell = \lfloor \log n \rfloor$ and assume without loss of generality that $k \leq \ell$. From this set system, we construct a bipartite graph G with bipartition $B \cup W$ as follows. For every element $x \in X$, we introduce a vertex $b(x)$. For every set $C \in \mathcal{F}$, we introduce a vertex $w(C)$. There is an edge between a vertex $b(x)$ and a vertex $w(C)$ if and only $x \in C$. Moreover, we introduce a set B^* of ℓ vertices $b_1, \ldots, b_\ell$. For every subset B' of B^*, a vertex $w(B')$ is introduced and connected to all vertices in B'. For every subset B' and every set $C \in \mathcal{F}$, a vertex $w(B', C)$ is introduced and connected to all vertices corresponding to $B' \cup C$. Observe that $|W| = n + \ell < 2n$. We claim that $\mathrm{VC}_{nbd}(G) \geq \ell + k$ if and only if the VC-dimension of the set system $\mathcal{F}$ is at least k.

(If). Let X^* be a subset of X shattered by $\mathcal{F}$ with $|X^*| \geq k$. It is straight forward to check that then $X^* \cup B^*$ is shattered by the neighborhoods.

(Only if). A set N^* of cardinality $k + \ell$ that is shattered by the neighborhoods, is either a subset of B or a subset of W. In case N^* is a subset of B, it contains at least k vertices outside of B^*. Check that the elements in X that correspond to these k vertices are shattered by $\mathcal{F}$. In case N^* is a subset of B, $|W| \geq 2^{k+\ell}$ must hold just to shatter N^*. This yields $|W| \geq 2n$, a contradiction. ■

Next, we deal with the VC-dimension of set systems induced by stars. We start with a precise definition of the term 'star'.

Definition 3.1 *Given a graph $G = (V, E)$. For any vertex $u \in V$, a* star *of u is a subset of $\{v \mid (u, v) \in E\} \cup \{u\}$. An* open-star *of u is a subset of $\{v \mid (u, v) \in E\}$.*

Theorem 3.3 *If G is a graph with maximum degree d then*

1. $d \leq \mathrm{VC}_{star}(G) \leq d + 1$,
2. $d = \mathrm{VC}_{open-star}(G)$.

Proof Let $u \in V$ denote a vertex of degree d. Then the neighborhood of u, excluding u itself, is star-shattered. Depending on the graph, u might be added to the shattered set. This shows part 1 of the theorem. Part 2 is trivial. ■

Theorem 3.4 *There is an $O(nd^2)$ algorithm for computing a maximum size set of vertices shattered by stars, for an arbitrary graph with n vertices and maximum degree d.*

Proof Let G be an arbitrary graph with maximum degree d. For each vertex x, let $N'(x)$ (respectively, $N(x)$) be the open (respectively, closed) neighborhood of x, i.e. the set of vertices adjacent to x ($N'(x) \cup \{x\}$). Consider the set M of vertices of G of maximum degree. For each $x \in M$, the maximum size shattered set is either $N'(x)$ or $N(x)$. It is clear that every subset B of $N(x)$ such that $x \in B$ is shattered. It is therefore sufficient to find a "polynomial" condition guaranteeing that $N'(x)$ as well is shattered.

Let $x \in M$ be fixed. Define $X_0 := N'(x) := \{x_1, x_2, \ldots, x_d\}$ to be the set of neighbors of x. The idea of the algorithm is as follows. Check if there is a vertex $u \neq x$ such that $N'(x) = N'(u)$. If yes, the whole set $N(x)$ is shattered and $\mathrm{VC}_{star}(G) = d + 1$. If no, then compute the set $S_0 := \{v \in X_0 : X_0 \subseteq N(v)\}$. If S_0 is empty then $N(x)$ is not shattered. If $S_0 \neq \emptyset$ then any set $B \subseteq N'(x)$ such that $B \cap S_0 \neq \emptyset$ is shattered. Hence, it is enough to test whether or not all subsets of the set $X_1 = N'(x) \setminus S_0$ are shattered. Now replace X_0 by X_1 above and repeat.

Formally we define two sequences of sets: A strictly decreasing sequence of sets $X_0 := N'(x) \supset X_1 \supset \cdots \supset X_k$ and sets $S_0 \subseteq S_1 \subseteq \cdots \subseteq S_k \subseteq N'(x)$ such that $S_i := \{v \in X_i : X_i \subset N(v)\}$, for $i \leq k$, and $X_{i+1} := S_i \setminus X_i$, for $i < k$. It is clear that $k \leq d$. The algorithm is as follows.

Input: Vertex x of maximum degree d.
while $X_i + 1 \neq \emptyset$ **do**:

1. Check if there is a vertex $u \neq x$ such that $X_i \subseteq N'(u)$. If yes, the whole set $N(x)$ is shattered and $\mathrm{VC}_{star}(G) = d + 1$. If no, then go to step 2.

2. Compute the set $S_i := \{v \in X_i : X_i \subseteq N(v)\}$. If $S_i = \emptyset$ then $N(x)$ is not shattered. If $S_i \neq \emptyset$ then any set $B \subseteq N'(x)$ such that $B \cap S_i \neq \emptyset$ is shattered. Hence, it is enough to test whether or not all subsets of the set $X_{i+1} = S_i \setminus X_i$ are shattered.

3. $i := i + 1$ and repeat step 1.

Clearly, on any given input x the maximum number of iterations is d. Each step may take time $O(n)$. The above algorithm must be executed on all vertices of maximum degree d. It is easy to check that its complexity is $O(n^2 d)$.

We can improve this complexity to $O(nd^2)$ by using a more sophisticated data structure for testing "neighborhood equality", namely whether or not $N'(x) = N'(u)$, for $x \neq u$. The idea is to look at the adjacency matrix of the graph. Now neighborhood equality corresponds to equality of two rows of the adjacency matrix and can be tested in time linear in the number of edges of the graph, which is $O(nd)$. Since the above algorithm requires $O(d)$ iterations the proof of the theorem is complete. ■

For the case of planar graphs it is easy to see that the number of iterations of the previous algorithm is $O(1)$. Hence we can also claim as a corollary the following result.

