The VC-Dimension of Graphs with Respect to k-Connected Subgraphs

Andrea Munaro*

Abstract

We study the VC-dimension of the set system on the vertex set of some graph which is induced by the family of its k-connected subgraphs. In particular, we give upper and lower bounds for the VC-dimension. Moreover, we show that computing the VC-dimension is NP-complete and that it remains NP-complete for planar graphs in the case $k = 2$. This is done by a reduction from a variant of Planar 1-In-3-Sat which we prove to be NP-complete.

1 Introduction

The notion now called VC-dimension of a set system was introduced by Vapnik and Chervonenkis [12]. The initial interest was in the contexts of empirical process theory and learning theory, where it proved to be a fundamental concept. It represents a prominent measure of the "complexity" of the set system. Let $\mathcal{H}$ be a set system on a finite set X. A subset $Y \subseteq X$ is *shattered by* $\mathcal{H}$ if $\{E \cap Y : E \in \mathcal{H}\} = 2^Y$. The *VC-dimension of* $\mathcal{H}$ is defined as the maximum size of a set shattered by $\mathcal{H}$. One might think to apply the abstract notion of VC-dimension to some concrete settings. A natural choice is the study of the VC-dimension associated to graphs. Given a graph, we consider set systems induced by a certain family of subgraphs. In this way we obtain several different notions of VC-dimension, each one related to a special family of subgraphs. This study was first initiated in a seminal paper by Haussler and Welzl [6]. They considered the set system induced by closed neighbourhoods of the vertices. Kranakis et al. [8] investigated the VC-dimensions induced by other families of subgraphs. They adapted the definition of VC-dimension to the graph theoretic setting as follows.

Definition 1. Let $G = (V, E)$ be a graph and let $\mathcal{P}$ be a family of subgraphs of G. A subset $A \subseteq V$ is $\mathcal{P}$-shattered if every subset of A can be obtained as the intersection of $V(H)$, for $H \in \mathcal{P}$, with A. The VC-dimension of G with respect to $\mathcal{P}$, denoted by $\mathrm{VC}_{\mathcal{P}}(G)$, is defined as the maximum size of a $\mathcal{P}$-shattered subset.

*Department of Computer Science and Hausdorff Center for Mathematics, University of Bonn. Email: aerdna.munaro@gmail.com

Note that in this paper we consider only finite undirected simple graphs and we use standard graph theoretic terminology from [13], unless stated otherwise.

According to Definition 1, we denote by $\mathrm{VC}_{\mathrm{tree}}$, $\mathrm{VC}_{\mathrm{con}}$, $\mathrm{VC}_{k-\mathrm{con}}$, $\mathrm{VC}_{\mathrm{nbd}}$, $\mathrm{VC}_{\mathrm{path}}$, $\mathrm{VC}_{\mathrm{cycle}}$ and $\mathrm{VC}_{\mathrm{star}}$ the VC-dimensions with respect to families of subgraphs which are tree, connected, k-connected, closed neighbourhood, path, cycle and star, respectively. Note that the VC-dimension with respect to some families of subgraphs is equal to well established quantities in graph theory: if $\mathcal{P}$ is the family of complete subgraphs then $\mathrm{VC}_{\mathcal{P}}$ is the clique number, while if $\mathcal{P}$ is the family of subgraphs induced by independent sets then $\mathrm{VC}_{\mathcal{P}}$ is the independence number.

Since a graph of order n has n closed neighbourhoods, then its VC-dimension is at most $\lfloor \log_2 n \rfloor$ [6]. It is not difficult to show that this bound is tight [1]. Indeed, consider the graph H built as follows. Take a set S of $\lfloor \log_2 n \rfloor$ independent vertices. For each non-singleton subset $R \subseteq S$, add a vertex v_R adjacent to precisely the vertices of R. The resulting graph H has at most n vertices and $\mathrm{VC}_{\mathrm{nbd}}(H) = \lfloor \log_2 n \rfloor$. If G is a graph with maximum degree Δ, then it is easy to see that $\Delta \leq \mathrm{VC}_{\mathrm{star}} \leq \Delta + 1$ [8]. The VC-dimension with respect to trees is the same as the VC-dimension with respect to connected subgraphs [8]. This is an immediate consequence of the fact that a connected graph contains a spanning tree.

Kranakis et al. [8] related the VC-dimension of a graph G with respect to connected subgraphs to the number of leaves $\ell(G)$ in a maximum leaf spanning tree of G.

Theorem 2 (Kranakis et al. [8]). *$\ell(G) \leq \mathrm{VC}_{\mathrm{con}}(G) \leq \ell(G) + 1$, for any graph G.*

From another perspective, a natural question is to investigate the computational complexity of computing $\mathrm{VC}_{\mathcal{P}}(G)$ for a given graph G and a family of its subgraphs $\mathcal{P}$. We formulate the decision problem as follows.

Graph $\mathrm{VC}_{\mathcal{P}}$ Dimension	
Instance:	A graph G and a number $s \geq 1$.
Question:	Does $\mathrm{VC}_{\mathcal{P}}(G) \geq s$ hold?

1.1 Our results

In Section 2 we extend Theorem 2 by giving upper and lower bounds on the VC-dimension with respect to k-connected subgraphs, for $k \geq 2$. These, similarly to Theorem 2, are given in terms of the number of leaves in a maximum leaf spanning tree. In Section 3 we prove that the related decision problem Graph $\mathrm{VC}_{k-\mathrm{con}}$ Dimension is NP-complete. Moreover, we show that Graph $\mathrm{VC}_{2-\mathrm{con}}$ Dimension remains NP-complete for planar graphs. In order to do this we first introduce a variant of Planar Monotone 3-Sat [2] and we show that it is NP-complete. The following table summarizes the known results about the complexity of Graph $\mathrm{VC}_{\mathcal{P}}$ Dimension.