Theorem 3.5 *There is an $O(n)$ algorithm for computing a maximum size set of vertices shattered by stars, for an arbitrary planar graph with n vertices.* ∎

4 Connected Sets and Trees

We derive three types of results in this section. First, we show that VC_{tree} and VC_{con} are identical. Secondly, we investigate the close relationship between VC_{con} and the so-called maximum leaf spanning tree. Thirdly, we prove that computing VC_{con} for a graph is NP-complete. Unless otherwise specified, in this section we will deal with VC dimensions for *connected* sets. Thus when speaking of a shattered set we always mean that it is shattered by connected sets.

Lemma 4.1 *For any graph $G = (V, E)$, $\mathrm{VC}_{tree}(G) = \mathrm{VC}_{con}(G)$ holds.*

Proof In view of Lemma 2.1 it is sufficient to show that $\mathrm{VC}_{tree}(G) \geq \mathrm{VC}_{con}(G)$. Consider a set $A \subseteq V$ which is shattered by connected subsets of G. Then for every $B \subseteq A$ there exists a connected set C with $B = C \cap A$. Replace the edges that connect C by a subset of these edges forming a spanning tree for C. The claim follows. ∎

Lemma 4.2 *For any tree T with l leaves, $\mathrm{VC}_{con}(T) = l$.*

Proof Let L denote the set of leaves of T. For any $B \subseteq L$ we can find a subtree of T whose leaves are exactly the elements of B. Thus L is shattered and $\mathrm{VC}_{con}(T) \geq l$. For proving $\mathrm{VC}_{con}(T) \leq l$, we consider an arbitrary set A which is shattered by connected sets of G. For each vertex u of T let $u_1, \ldots, u_k$ denote the children of u and $T_1, \ldots, T_k$ the corresponding subtrees of T rooted at $u_1, u_2, \ldots, u_k$ (see Figure 1). If $u \in A$ then there can be at most one index i, $1 \leq i \leq k$, with $T_i \cap A = \emptyset$ (two vertices $x \in T_i \cap A$ and $y \in T_j \cap A$ with $i \neq j$ cannot be connected without using u). ∎

There is a nice characterization of the VC_{path} dimension by relating it to maximum leaf spanning trees (abbreviated, MLST). A *maximum leaf spanning tree* is a spanning tree with a maximum number of leaves among all spanning trees.

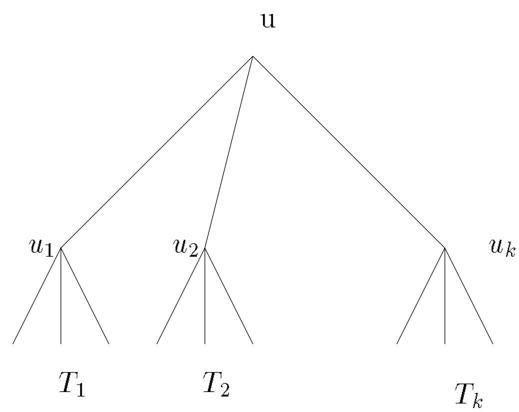

Figure 1: Illustration of the proof of lemma 4.2

Definition 4.1 *For any arbitrary graph G let*

$$l(G) := \max\{k \mid \textit{there exists a spanning tree } T \textit{ with } k \textit{ leaves}\}.$$

Theorem 4.3 *For any graph G*

$$l(G) \leq \mathrm{VC}_{con}(G) \leq l(G) + 1. \tag{2}$$

PROOF The inequality in the lefthand side follows from Lemma 4.2. To prove the inequality in the righthand side, consider a shattered set A of maximum cardinality. We show that there exists a spanning tree T with at least $|A| - 1$ leaves. Choose any vertex $r \in A$ as the root of T. Since A is shattered, there exists a path in G between any two vertices in A avoiding all the other vertices in A. Connect all $v \in A$, $v \neq r$ by these paths to r. This yields a connected subgraph G' of G where all vertices in $A \setminus \{r\}$ are of degree one. Destroy all cycles in G' by removing appropriate edges while keeping the subgraph connected. This eventually results in a tree T with $|A| - 1$ leaves. So far T is not necessarily a spanning tree. While there exist vertices not connected to the subgraph, perform the following procedure: Find an edge between some vertex that belongs to the subgraph and another vertex that does not belong to the subgraph and add it to the subgraph. This procedure cannot decrease the number of leaves in the tree (the just connected vertex always is a leaf). Finally, we will end up with a spanning tree with $|A| - 1$ leaves. ■

Let us define $\mathrm{VC}^k_{path}(G)$, as the maximum size of a set A of vertices of G such that for all subsets of A of size $\leq k$ there exists a path P such that $B = P \cap A$. It is clear that for all k, $\mathrm{VC}^{k+1}_{path}(G) \leq \mathrm{VC}^k_{path}(G)$. Moreover, the proof of Theorem 4.3 implies the following result.

Corollary 4.4 *For any graph G, $\mathrm{VC}^2_{path}(G) = \mathrm{VC}_{con}(G)$.* ■

This result does not generalize to $k \geq 3$: E.g. if T is a tree then $\mathrm{VC}^3_{path}(T) = 2$ whereas $\mathrm{VC}_{con}(T)$ equals the number of leaves.

It remains to determine when $\mathrm{VC}_{con}(G)$ and $l(G)$ differ by exactly one. The following theorem gives one possible characterization.

Theorem 4.5 *Let G be a graph and T any maximum leaf spanning tree on G whose set of leaves is L. Then $\mathrm{VC}_{con}(G) = |L| + 1$ if and only if the following condition is satisfied: "There exists a rooted subtree (i.e., all its leaves are also leaves in the original tree) T' of T (which may be T itself) with root r' whose set of leaves is L' and for all $l \in L$ and for all $l' \in L'$ there exists a path from l to l' in G avoiding r'."*

Proof We know from Lemma 4.2 that L is shattered. The condition above guarantees that r' can be included to the shattered set, because it is not necessary for shattering L. This proves the "if" part.

For the "only if" part we have to show that if $\mathrm{VC}_{con}(G) = |L| + 1$ then the condition is true for T. Therefore consider any shattered set A of maximum cardinality. We may assume that all elements in A but one, are leaves in T. To achieve this we argue as follows: The elements of A are vertices of the spanning tree T. Replace, one at a time, each element $a \in A$ which is not a leaf, by a leaf which has a path to a avoiding all remaining elements of A. It is easy to see that each time we do this the resulting set remains shattered. Finally, we end up with a shattered set A in which all but one element are leaves in T.

But now it is obvious that the condition in the statement of the theorem has to be true: The single non-leaf element r' gives rise to a rooted subtree with root r'. All its leaves must have paths to all $l \in L$ avoiding r' (since otherwise A would not be shattered). ■

Theorem 4.6 *It is* NP*-complete to decide for an input consisting of some graph $G = (V, E)$ and a number $k \geq 1$, whether $\mathrm{VC}_{con}(G) \geq k$ holds.*

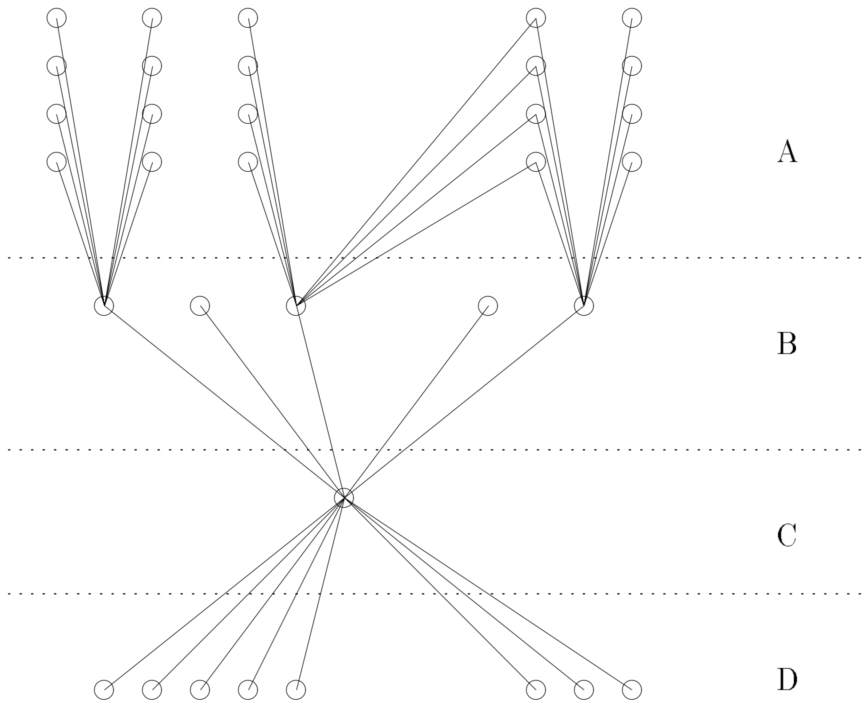

Figure 2: Construction of the graph

Proof First we show that the VC-problem is in NP: Just guess a subset $A \subseteq V$ with $|A| \geq k$. Corollary 4.4 allows us to check in polynomial time whether A is shattered: Test for all $O(n^2)$ possible pairs $a, b \in A$ if they are connected by a path avoiding A.

NP-hardness is proved by a transformation from the MINIMUM SET COVER problem [4], which is defined as follows: Given a finite set $S = \{a_1, \ldots, a_n\}$, a collection of m subsets $S_1, \ldots, S_m \subseteq S$ and an integer $t \leq m$, one wants to know whether there exists an index set $I \subseteq \{1, \ldots, m\}$ such that $|I| \leq t$ and $\bigcup_{i \in I} S_i = S$.

Consider the following graph $G = (V, E)$ for a given instance of the MINIMUM SET COVER problem: The set of vertices is given by four pairwise disjoint sets A, B, C and D with $V = A \cup B \cup C \cup D$ (see figure 2 for an illustration). A has $n \cdot (m + 1)$ vertices, arranged in n columns of

$m + 1$ vertices each. $B = \{v_1, \ldots, v_m\}$, where v_i corresponds to the set S_i. C consists of only one vertex and D of $m + 2$ vertices. The vertices are connected as follows: The vertex from C is directly connected to all vertices in B and D. There are edges between $v_i \in B$ and all vertices of the jth column of A iff $a_j \in S_i$. Note that $|V| = mn + 2m + n + 3$.

We claim that $\mathrm{VC}_{con}(G) \geq |V| - (t + 1)$ if and only if the instance of MINIMUM SET COVER has a solution. (If). Assume that there exists a solution $I \subseteq \{1, \ldots, m\}$ of the MINIMUM SET COVER instance. Then the set $V' = D \cup A \cup \{v_i \in B \mid i \notin I\}$ is shattered in G: Every vertex in V' is connected to C by a path that avoids V'. This trivially holds for the vertices in B and in D. For every vertex in A, there is an adjacent vertex in $v_i \in B \setminus V'$ (since the corresponding element in S is contained in some S_i with $i \in I$) and thus it is connected to C via this vertex v_i. Every subset of V' can be covered by the corresponding set of paths. This set of paths is connected (via the vertex in C) and avoids $V \setminus V'$.

(Only if). Consider a shattered set V' with $V' \geq |V| - t - 1$. How does V' look like? The single vertex in C cannot be in V' since otherwise none of the $m + 2 > t + 1$ vertices in D could be included in V'. Hence, we may assume w.l.o.g. that $D \subseteq V'$. Since all elements in any fixed column in A have identical neighborhoods, we may assume that either all or no elements from any column are in V'. Since every column contains $m + 1$ vertices, it follows that *all* elements in *all* columns have to be in V' (otherwise at least $m + 2 \geq t + 2$ vertices would be outside of V'). Now $A \subseteq V'$, $D \subseteq V'$ and $C \not\subseteq V'$, and consequently at least $m - t$ vertices $\in B$ have to be in V'. Since V' is shattered, for every vertex in A there must exist an adjacent $v_i \in B \setminus V'$. With this it is straightforward to see that $I = \{i | v_i \in B \setminus V'\}$ constitutes a solution for the given instance of the MINIMUM SET COVER problem. ■

Finally, we observe that in case the number k is *not part of the input* the problem is solvable in polynomial time. Simply check all $O(n^k)$ subsets on k vertices whether they are shattered. Checking shatteredness can be done efficiently with the help of Corollary 4.4.