Family $\mathcal{P}$	Graph G	Computational Complexity	Reference
star		in P	Kranakis et al. [8]
neighbourhood		LOGNP-complete	Kranakis et al. [8]
path		Σ_3^p-complete	Schaefer [10]
cycle		Σ_3^p-complete	Schaefer [10]
k-connected		NP-complete	Theorem 7
2-connected	planar	NP-complete	Theorem 12

2 Bounds on the VC-dimension

We extend Theorem 2 by considering families of k-connected subgraphs, for $k \geq 2$. Concerning the upper bound, the idea is to construct a spanning tree with at least $\mathrm{VC}_{k-\mathrm{con}}(G) + k - 1$ leaves. We fix a shattered set A of maximum cardinality and choose an arbitrary vertex $r \in A$ as the root. Then we consider some k neighbours of r in A, say $u_1, \ldots, u_k$, and we try to "attach" the remaining vertices in A to the graph $(\{r, u_1, \ldots, u_k\}, \{ru_1, \ldots, ru_k\})$ via appropriate paths.

Theorem 3. $\mathrm{VC}_{k-\mathrm{con}}(G) \leq \ell(G) - k + 1$, *for any graph G and $k \geq 2$.*

Proof. Without loss of generality, we may assume G to be connected. Let A be a shattered set of maximum cardinality. Our aim is to construct a spanning tree T with at least $|A| + k - 1$ leaves. Choose any vertex $r \in A$ as the root of T. Since A is shattered, there exists a k-connected subgraph H such that r is the only element of A contained in $V(H)$. Clearly $d_H(r) \geq k$. Let $u_1, \ldots, u_k$ be arbitrary vertices in $N_H(r)$. We select the edges $u_1r, \ldots, u_kr$. For any $w \in A \setminus \{r\}$, let H_w be the k-connected subgraph such that $V(H_w) \cap A = \{r, w\}$. By Menger's theorem [13], there exist k independent $w - r$ paths in H_w, say $P_{w,1}, \ldots, P_{w,k}$. We call $w \in A \setminus \{r\}$ an *upper leaf* if there exists $i \in [k]$ such that $E(P_{w,i}) \cap \{u_1r, \ldots, u_kr\} = \emptyset$. Otherwise, i.e. if distinct $w - r$ paths in $\{P_{w,1}, \ldots, P_{w,k}\}$ contain distinct edges in $\{u_1r, \ldots, u_kr\}$, we call w a *lower leaf*.

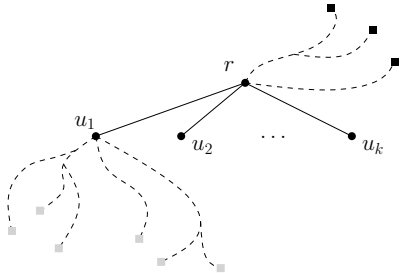


Figure 1: The black square vertices are the upper leaves, while the gray square vertices are the lower leaves. The selected paths are dashed.

We set $L := \{u_1, \ldots, u_k\}$ and we view L as the set of potential leaves for T. For any $w \in A \setminus \{r, u_1, \ldots, u_k\}$, we do the following:

- If w is an upper leaf, select the path $P_{w,i}$ of maximal length and add w to L;
- If w is a lower leaf, select the path $P_{w,1}$, add w to L and remove u_1 from L.

After removing cycles and appropriate edges from the selected subgraph, we get a tree T, rooted at r, in which the elements of L are leaves. From now on, we assume that the procedure of selecting paths is followed by the procedure just described, so that T is indeed a tree. Moreover, L has cardinality at least $|A| + k - 2$. Note that u_1 is the vertex in which we "attach" the paths coming from lower leaves and that it could be replaced by any other u_j. The construction above works for any k, the case $k = 1$ giving the upper bound in Theorem 2. Now we consider $k \geq 2$.

Claim 4. *We may assume that T has no upper leaves.*

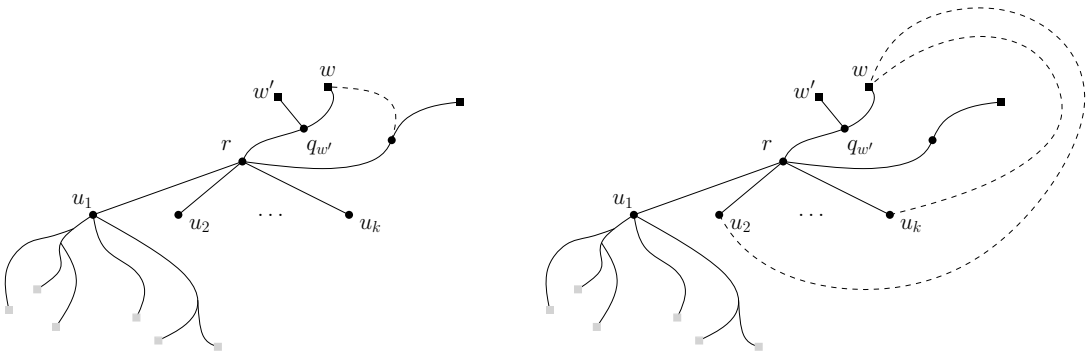
Proof of Claim 4. Let T' be a tree constructed as above and with minimal number of upper leaves. Namely, T' satisfies the following properties:

- T' is rooted at r;
- $\{u_1r, \ldots, u_kr\} \subseteq E(T')$;
- T' has at least $|A| + k - 2$ leaves;
- T' has the minimal number of upper leaves.