5 Paths

We derive three types of results in this section. First, we give a precise characterization of all graphs fulfilling $\mathrm{VC}_{path}(G) = 3$. Then we derive

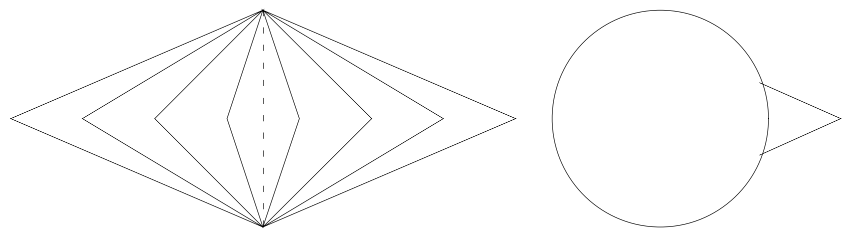

Figure 3: Graphs G with $\mathrm{VC}_{path}(G) = 3$.

upper and lower bounds for the number $|E|$ of edges in terms of the number $|V|$ of vertices and the VC-dimension $\mathrm{VC}_{path}(G)$. Finally, we deal with the computational complexity of computing VC_{path}. This problem is NP-hard in general, but it can be solved in polynomial time if the dimension is not part of the input.

5.1 Graphs with VC_{path} Dimension Three

In this subsection we characterize those connected graphs $G = (V, E)$ with $\mathrm{VC}_{path}(G) = 3$ (observe that $\mathrm{VC}_{path}(G) = 2$ iff G is a tree).

Theorem 5.1 *The graphs G having $\mathrm{VC}_{path}(G) = 3$ are the graphs depicted in Figure 3, where from each of the vertices may emanate trees and the cycles depicted in the right hand side are adjacent on a single edge.*

The proof of the theorem will follow after several lemmas. First of all observe the following result.

Lemma 5.2 *For a graph on n vertices,*

1. $\mathrm{VC}_{path}(G) = 2$ *if and only if G is a tree.*
2. $\mathrm{VC}_{path}(G) = n$ *if and only if $G = K_n$.*

Proof Immediate. ∎

Let G be a connected n-vertex graph with $\mathrm{VC}_{path}(G) = 3$. In view of Lemma 5.2 G must have a cycle. All the cycles we consider in the sequel are simple.

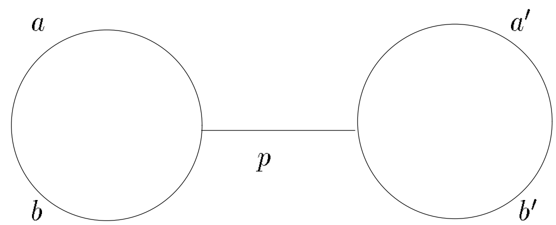

Figure 4: Two edge-disjoint cycles and a shattered set of size 4.

Lemma 5.3 *Any two cycles have at least two vertices in common.*

PROOF If the cycles are not edge-disjoint cycles the theorem is obvious. Without loss of generality assume that the two cycles, say C, C', are edge-disjoint. Assume on the contrary that $|C \cap C'| \leq 1$. As in Figure 4 the cycles are connected by a path p which might be of length 0. Pick vertices a, b in $C \setminus C'$ and a', b' in $C' \setminus C$. It is easily checked that the set $\{a, b, a', b'\}$ is shattered by paths. This proves the lemma. ■

Lemma 5.4 *Any two edge-disjoint cycles C, C' must have exactly two vertices in common and either $(|C|, |C'|) = (3, 4)$ or $(|C|, |C'|) = (4, 4)$.*

PROOF Let C, C' be the cycles. In view of Lemma 5.3 $|C \cap C'| \geq 2$. First of all consider the case where $|C \cap C'| \geq 3$. In this case it is easy to find two edge-disjoint cycles C_1, C_1' such that $|C_1 \cap C_1'| \geq 1$, contradicting Lemma 5.3. This proves that $|C \cap C'| = 2$. Suppose that none of the cycles has size 4. We consider two cases. First suppose that C' is of size 3. In this case and since C, C' have two vertices in common and are edge-disjoint it must be the case that C has size at least 5. It is easy to see from Figure 5 that in this case the set $\{a, b, u, v\}$ is path-shattered, thus contradicting the assumption that the graph has VC_{path}-dimension equal to 3. Hence, without loss of generality we may assume that C' has size at least 5. Let us consider the case where C has size 3. The case when C has size ≥ 5 is similar. Consider Figure 6. The vertex v must exist because the two cycles share no edge. But then it is clear that the set $\{a, b, u, v\}$ is path-shattered, thus again contradicting the assumption that the graph has VC_{path}-dimension equal to 3. This completes the proof of the lemma. ■

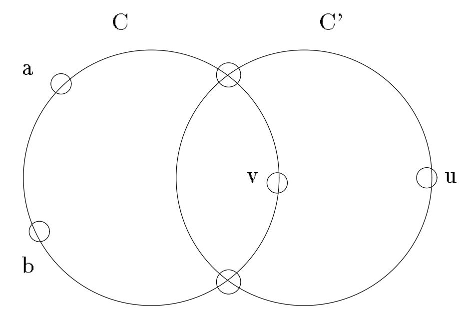

Figure 5: A shattered set of size 4.

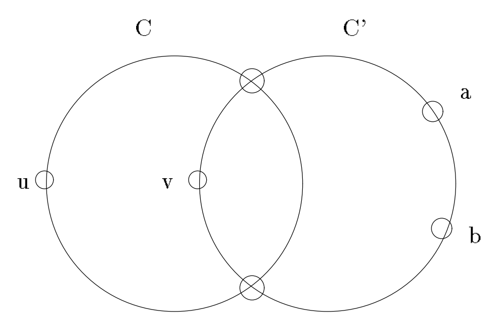

Figure 6: Graph with a shattered set of size 4.

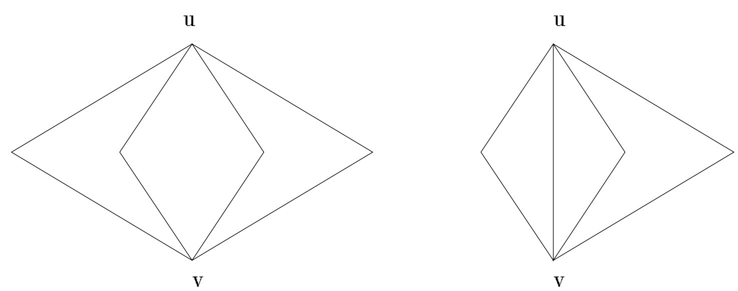

Figure 7: Graphs with a shattered set of size 4.

Lemma 5.5 *The edge-disjoint cycles have two vertices in common, thus forming the configuration depicted in the left side of Figure 3, where the dashed edge is optional.*

PROOF Take any two edge-disjoint cycles C, C'. Their sizes are either $(3,4)$ or $(4,4)$. It is easy to see that this gives rise to one of the configurations depicted in Figure 7. We must show that the all other cycles must pass through the vertices u, v. However it is not hard to check that the addition of any path between two vertices (that uses a vertex other than either u or v) to the graph will create a graph with VC_{path} greater or equal to 4. This proves the lemma. ■

PROOF of Theorem 5.1. Consider cycles that have edges in common. If each cycle has at least two vertices not belonging to the other then as in the proof depicted in Figure 5 we can find a set of size 4 which is shattered by paths. If the cycles have more than one edge in common then as in the proof depicted in Figure 6 we can find a set of size 4 which is shattered by paths. It follows that the only possible configurations are the ones depicted in Figure 3. This completes the proof of Theorem 5.1.

5.2 VC_{path} and the number of edges

For each $k \leq n$ let e_k be the minimal number of edges of a connected graph G with $\mathrm{VC}_{path}(G) \geq k$. It is clear that $e_2 = n-1$ and $e_n = n(n-1)/2$. We can prove the following result.

Theorem 5.6 $\Omega(n^2/k^2) \leq e_{n-k} \leq O(n^2/k)$.

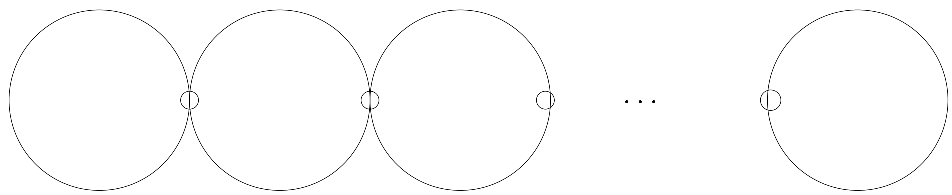

Figure 8: A graph G with $\mathrm{VC}_{path}(G) \geq k$.

PROOF First we prove the upper bound. Consider the graph of Figure 8. It consists of k complete graphs. They are connected at the $k-1$ vertices indicated. The first $k-1$ graphs each consist of $\lfloor n/k \rfloor$ vertices; the kth graph has $n-(k-1)\lfloor n/k \rfloor$ vertices. The total number of vertices is

$$(k-1)\frac{\lfloor n/k \rfloor(\lfloor n/k \rfloor - 1)}{2} + \frac{(n-(k-1)\lfloor n/k \rfloor)(n-(k-1)\lfloor n/k \rfloor) + 1}{2}.$$

Since this last term is $O(n^2/k)$ the proof of the upper bound is complete.

To prove the lower bound we argue as follows. Let G be a graph such that $\mathrm{VC}_{path}(G) = n-k$ and let A be a set of vertices of size $n-k$ which is shattered by paths. Choose integers $d_1, d_2, \ldots, d_k \leq n$; their precise value will be determined in the sequel. If there exist $k+1$ vertices of degree $< d_1$ then pick a vertex $v_1 \in A$ of degree $< d_1$. Otherwise the graph has $\Omega((n-d_1)d_1)$ edges. If there exist $n-d_1-k$ vertices of degree $< d_2$ then pick a vertex $v_2 \in A$ of degree $< d_2$ non-adjacent to v_1. Otherwise the graph has $\Omega(n-d_1-k)d_2)$ edges.

Proceed by induction to construct a sequence $v_1, v_2, \ldots, v_k$ of pairwise non-adjacent vertices in A. Assuming $v_1, \ldots, v_i$ have been constructed we construct v_{i+1} as follows. If there exist $n-d_1-d_2-\cdots-d_i-k$ vertices of degree $< d_{i+1}$ then pick a vertex $v_{i+1} \in A$ which is not adjacent to any of the vertices $v_1, \ldots, v_i$. Otherwise the graph has $\Omega(n-d_1\cdots-d_i-k)d_{i+1})$ edges. Since the graph is shattered it follows that any two vertices not adjacent to any of $v_1, v_2, \ldots, v_k$ must be adjacent. This gives a total of

$$\Omega\left(\binom{n-d_1-\cdots-d_k-k}{2}\right)$$

edges.

If we let $d_i = n/(k-1)$ then we obtain the desired lower bound. ■

5.3 Complexity of Computing VC_{path}

We do not know whether computing VC_{path} is in the class NP (since there is an exponential number of conditions that have to be checked in order to verify that some set is shattered). Hence, we only prove NP-hardness of the problem.