Let $w' \in V(T')$ be an upper leaf and let $P_{w',i'}$ be the corresponding $w' - r$ path. Suppose first that $N_{P_{w',i'}}(w') = q_{w'} \neq r$. Let W be the set of upper leaves whose corresponding selected paths contain $q_{w'}$. Note that if $q_{w'}$ is contained in some $w'' - r$ path, for a lower leaf w'', we trivially get a tree with fewer upper leaves. Therefore, we may assume this is not the case. For any $w \in W$, there exist $k - 1 \geq 1$ independent $w - r$ paths not containing $q_{w'}$. We distinguish two cases, proceeding as follows (see Figure 2):

1. If one of these paths intersects for the first time an already selected path in a vertex not in $\{u_2, \ldots, u_k\}$, then select this path;
2. Otherwise, we have a $w - \{u_2, \ldots, u_k\}$ fan. Then select the path containing u_2.

If 1. holds for each $w \in W$, then $q_{w'}$ becomes an additional leaf and we get a tree with at least $|A| + k - 1$ leaves. Otherwise, we get a tree with at least $|A| + k - 2$ leaves (we have replaced u_2 by $q_{w'}$) and with fewer upper leaves, a contradiction. Therefore, it remains to consider the case $N_{P_{w',i'}}(w') = r$. By maximality of the length of $P_{w',i'}$, we have that there exists an $w' - r$ path containing u_jr, for some $j \in [k]$. Then, after "moving" the lower leaves to u_j, we get a contradiction again. □



(a) Case 1. **(b)** Case 2.

Figure 2: Illustration of the two possible cases in the Proof of Claim 4. The subpaths of paths not containing $q_{w'}$ are dashed.

Now we want to build a tree, rooted at u_1, with at least $|A| + k - 1$ leaves. By Claim 4, it is enough to do the following: for any $1 \neq i \in [k]$, select a $u_1 - u_i$ path in H not containing $\{u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_k\} \cup \{r\}$ and add r to L. If the resulting tree T is not spanning, then find an edge between a vertex not in $V(T)$ and a vertex in $V(T)$ and add it to $E(T)$. Since this procedure does not decrease the number of leaves in T, we get a spanning tree with at least $|A| + k - 1$ leaves. □

By considering the complete graph on $k + 1$ vertices, our bound is tight in the sense of the following.

Proposition 5. *For any $k \geq 2$, there exists a graph G such that* $\mathrm{VC}_{k-\mathrm{con}}(G) = \ell(G) - k + 1$.

As for a lower bound, we note that having a sufficiently large complete subgraph is enough to guarantee shattering by k-connected subgraphs.

Theorem 6. *Let G be a graph of order n, size m, and maximum degree Δ. For $k \geq 2$,*

$$\mathrm{VC}_{k-\mathrm{con}}(G) \geq \ell(G) - k + 1 - \left(n + 2 - \left\lceil \frac{n-2}{\Delta - 1} \right\rceil - \frac{n^2}{n^2 - 2m} \right).$$

Proof. By Turán's theorem [13], if $m > \left(1 - \frac{1}{r}\right) \frac{n^2}{2}$, then G contains K_{r+1} as a subgraph. Therefore, a set of size $r + 1 - (k + 1)$ can be shattered by k-connected subgraphs. The condition above is equivalent to $r < \frac{n^2}{n^2 - 2m}$ and so, taking $r = \left\lceil \frac{n^2}{n^2-2m} - 1 \right\rceil$, we get

$$\mathrm{VC}_{k-\mathrm{con}}(G) \geq \left\lceil \frac{n^2}{n^2 - 2m} - 1 \right\rceil + 1 - (k + 1).$$

Now we want to find an upper bound for $\ell(G)$. Let T be a spanning tree of G and let $d_i = |\{v \in V(T) : d_T(v) = i\}|$ be the number of vertices of degree i in T. We have that

$$\sum_{i=1}^{\Delta} d_i = n, \quad \text{and} \quad 2(n-1) = \sum_{v \in V(T)} d_T(v) = \sum_{i=1}^{\Delta} i d_i.$$

Therefore, we consider the following integer linear programming

$$\begin{aligned} \max \quad & d_1 \\ \text{s.t.} \quad & n-2 = \sum_{i=2}^{\Delta} (i-1)d_i. \end{aligned}$$

Equivalently, we seek to find

$$\begin{aligned} \min \quad & \sum_{i=2}^{\Delta} d_i \\ \text{s.t.} \quad & n-2 = \sum_{i=2}^{\Delta} (i-1)d_i. \end{aligned}$$

It is not difficult to see that the minimum is equal to $\left\lceil \frac{n-2}{\Delta-1} \right\rceil$. Summarizing, we get

$$\begin{aligned} \mathrm{VC}_{k-\mathrm{con}}(G) &\geq \left\lceil \frac{n^2}{n^2-2m} - 1 \right\rceil - k \\ &\geq \frac{n^2}{n^2-2m} - 1 - k + \left(\ell(G) - n + \left\lceil \frac{n-2}{\Delta-1} \right\rceil \right) \\ &\geq \ell(G) - k + 1 - \left(n + 2 - \left\lceil \frac{n-2}{\Delta-1} \right\rceil - \frac{n^2}{n^2-2m} \right). \end{aligned}$$

□

3 The decision problem

In this Section we investigate the computational complexity of Graph $\mathrm{VC}_{k-\mathrm{con}}$ Dimension. Consider the following decision problem, usually called Set Multicover:

Set Multicover	
Instance:	A set $S = \{a_1, \ldots, a_n\}$, a collection of subsets $S_1, \ldots, S_m \subseteq S$, and integers k and t.
Question:	Is there an index set $I \subseteq \{1, \ldots, m\}$ such that $\bigcup_{i \in I} S_i = S$, each a_i is covered by at least k distinct subsets, and $\lvert I \rvert \leq t$?

Being a generalization of the well-known Set Cover (also known as Minimum Cover [4]), it is NP-complete. We use it to prove the following Theorem.

Theorem 7. Graph $\mathrm{VC}_{k-\mathrm{con}}$ Dimension *is* NP*-complete.*

Proof. First we show that the problem is in NP. Our proof is based on the following elementary Lemma. Since we could not find it in the literature, we give its short proof.

Lemma 8. *Let G and G' be two k-connected graphs such that $|V(G) \cap V(G')| \geq k$. Then $G \cup G'$ is k-connected as well.*

Proof. Let $S \subset (G \cup G')$ be a subset such that $|S| < k$. Let v and w be two distinct vertices in $(G \cup G') - S$. If both v and w are in G or in G', then there is a $v - w$ path in $(G \cup G') - S$ by assumption. Otherwise, since $|V(G) \cap V(G')| \geq k$, there exists $u \in V(G) \cap V(G') \cap V((G \cup G') - S)$. Moreover, since $G - S$ and $G' - S$ are connected, there exist a $v - u$ path in $G - S$ and a $u - w$ path in $G' - S$. But then there is a $v - w$ walk in $(G \cup G') - S$ and so a $v - w$ path as well. □

Let $G = (V, E)$ and $s \geq 1$ be an instance of Graph $\text{VC}_{k-\text{con}}$ Dimension. For a subset $V' \subseteq V$ with $|V'| \geq s$, by Lemma 8, we can check in polynomial time whether V' is shattered. Indeed, it is enough to check all the $O(|V|^{k+1})$ subsets of V' of size at most $k+1$.

Now we prove NP-hardness by a reduction from Set Multicover. Given an instance of Set Multicover, our reduction constructs a graph $G = (V, E)$ as follows (see Figure 3). The set of vertices V is formed by four pairwise disjoint sets A, B, C and D. The set A has $n \cdot (t + k + 1)$ vertices, arranged in n columns of $t + k + 1$ vertices each (for $1 \leq j \leq n$, each element in the j-th column corresponds to a copy of a_j). $B = \{v_1, \ldots, v_m\}$, where v_i corresponds to the set S_i, C consists of k vertices and D of $t + m + 1$ vertices. The vertices in C are connected to each other in order to form the complete graph K_k and each vertex in C is connected to all vertices in B and D. Finally, $v_i \in B$ is connected to every copy of $a_j \in A$ if and only if $a_j \in S_i$.

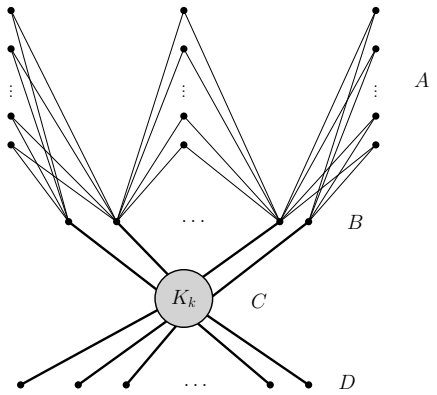


Figure 3: The graph G for the reduction. A thick edge joining a vertex v to K_k means that v is adjacent to all the vertices of K_k.

We claim that there is an index set $I \subseteq \{1, \ldots, m\}$ such that $\bigcup_{i \in I} S_i = S$, each a_i is covered by at least k distinct subsets and $|I| \leq t$ if and only if $\text{VC}_{k-\text{con}}(G) \geq |V| - (t + k)$.

First suppose that such an index set I exists. We claim that the set

$$V' = A \cup D \cup \{v_i \in B : i \notin I\}$$

is shattered. Indeed, any $W \subseteq V'$ is of the form $W = W' \cup W''$, where $W' \subseteq A$ and $W'' \subseteq D \cup \{v_i \in B : i \notin I\}$. But then W can be written as $W = V(G[C \cup B \cup D \cup W']) \cap V'$ and $G[C \cup B \cup D \cup W']$ is k-connected by the well-known Expansion Lemma [13]. Finally, since $|I| \leq t$, we have $|V'| \geq |V| - t - k$.

Conversely, let V' be a shattered set of maximum cardinality. Then $|V'| \geq |V| - (t+k)$. Suppose there exists $c \in C \cap V'$. Then no vertex in D can be shattered, and so $|V'| \leq |V| - (t+m+1) < |V| - (t+k)$. Therefore, no vertex in C is in V' and $D \subseteq V'$. By the Expansion Lemma and since all the vertices in any fixed column in A have identical neighbourhoods, then either all or no vertices in any column are in V'. Since any column contains $t+k+1$ vertices, we have that $A \subseteq V'$. Therefore, the number of vertices in B which are in V' is at least $|V| - (t+k) - |A| - |D| = m - t$. We claim that $I = \{i : v_i \in B \setminus V'\}$ is a yes-instance of the Set Multicover problem. Indeed, since V' is shattered, any vertex in A has at least k neighbours in $B \setminus V'$. In other words each $a_i \in S$ is contained in at least k of the subsets $S_1, \ldots, S_m$. Moreover, $|I| \leq m - (m-t) = t$. □

3.1 Planar Monotone 1-In-3-Sat

A natural way to prove NP-hardness of a planar problem is to reduce from another planar problem, such as Planar 3-Sat. Our hardness proof for Graph VC_{2-con} Dimension is indeed based on a variant of Planar Monotone 3-Sat, which was shown to be NP-complete by de Berg and Khosravi [2]. Since their problem has already found many and diverse applications in NP-hardness proofs [3, 7, 5, 11], we think that our variation may be useful as well.