Theorem 5.7 *It is* NP*-hard to decide for an input consisting of some graph $G = (V, E)$ and a number $k \geq 1$, whether $\mathrm{VC}_{path}(G) \geq k$ holds.*

Proof The proof is done by a transformation from the HAMILTONIAN PATH problem in bipartite graphs: Given an undirected, bipartite graph $H = (B \cup W, F)$ with $F \subseteq B \times W$ and with $|B| = a + 1$, $|W| = a$, the question is to decide whether there exists a Hamiltonian path for H (i.e. a path that visits every vertex exactly once). This problem is known to be NP-complete [4].

We construct from H another undirected graph G as follows. For every vertex $b \in B$, we introduce a new vertex b^*. This new set of vertices is denoted by B^*, $|B^*| = a + 1$. We connect every vertex $b \in B$ by an edge with its corresponding vertex b^*. Moreover, we connect all vertices in $B^* \cup W$ to each other. The resulting graph is graph G. We claim that $\mathrm{VC}_{path}(G) \geq 2a + 2$ if and only if H possesses a Hamiltonian path.

(If). Assume that H has a Hamiltonian path. Then the set $B \cup B^*$ is shattered: Consider arbitrary subsets $X \subseteq B$ and $X^* \subseteq B^*$. We must show that there is a path in $W \cup X \cup X^*$ that contains all of $X \cup X^*$. If $X = B$, we use the Hamiltonian path and append a path through X^* to the Hamiltonian path. If $X \neq B$, we select for every vertex in X its two incident edges in the Hamiltonian path (respectively, we select its unique incident edge, if it is an endvertex of the path). Since $X \neq B$, some of the selected edges do not meet other selected edges in W but have dangling ends. Select a Hamiltonian path for X^* with edges from $X^* \times X^*$ and paste it between two of these dangling ends (be careful not to form a cycle). Finally, select an appropriate subset of the edges in $W \times W$ to get a complete path.

(Only if). Now assume that there is a shattered set V' with $|V'| = 2a+2$. Define $B' = B \cap V'$, $|B'| = x$. Then $W \cup B^*$ contains $2a + 2 - x$ vertices in V' and $x - 1$ vertices that are not in V'. First suppose that $x = 1$ holds. Then $V' = W \cup B^* \cup \{b_1\}$ for some $b_1 \in B$. But now for any $b_2 \in B$, $b_2 \neq b_1$, the set $\{b_1, b_2^*\}$ is a subset of V' but not covered by any path that avoids V'; a contradiction. Hence, $x \geq 2$ must hold.

Since V' is shattered, there exists a path P through B' that avoids the rest of V'. Since $B' \subseteq B$ is an independent set, there are at least $x-1$ vertices on this path P that do not belong to V' and all of them must be in $W \cup B^*$. Since $|(W \cup B^*) \setminus V^*| = x - 1$, the path contains *all* vertices that are not in V'. Since $x \geq 2$, all these $x-1$ vertices must have two neighbors in B' and consequently, they are not in B^* (every vertex in B^* has a unique neighbor in B). Hence, $B^* \subseteq V'$ must hold.

Now suppose that there exists a $b \in B$, $b \notin V'$. Consider the set $B' \cup \{b^*\} \subseteq V'$. This set is independent and of cardinality $x+1$. Then any path through this set must contain at least x vertices in $W \cup B^*$ and therefore it cannot avoid V^*. This is a contradiction and yields $B \subseteq V'$.

Finally, consider the set $B \subseteq V'$. Since V' is shattered, there exists a path through B that avoids B^*. All $|B| - 1 = |W|$ intermediate vertices of this path must be in W. Clearly, this yields a Hamiltonian path for H. ■

Theorem 5.8 *For any fixed number k (that is not part of the input), the problem of deciding whether* $\mathrm{VC}_{path}(G) \geq k$ *for some input graph $G = (V, E)$ is solvable in polynomial time.*

Proof The idea is to check all $O(n^k)$ subsets of V whether they are shattered. For a fixed subset V', $|V'| = k$, we must check all 2^k subsets $W \subseteq V'$ whether there is a cycle through W that avoids V'. This problem is just a special case of the FIXED-VERTEX SUBGRAPH HOMEOMORPHISM(H) problem (i.e. given a graph $G_1 = (V_1, E_1)$ and a map f from the vertices of the fixed pattern graph H to the vertices of G_1, does G_1 contain a homeomorphic image of H in which each vertex of H is identified with its image under f?). The FIXED-VERTEX SUBGRAPH HOMEOMORPHISM(H) problem is solvable in polynomial time (cf. Johnson [7]).

Since for a constant number k all involved numbers are polynomial in n, this yields a polynomial time algorithm for computing $\mathrm{VC}_{path}(G)$. ■

6 Cycles

We say that every single vertex $v \in V$ constitutes a cycle of length one, and that every edge $[u, v] \in E$ constitutes a cycle of length two spanning the vertices u and v.

Lemma 6.1 *For a graph on n vertices,*

1. $\mathrm{VC}_{cycle}(G) = 1$ *if and only if E is empty.*

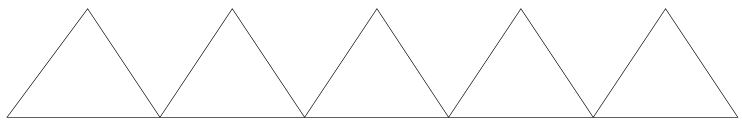

Figure 9: A graph G for which $\text{Cycles}(G)$ is acyclic.

2. $\text{VC}_{cycle}(G) = n$ *if and only if* $G = K_n$.

3. $\text{VC}_{cycle}(G) \leq n$. ■

For any graph G let $\text{Cycles}(G)$ be the graph of cycles of G: its vertices are the simple cycles of G and two cycles are adjacent if and only if they have at least one vertex of G in common.

Theorem 6.2 *If* $\text{Cycles}(G)$ *has no triangles then* $\text{VC}_{cycle}(G) \leq 2$.