Let $\mathcal{U} = \{x_1, \ldots, x_n\}$ be a set of n boolean variables and let $\mathcal{C} = C_1 \wedge \cdots \wedge C_m$ be a 3-CNF formula defined over $\mathcal{U}$, where each clause C_i is the disjunction of exactly three literals. 1-In-3-Sat is the problem of deciding whether $\mathcal{C}$ is satisfiable in 1-in-3 sense, namely if there exists a truth assignment to the variables such that exactly one literal in each clause is true. Given an instance $\mathcal{C}$ of 1-In-3-Sat, the *associated graph* $G(\mathcal{C})$ is the bipartite graph having one vertex for each variable v_i and one vertex for each clause C_j, and an edge $\{v_i, C_j\}$ if and only if v_i or $\overline{v_i}$ appears in C_j. An instance $\mathcal{C}$ of 1-In-3-Sat is *planar* if $G(\mathcal{C})$ is planar. A planar instance of 1-In-3-Sat has a *rectilinear representation* if $G(\mathcal{C})$ can be drawn as follows: the vertices are drawn as rectangles, with all the rectangles representing the variables on a horizontal line, and the edges are vertical segments (see Figure 4 for an example).

Mulzer and Rote [9] showed that the following problem is NP-complete.

Planar 1-In-3-Sat	
Instance:	A planar 3-CNF formula $\mathcal{C}$ defined over $\mathcal{U}$, together with a rectilinear representation.
Question:	Is $\mathcal{C}$ satisfiable (in 1-in-3 sense)?

A clause with only positive literals is a *positive clause*, a clause with only negative literals is a *negative clause* and a clause with both positive and negative literals is a *mixed*

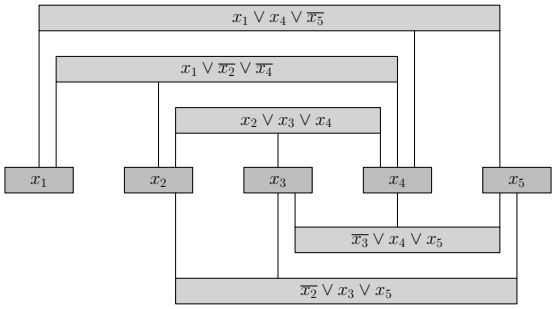


Figure 4: A rectilinear representation of a planar instance of 1-In-3-Sat.

clause. An instance of 1-In-3-Sat is called *monotone* if it does not contain any mixed clause. Given a planar monotone instance $\mathcal{C}$ of 1-In-3-Sat, a *monotone rectilinear representation* of $\mathcal{C}$ is a rectilinear representation where all the positive clauses are drawn above the variables and all the negative clauses are drawn below the variables. Therefore, we can define the following variant of Planar 1-In-3-Sat:

Planar Monotone 1-In-3-Sat	
Instance:	A planar monotone 3-CNF formula $\mathcal{C}$ defined over $\mathcal{U}$, together with a monotone rectilinear representation.
Question:	Is $\mathcal{C}$ satisfiable (in 1-in-3 sense)?

As pointed out before, the 3-Sat variant of the problem above was shown to be NP-complete by de Berg and Khosravi [2]. With a slight modification of their idea we can show the following Theorem. For the sake of completeness we include a detailed proof.

Theorem 9. Planar Monotone 1-In-3-Sat *is* NP*-complete.*

Proof. Obviously, Planar Monotone 1-In-3-Sat is in NP. We prove NP-hardness by a reduction from Planar 1-In-3-Sat. Let $\mathcal{C} = C_1 \wedge \cdots \wedge C_m$ be an instance of Planar 1-In-3-Sat defined over the variable set $\mathcal{U} = \{x_1, \ldots, x_n\}$, together with a rectilinear representation. A literal-clause pair $(\overline{x_i}, C_j)$ is *inconsistent* if $\overline{x_i}$ appears in C_j and C_j is above the variables in the rectilinear representation. Similarly, (x_i, C_j) is *inconsistent* if x_i appears in C_j and C_j is below the variables in the rectilinear representation. Note that a rectilinear representation without inconsistent literal-clause pairs is monotone. The idea is to convert the given instance into an equivalent instance with a monotone rectilinear representation by decreasing successively the number of inconsistent literal-clause pairs. Given two variables x and y, the *inequality gadget* $x \neq y$ is defined as the formula $(x \vee a \vee y) \wedge (a \vee b \vee c) \wedge (\overline{b} \vee \overline{c} \vee \overline{d})$, where a, b, c and d are new variables used only inside the gadget. Note that it has a monotone rectilinear representation, as depicted in Figure 5.

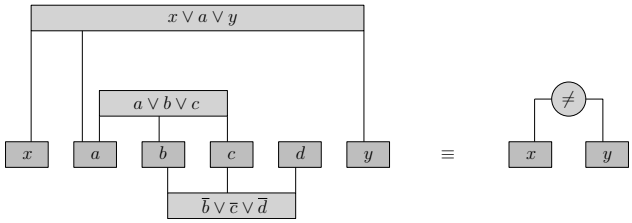


Figure 5: Inequality gadget $x \neq y$, together with a monotone rectilinear representation.

Lemma 10. *$x \neq y$ is satisfiable (in* 1*-in-*3 *sense) if and only if exactly one between x and y is* TRUE*.*

Proof. If $x \neq y \equiv (x \vee a \vee y) \wedge (a \vee b \vee c) \wedge (\overline{b} \vee \overline{c} \vee \overline{d})$ is satisfiable, then $a =$ FALSE (otherwise there would be at least two TRUE literals in the last clause) and so exactly one between x and y is TRUE. Conversely, setting $a =$ FALSE, $b =$ TRUE, $c =$ FALSE and $d =$ TRUE gives an assignment that satisfies $x \neq y$. □

We now show how to reduce the number of inconsistent pairs of the form $(\overline{x_i}, C_j)$. The remaining case can be treated similarly. The process is as follows:

- Introduce two new variables x and y.
- In clause C_j, replace $\overline{x_i}$ by x.
- Introduce the inequality gadgets $x_i \neq x$ and $x \neq y$.
- In each clause containing x_i, drawn above the variables and connecting to x_i to the right of C_j, replace x_i by y (see Figure 6).