Proof Assume that on the contrary $\text{VC}_{cycle}(G) \geq 3$ holds and take a set A of three vertices which is shattered by cycles. For any two element subset of A there is a cycle passing through these two elements and avoiding the third. The resulting three cycles form a triangle, a contradiction. ■

There are graphs for which the quantities $\text{VC}_{cycle}(G)$ and $\text{VC}_{path}(G)$ are arbitrarily far apart. For example, the graph G depicted in Figure 9 has $\text{VC}_{cycle}(G) = 2$ (by Lemma 6.2) but $\text{VC}_{path}(G) \geq k + 2$, where k is the number of triangles.

For each $k \leq n$ let $\bar{e}_k$ be the minimal number of edges of a connected graph G with $\text{VC}_{cycle}(G) \geq k$. It is clear that $\bar{e}_1 = n - 1, \bar{e}_2 = n$ and $\bar{e}_n = n(n-1)/2$. Moreover, $\bar{e}_k \leq e_k$, for all k. As in the proof of Theorem 5.6 we can prove the following result.

Theorem 6.3 $\Omega(n^2/k^2) \leq \bar{e}_{n-k} \leq O(n^2/k)$.

Proof (Outline) For the upper bound we use the idea of Figure 8 but this time we connect the complete subgraphs into a cycle. The lower bound proof is the same as before. ■

7 VC Dimensions for Edges

So far we considered the VC-dimensions in a graph only for vertices. Next, we give an analogous definition for edges.

Definition 7.1 *Let $G = (V, E)$ be a graph and let $\mathcal{P}$ be a family of sets of edges of the graph. We say that a subset $A \subseteq E$ is $\mathcal{P}$-*edge-shattered *(or shattered by sets of edges) if and only if for all $B \subseteq A$ there exists a set $C \subseteq E$ satisfying property $\mathcal{P}$ such that $B = C \cap A$. Then the* EVC-*dimensions of G with respect to $\mathcal{P}$ are given by*

$$\mathrm{EVC}_{\mathcal{P}}(G) := \max\{|A| : A \textit{ is P-edge-shattered } \} \tag{3}$$

Defining the EVC-dimensions for connectedness, trees, paths and stars yields $\mathrm{EVC}_{con}(G)$, $\mathrm{EVC}_{tree}(G)$, $\mathrm{EVC}_{path}(G)$ and $\mathrm{EVC}_{star}(G)$. If clear from the context we will simply say "shattered" instead of "edge-shattered".

Lemma 2.1 also holds for the EVC-dimensions. However EVC_{tree} and EVC_{con} are not necessarily equal. For example $\mathrm{EVC}_{tree}(C_3) = 2$, but $\mathrm{EVC}_{con}(C_3) = 3$, where C_3 is the cycle on three vertices. For $n \geq 3$, $\mathrm{EVC}_{tree}(C_n) = \mathrm{EVC}_{con}(C_n)$. Also, as in Example 2.1, we get that $\mathrm{EVC}_{con}(P_n) = 2$. However for complete graphs the situation is quite different as the following theorem indicates.

Theorem 7.1 *Let K_n be the complete graph on n vertices, $n \geq 5$. Then*

$$\mathrm{EVC}_{con}(K_n) = \frac{n(n-1)}{2} - (n-2).$$

PROOF First we show that $\mathrm{EVC}_{con}(K_n) \geq \frac{n(n-1)}{2} - (n-2)$. Consider a set $B = \{e_1, \ldots, e_{n-1}\}$ of all edges being adjacent to one arbitrary vertex u. Let A denote the set of all remaining edges of K_n. Clearly $|A| = \frac{n(n-1)}{2} - (n-1)$. For any subset $\{c_1, \ldots, c_k\} \subseteq A$ we can find for every c_i an edge $e_{j_i} \in B$ connecting c_i to u. Now we can choose $C = \{c_1, \ldots, c_k\} \cup \{e_{j_1}, \ldots, e_{j_k}\}$ and we get $C \cap A = \{c_1, \ldots, c_k\}$. Thus A is shattered. However we can add one more edge to A because in the construction above we have for every c_i two choices for e_{j_i}. Thus one element from B can be moved to A.

With the help of Turan's theorem we can now show that a set of cardinality greater than $\frac{n(n-1)}{2} - (n-2)$ cannot be shattered. This theorem states that a simple graph on n vertices, having more than $\frac{p-2}{2(p-1)}n^2$ edges contains K_p (see for example [10]). If we consider an arbitrary graph $G = (V, E)$

with $|V| = n$ and $|E| = \frac{n(n-1)}{2} - (n-3)$ we can apply Turan's theorem with $p = n-1$ because then $|E| > \frac{p-2}{2(p-1)}n^2$. Hence G contains K_{n-1} and there is one vertex u of degree 2, since $\frac{n(n-1)}{2} - (n-3) - \binom{n-1}{2} = 2$.

Now let us assume that we could shatter a set of cardinality $|E| = \frac{n(n-1)}{2} - (n-3)$ in K_n. As we have shown there must be in this case a vertex u which is adjacent to exactly two edges (u, u_1) and (u, u_2) in E. Note that the edge (u_1, u_2) must be in E. Now take an edge $(u_3, u_4) \in E$ not adjacent to (u_1, u_2). For $n \geq 5$ such an edge always exists. This edge cannot be connected to (u_1, u_2) without using other edges from E. Thus a set of cardinality $|E|$ cannot be shattered. ∎

For any connected graph, the EVC-dimensions for trees and connected sets can lie only within a small interval, as the follwing theorems indicate.