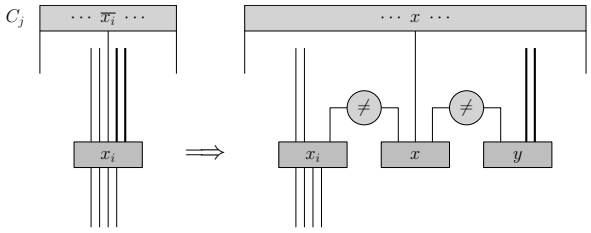


Figure 6: Removing an inconsistent pair $(\overline{x_i}, C_j)$.

Lemma 11. *Let $\mathcal{C}'$ be the formula obtained after the process above. Then $\mathcal{C}$ is satisfiable (in* 1*-in-*3 *sense) if and only if $\mathcal{C}'$ is satisfiable (in* 1*-in-*3 *sense).*

Proof. Suppose there is a truth assignment to $\mathcal{U}$ that satisfies $\mathcal{C}$. $\mathcal{C}'$ is defined over $\mathcal{U} \cup \mathcal{F}$, where $\mathcal{F}$ is the set of new variables appearing in the gadgets $x_i \neq x$ and $x \neq y$. We use the same truth assignment for $\mathcal{U}$ and we set $x := \overline{x_i}$ and $y := x_i$. By Lemma 10 there exists

a truth assignment that satisfies $\mathcal{C}'$. Conversely, suppose there is a truth assignment to $\mathcal{U} \cup \mathcal{F}$ that satisfies $\mathcal{C}'$. We claim that such an assignment restricted to $\mathcal{U}$ satisfies $\mathcal{C}$. Indeed, by Lemma 10, we have that $x = \overline{x_i}$ and $x_i = y$. But then all the modified clauses in $\mathcal{C}$ are satisfied. □

By eventually growing or shrinking some of the rectangles in $\mathcal{C}$ to make room for the gadgets $x_i \neq x$ and $x \neq y$, we can find a monotone rectilinear representation for the new instance $\mathcal{C}'$. Finally, since there are at most $3m$ inconsistent literal-clause pairs, it is clear that the reduction is polynomial time. □

3.2 The decision problem for planar graphs

Finally, we show that Graph $\mathrm{VC}_{2-\mathrm{con}}$ Dimension remains NP-complete for planar graphs by a reduction from Planar Monotone 1-In-3-Sat. For an instance $\mathcal{C}$ of Planar Monotone 1-In-3-Sat, our reduction constructs a planar graph $G_{\mathcal{C}}$ by adding several planar gadgets to the graph associated to $\mathcal{C}$.

Theorem 12. Graph $\mathrm{VC}_{2-\mathrm{con}}$ Dimension *is* NP*-complete for planar graphs.*

Proof. Clearly, membership in NP follows from Theorem 7. To prove NP-hardness we use a reduction from Planar Monotone 1-In-3-Sat. Let $\mathcal{C}$ be an instance of Planar Monotone 1-In-3-Sat. Suppose that $\mathcal{C}$ is the conjunction of m clauses, say $C_1, \ldots, C_m$, defined over the variable set $\mathcal{U} = \{x_1, \ldots, x_n\}$. We modify the associated graph $G(\mathcal{C})$ to get a planar graph $G_{\mathcal{C}}$ as follows. First we define a *shamrock* as the graph constructed successively from a triangle C, called the *centre* of the shamrock, by adding q ears of length two to every pair of vertices of C. These q ears constitute a *leaf* of the shamrock. For a fixed planar drawing of a shamrock, the exterior vertex in a leaf is called the *peak* of the leaf, while the other vertices are called the *veins* of the leaf. For every clause, we introduce a shamrock with p veins in every leaf (for a p to be chosen later) as depicted in Figure 7(a). Each peak of a leaf in the shamrock corresponds to a literal in the clause. For every variable x, we introduce a variable gadget as depicted in Figure 7(b). It consists of two parts. The first part is a 4-cycle with two specified opposite vertices corresponding to the literals x and $\overline{x}$, called the *literal* vertices, and with two *horizontal* vertices. The second part is a shamrock with p veins in every leaf and no peaks which is connected to the 4-cycle via three edges with endpoints in the centre of the shamrock. Note that two of these edges join x and $\overline{x}$ to the centre of the shamrock. Clearly, a monotone rectilinear representation of $\mathcal{C}$ defines a total ordering of the variables according to their position on the horizontal line. We connect every pair of consecutive variables x and y (y is the successor of x in the total ordering) through a variable connector gadget as depicted in Figure 7(c). It consists of $p + 1$ independent paths: one of them is of length one while the remainings are all of length two. Finally, we connect clause and variable gadgets via edges joining every peak in a shamrock with the corresponding literal in the variable gadget. We do this according to the monotone rectilinear representation of $G(\mathcal{C})$ and we

define $p := 5m + 6n + 1$. Clearly, the above construction can be done in polynomial time and the resulting graph $G_{\mathcal{C}} = (V, E)$ is planar (see Figure 8 for an example). We claim that $\mathrm{VC}_{2-\mathrm{con}}(G_{\mathcal{C}}) \geq |V| - (5m + 6n)$ if and only if $\mathcal{C}$ is satisfiable (in 1-in-3 sense).