Theorem 7.2 *For any graph G with n vertices,*

$$n - 1 \geq \mathrm{EVC}_{tree}(G) \geq \mathrm{VC}_{tree}(G) - 1$$

Proof First we prove the lower bound. Consider a spanning tree of G with $l(G)$ leaves. Taking the edges associated to these leaves the result follows from Theorem 4.3 immediately. The upper bound is trivial since a shattered set must be a tree and cannot have more than $n-1$ edges. ∎

Theorem 7.3 *For any graph G with n vertices,*

$$|E| \geq \mathrm{EVC}_{con}(G) \geq |E| - (n-1).$$

Proof The upper bound is trivial. For the lower bound consider a spanning tree T through all n vertices. Then all remaining $|E| - (n-1)$ edges can be shattered, since each of these edges is adjacent to T. Thus any two edges not in T can be connected via T. ∎

The precise relationship between the VC-dimensions and the EVC-dimensions can be characterized via line graphs. For a good overview on line graphs see e.g. [6]. We say $L(G) = (V^*, E^*)$ is the line graph of $G = (V, E)$, if $V^* = E$ and $E^* = \{\{e_1, e_2\} | e_1, e_2 \in E \text{ and } e_1, e_2 \text{ have a common vertex}\}$

Lemma 7.4 *If $L(G)$ is the line graph of a graph G, then*

$$G \text{ is connected} \Leftrightarrow L(G) \text{ is connected}$$

Proof By definition in a connected graph there exists a path between any pair of vertices. Thus also any pair of edges is connected. By the definition of a line graph, it immediately follows that two edges are connected in G if and only if the corresponding vertices in $L(G)$ are connected. ■

Now it is easy to show that the EVC_{con}-dimensions of a graph are equal to the VC_{con}-dimensions of its line graph.

Theorem 7.5 *If $L(G)$ is the line graph of a graph G then*

$$\mathrm{VC}_{con}(L(G)) = \mathrm{EVC}_{con}(G)$$

Proof We only show that $\mathrm{VC}_{con}(L(G)) \geq \mathrm{EVC}_{con}(G)$. The $\leq$ case can be done in a very similar way.

Consider a set $A^* \subseteq V^*$, which is shattered in $L(G)$. This means by definition that for all $B^* \subseteq A^*$ there exists a connected C^* with $B^* = A^* \cap C^*$. Now consider B and C. Because of Lemma 7.4 also C is connected. It follows that also A is shattered. ■

Theorem 7.6 *It is* NP*-complete to decide for an input consisting of some graph $G = (V, E)$ and a number $k \geq 1$, whether* $\mathrm{EVC}_{con}(G) \geq k$ *holds.*

Proof The proof is very similar to the NP-completeness proof for VC_{con} in Theorem 4.6. We use the same notation as in Theorem 4.6 and just sketch the construction.

For a MINIMUM SET COVER instance, we construct a graph $G = (V, E)$ consisting of four parts A, B, C and D. Part C contains two vertices c_1 and c_2 that are connected by an edge. Part D contains $m + 2$ vertices that are all connected to vertex c_2. Part B has vertex set $\{v_1, \ldots, v_m\}$, where v_i corresponds to the set S_i; every v_i is connected to c_1. Finally, part A consists of n stars with roots $\{w_1, \ldots, w_n\}$. Every star has $m + 1$ edges. There is an edge between v_i and w_j iff $a_j \in S_i$. It can be shown that $\mathrm{EVC}_{con}(G) \geq |E| - (n + t + 1)$ if and only if the MINIMUM SET COVER instance has a solution. ■

8 Conclusion

We have investigated several set systems resulting from special graph properties of simple loopless graphs and the associated VC-dimensions for vertices (like VC_{con}, VC_{path}, VC_{star}, VC_{tree}, VC_{nbd} and VC_{cycle}) as well as for edges (like EVC_{con}, EVC_{path}, EVC_{star}, EVC_{tree}). We studied the computational complexity of $VC_{\mathcal{P}}$ for several graph properties $\mathcal{P}$ and showed that they all are NP-hard with the exception of the neighborhood property (which is complete for the class LogNP) and the star property (which is computable in polynomial time). We derived several combinatorial properties of these set systems and related them to special graph parameters (like the maximum number of leaves in any spanning tree). In addition, for the path and cycle properties we constructed graphs G with the minimum number of edges under the condition $VC_{\mathcal{P}}(G) \geq k$.

This paper is just a first step towards a systematic investigation of the Vapnik-Chervonenkis dimension on graphs. Problems that deserve further studies are e.g. the investigation of set systems induced by other graph properties (like planarity, bounded genus, k-connectivity, bounded diameter, k-colorability, or forbidden subgraphs) or the problem of determining the complexity of computing VC_{con}, VC_{path}, etc. for specially structured graph classes (like interval graphs, cographs, partial k-trees, or planar graphs).

References

[1] M. Anthony, G. Brightwell, and C. Cooper, "On the Vapnik-Chervonenkis Dimension of a Graph", Technical Report, London School of Economics, 1993.

[2] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, "Learnability and the Vapnik-Chervonenkis Dimension", Journal of the ACM, 36: 929–965, 1989.

[3] B. Chazelle and E. Welzl, "Quasi-Optimal Range Searching and VC-Dimensions", Discrete & Computational Geometry, 4: 467–490, 1989.

[4] M. R. Garey and D. S. Johnson, "Computers and intractability" Freeman, San Francisco, 1979.

[5] D. Haussler and E. Welzl, "Epsilon-nets and Simplex Range Queries", Discrete & Computational Geometry, 2: 127–151, 1987.

[6] R. L. Hemminger and L. W. Beineke, "Line Graphs and Line Digraphs", in: L.W. Beineke and R.J. Wilson, editors, Selected topics in graph theory, pages 271–305, Academic Press, London, 1978.

[7] D. S. Johnson, "The NP-completeness column: An ongoing guide", Journal of Algorithms, 8: 285–303 (1987).

[8] C. H. Papadimitriou and M. Yannakakis, "On limited nondeterminism and the complexity of VC-dimension", Proceedings of the eighth annual Conference on Structure in Complexity Theory, 12–18, IEEE, 1993.

[9] H.-I. Lu and R. Ravi, "The Power of Local Optimization: Approximation Algorithms for Maximum-Leaf Spanning Tree", Proceedings of 1992 Allerton Conference.

[10] J. H. van Lint and R. M. Wilson "A Course in Combinatorics" Cambridge University Press.

[11] M. Yannakakis, "Node- and Edge-deletion NP-complete Problems", Proc. 10th Annual ACM Symposium on Theory of Computing, Association for Computing Machinery, New York, 253–264, 1978.

[12] V. N. Vapnik and A. 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.