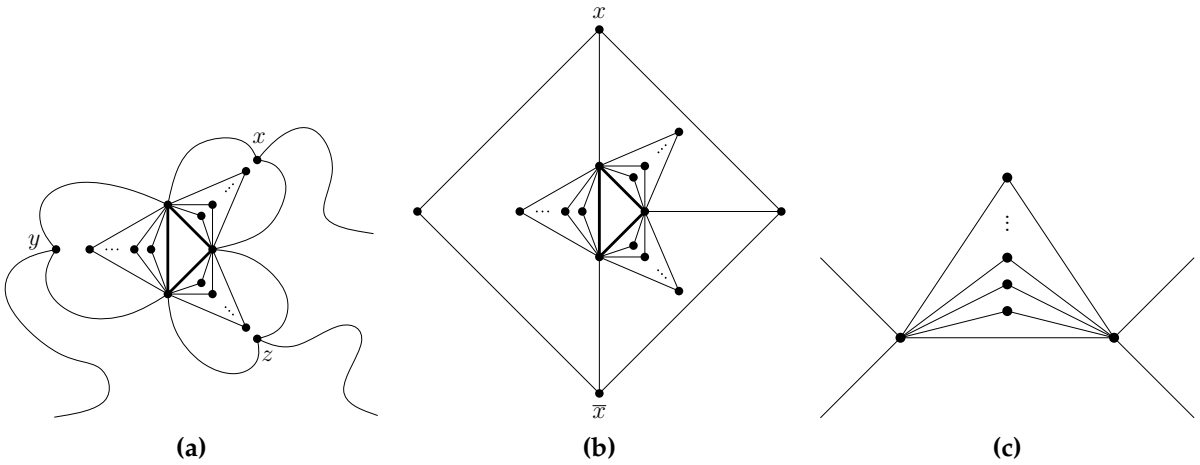


Figure 7: Gadgets for the reduction of Planar Monotone 1-In-3-Sat to Graph $\mathrm{VC}_{2-\mathrm{con}}$ Dimension. (a) Clause gadget corresponding to the clause $(x \vee y \vee z)$. (b) Variable gadget. (c) Variable connector gadget.

First suppose that φ is a satisfying truth assignment for $\mathcal{C}$. We say that a literal vertex in a variable gadget or a peak in a clause gadget is a *true* (*false*) vertex if the corresponding literal is assigned to TRUE (FALSE) by φ. Let $A \subseteq V$ be defined as follows. It contains the vertex set of the centre of any shamrock in $G_{\mathcal{C}}$. Moreover, for any variable gadget, it contains the horizontal vertices as well as the false vertex and, for any clause gadget, it contains the false vertices. Note that $|A| = 5m + 6n$. Therefore, it is enough to show that $V \setminus A$ is shattered. Consider the induced subgraph $G_{\mathcal{C}}[A]$. It consists of a $3n$-cycle through the false vertices in the variable gadgets, together with the centres of any variable and clause gadget in $G_{\mathcal{C}}$ joined to the cycle via exactly two edges (since for any variable x exactly one between x and $\overline{x}$ is assigned to FALSE and since in any clause exactly two literals are assigned to FALSE). Clearly, $G_{\mathcal{C}}[A]$ is 2-connected. Note that any vertex in $V \setminus A$ belonging to a shamrock can be joined to $G_{\mathcal{C}}[A]$, independently on any other vertex of $V \setminus A$, via its two neighbours on the centre. Moreover, any true literal in a variable gadget can be joined to $G_{\mathcal{C}}[A]$, independently on any other vertex of $V \setminus A$, via the two horizontal vertices adjacent to it. In both cases, by the Expansion Lemma [13], the resulting subgraphs are 2-connected. Therefore, $V \setminus A$ can be shattered.

Conversely, let V' be a shattered set of maximum cardinality. Then $|V'| \geq |V| - (5m + 6n)$. Consider a shamrock in $G_{\mathcal{C}}$ and one of its three leaves. Then, by maximality, either all or no veins of the leaf are in V'. Since there are $5m + 6n + 1$ veins, we have that every vein in every shamrock in $G_{\mathcal{C}}$ is in V'. Therefore, no vertex w belonging to the centre of a shamrock is in V', otherwise it would not be possible to shatter the singleton set consisting of a vein adjacent to w. Similarly, all the vertices in the variable connector gadgets are in V' and no horizontal vertex is in V'. Note that at most one literal vertex in

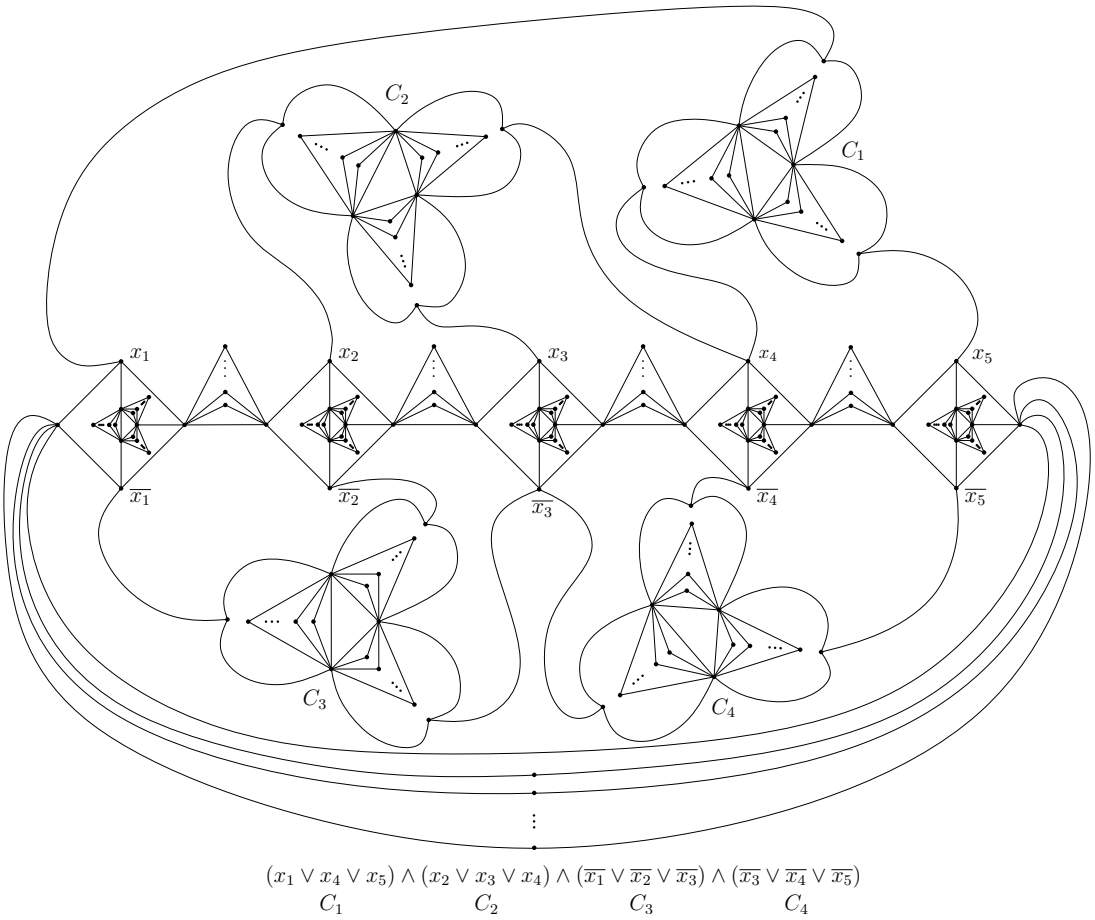


Figure 8: Example reduction of Planar Monotone 1-In-3-Sat to Graph $\mathrm{VC}_{2-\mathrm{con}}$ Dimension.

a variable gadget is in V', otherwise it would not be possible to shatter the set consisting of a vein in a variable gadget and a vertex in a variable connector gadget. Moreover, at most one peak in a clause gadget is in V', otherwise it would not be possible to shatter the set consisting of a vein in the clause gadget and a vertex not in the clause gadget. But then, since $|V'| \geq |V| - (5m + 6n)$, exactly one literal vertex in every variable gadget and exactly one peak in every clause gadget are in V'. Without loss of generality we may assume that no literal vertex which does not appear in any clause (and therefore having degree 3 in $G_{\mathcal{C}}$) is in V'. But then, if a literal vertex is in V', the corresponding peaks are in V' as well, otherwise it would not be possible to shatter vertices in a clause and vertices not in that clause. Therefore, based on the shattered set of maximum size V', we set a literal to true if the corresponding literal vertex is in V' and to false otherwise. From the considerations above it is clear that the resulting assignment is well-defined and that it satisfies $\mathcal{C}$ (in 1-in-3 sense). □

4 Conclusion and further work

This paper represents a little step in the systematic study of the VC-dimensions of graphs initiated by Kranakis et al. [8]. We have concentrated on the VC-dimension with respect to k-connected subgraphs. First we have given upper and lower bounds. In this context it would be interesting to improve the lower bound in Theorem 6. Then we have proved NP-completeness results for the associated decision problem. It would be interesting to check whether the problem for planar graphs becomes easier for large values of k. We intend to further study the complexity of the problem when the input graph belongs to other classes than the one of planar graphs. Moreover, we intend to focus on the VC-dimension with respect to other classes of subgraphs.

References

[1] M. Anthony, G. Brightwell, and C. Cooper. The Vapnik-Chervonenkis Dimension of a Random Graph. *Discrete Mathematics*, 138(1):43–56, 1995.

[2] M. de Berg and A. Khosravi. Optimal Binary Space Partitions for Segments in the Plane. *Int. J. Comput. Geometry Appl.*, 22(3):187–206, 2012.

[3] E. D. Demaine and S. Eisenstat. Flattening Fixed-Angle Chains Is Strongly NP-Hard. In *Algorithms and Data Structures - 12th International Symposium, WADS 2011, New York, NY, USA, August 15-17, 2011. Proceedings*, pages 314–325. Springer, 2011.

[4] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[5] T. Hartmann, J. Rollin, and I. Rutter. Cubic Augmentation of Planar Graphs. In *Algorithms and Computation - 23rd International Symposium, ISAAC 2012, Taipei, Taiwan, December 19-21, 2012. Proceedings*, volume 7676 of *Lecture Notes in Computer Science*, pages 402–412. Springer, 2012.

[6] D. Haussler and E. Welzl. ε-Nets and Simplex Range Queries. *Discrete & Computational Geometry*, 2:127–151, 1987.

[7] B. Katz, I. Rutter, and G. J. Woeginger. An algorithmic study of switch graphs. *Acta Inf.*, 49(5):295–312, 2012.

[8] E. Kranakis, D. Krizanc, B. Ruf, J. Urrutia, and G. J. Woeginger. The VC-dimension of set systems defined by graphs. *Discrete Applied Mathematics*, 77(3):237–257, 1997.

[9] W. Mulzer and G. Rote. Minimum weight triangulation is NP-hard. In *Proceedings of the 22nd ACM Symposium on Computational Geometry, Sedona, Arizona, USA, June 5-7, 2006*, pages 1–10. ACM, 2006.

[10] M. Schaefer. Deciding the VC-dimension is Σ_3^p-complete, II. Technical Report TR00-006, DePaul University, 2000.

[11] L. Schlipf. Notes on Convex Transversals. *CoRR*, abs/1211.5107, 2012.

[12] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16(2):264–280, 1971.

[13] D. B. West. *Introduction to Graph Theory*. Prentice Hall, 2 edition, 2001.