



**QUEEN'S  
UNIVERSITY  
BELFAST**

## The VC-dimension of graphs with respect to k-connected subgraphs

Munaro, A. (2016). The VC-dimension of graphs with respect to k-connected subgraphs. *Discrete Applied Mathematics*, 211, 163-174. <https://doi.org/10.1016/j.dam.2016.04.016>

**Published in:**  
Discrete Applied Mathematics

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

### **Publisher rights**

Copyright 2016 Elsevier.

This manuscript is distributed under a Creative Commons Attribution-NonCommercial-NoDerivs License

(<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

### **General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

### **Open Access**

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

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

Andrea Munaro\*

*Laboratoire G-SCOP, Université Grenoble Alpes*

---

## 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 tight 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 split graphs and for some subclasses of planar bipartite graphs in the cases  $k = 1$  and  $k = 2$ . On the positive side, we observe it can be decided in linear time for graphs of bounded clique-width.

*Keywords:* VC-dimension,  $k$ -connected, NP-complete

---

## 1. Introduction

Let  $\mathcal{H}$  be a set system on a finite set  $X$ . A subset  $Y \subseteq X$  is *shattered by  $\mathcal{H}$*  if  $\{E \cap Y : E \in \mathcal{H}\} = 2^Y$  and the *VC-dimension* of  $\mathcal{H}$  is defined as the maximum size of a set shattered by  $\mathcal{H}$ . The VC-dimension of a set system was introduced by Vapnik and Chervonenkis [1]. 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 a set system. One might think applying 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 [2]. They considered the set system induced by closed neighbourhoods of the vertices. Kranakis et al. [3] 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  $V(H)$  with  $A$ , for  $H \in \mathcal{P}$ . The *VC-dimension* of  $G$  with respect to  $\mathcal{P}$ , denoted by  $VC_{\mathcal{P}}(G)$ , is defined as the maximum size of a  $\mathcal{P}$ -shattered subset.

Note that in this paper we consider only finite undirected simple graphs and we use standard graph theoretic terminology from [4], unless stated otherwise. For definitions and examples related to tree-width and clique-width we refer the reader to [5, 6], while for an introduction to monadic second-order logic of graphs see, for example, [7].

According to Definition 1, we denote by  $VC_{\text{tree}}$ ,  $VC_{\text{con}}$ ,  $VC_{k\text{-con}}$ ,  $VC_{\text{nbd}}$ ,  $VC_{\text{path}}$ ,  $VC_{\text{cycle}}$  and  $VC_{\text{star}}$  the VC-dimensions with respect to families of subgraphs which are trees, connected,  $k$ -connected, induced by the closed neighbourhoods of the vertices, paths, cycles and stars, 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

---

\*This work was partly done at Dept. of Computer Science, University of Bonn.  
Email address: [Andrea.Munaro@grenoble-inp.fr](mailto:Andrea.Munaro@grenoble-inp.fr) (Andrea Munaro)

family of complete subgraphs then  $\text{VC}_{\mathcal{P}}$  is the clique number, while if  $\mathcal{P}$  is the family of subgraphs induced by independent sets then  $\text{VC}_{\mathcal{P}}$  is the independence number.

Since a graph of order  $n$  has  $n$  closed neighbourhoods, its VC-dimension is at most  $\lfloor \log_2 n \rfloor$  [2]. It is not difficult to show that this bound is tight [8]. 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  $\text{VC}_{\text{nbd}}(H) = \lfloor \log_2 n \rfloor$ . If  $G$  is a graph with maximum degree  $\Delta$ , then it is easy to see that  $\Delta \leq \text{VC}_{\text{star}} \leq \Delta + 1$  [3]. The VC-dimension with respect to trees is the same as the VC-dimension with respect to connected subgraphs [3]. This is an immediate consequence of the fact that a connected graph contains a spanning tree.

Kranakis et al. [3] 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. [3]).**  $\ell(G) \leq \text{VC}_{\text{con}}(G) \leq \ell(G) + 1$ , for any graph  $G$ .

Recall that  $\ell(G) = |V(G)| - \gamma_c(G)$ , where  $\gamma_c(G)$  is the connected domination number, the minimum size of a connected dominating set in  $G$ . Moreover, given a graph  $G$  and an integer  $k$ , the problem of deciding whether  $\gamma_c(G) \leq k$  is NP-complete [9]. A natural question is to investigate the computational complexity of computing  $\text{VC}_{\mathcal{P}}(G)$  for a given graph  $G$  and a family of its subgraphs  $\mathcal{P}$ . The decision problem is formulated as follows:

GRAPH  $\text{VC}_{\mathcal{P}}$  DIMENSION

**Instance:** A graph  $G$  and a number  $s \geq 1$ .

**Question:** Does  $\text{VC}_{\mathcal{P}}(G) \geq s$  hold?

### 1.1. Our results

In Section 2 we extend Theorem 2 by giving tight upper and lower bounds on the VC-dimension with respect to  $k$ -connected subgraphs, for  $k \geq 2$ . These are given, similarly to Theorem 2, in terms of the number of leaves in a maximum leaf spanning tree. In Section 3 we prove that the related decision problem GRAPH  $\text{VC}_{k-\text{con}}$  DIMENSION is NP-complete even for split graphs. On the positive side, we show it can be decided in linear time for graphs of bounded clique-width and in polynomial time for the subclass of split graphs having Dilworth number at most 2. Finally, we prove that GRAPH  $\text{VC}_{\text{con}}$  DIMENSION and GRAPH  $\text{VC}_{2-\text{con}}$  DIMENSION remain NP-complete for some subclasses of planar bipartite graphs with maximum degree at most 4. The following table summarizes the known results about the complexity of GRAPH  $\text{VC}_{\mathcal{P}}$  DIMENSION (see [10] and [11] for the definitions of the classes LOGNP and  $\Sigma_3^P$ , respectively).

Family $\mathcal{P}$	Graph $G$	Computational Complexity	Reference
star		polynomial time	Kranakis et al. [3]
neighbourhood		LOGNP-complete	Kranakis et al. [3]
path		$\Sigma_3^P$ -complete	Schaefer [12]
cycle		$\Sigma_3^P$ -complete	Schaefer [12]
$k$ -connected	split	NP-complete	Theorem 17
$k$ -connected	bounded clique-width	linear time	Corollary 20
$k$ -connected	split, Dilworth number $\leq 2$	polynomial time	Theorem 21
connected	planar, bipartite, $\Delta(G) = 3$	NP-complete	Theorem 25
2-connected	planar, bipartite, $\Delta(G) = 4$	NP-complete	Theorem 29

## 2. Bounds on the VC-dimension

The main result of this section is an extension of Theorem 2 by considering families of  $k$ -connected subgraphs, for  $k \geq 2$ .

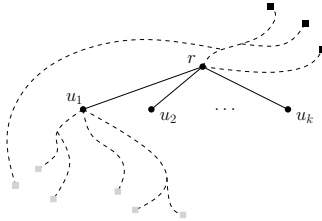
**Theorem 3.**  $\text{VC}_{k\text{-con}}(G) \leq \ell(G) - k + 1$ , for any connected graph  $G$  and  $k \geq 2$ .

The proof of Theorem 3 consists essentially of two parts. First, we show that  $\text{VC}_{k\text{-con}}(G) \leq \ell(G) - k + 2$ , for any connected graph  $G$  and  $k \geq 1$  (the case  $k = 1$  thus gives a proof for the upper bound of Theorem 2). The idea is to construct a spanning tree  $T$  with at least  $\text{VC}_{k\text{-con}}(G) + k - 2$  leaves. We fix a shattered set  $A$  of maximum cardinality and choose an appropriate vertex  $r \in A$  as the root. Then we consider some  $k$  neighbours of  $r$ , say  $u_1, \dots, u_k$ , and we try to “attach” the remaining vertices of  $A$  to the graph  $(\{r, u_1, \dots, u_k\}, \{ru_1, \dots, ru_k\})$  via appropriate paths. A crucial step is the distinction between two types of vertices of  $A$  that we are going to add as leaves of  $T$ : lower leaves and upper leaves (see below for definitions). In the second part, we provide the actual proof of Theorem 3 by contradiction. In particular, we suppose that  $\ell(G) \leq \text{VC}_{k\text{-con}}(G) + k - 2$  and we show how to modify the tree  $T$  constructed in the first part in order to obtain a contradiction.

*Proof of Theorem 3.* We begin by showing that  $\text{VC}_{k\text{-con}}(G) \leq \ell(G) - k + 2$ , for any connected graph  $G$  and  $k \geq 1$ . Let  $A$  be a shattered set of maximum cardinality. Since our aim is to construct a spanning tree with at least  $|A| + k - 2$  leaves, note that it is enough to construct a tree  $T \subseteq G$  with as many leaves. For  $w \in A$ , we denote by  $G_w$  a fixed  $k$ -connected subgraph such that  $V(G_w) \cap A = \{w\}$ . Similarly,  $G_{ww'}$  denotes a fixed  $k$ -connected subgraph such that  $V(G_{ww'}) \cap A = \{w, w'\}$ . We choose a vertex  $r \in A$  having the minimum number of neighbours in  $V(G) \setminus A$  as the root of  $T$ . Clearly,  $d_{G_r}(r) \geq k$ . Let  $U = \{u_1, \dots, u_k\}$  be a set of  $k$  arbitrary vertices in  $N_{G_r}(r)$ . We select the edges  $u_1r, \dots, u_kr$ . By Menger’s Theorem [4], there exist  $k$  independent  $w - r$  paths in  $G_{wr}$ . We call  $w \in A \setminus \{r\}$  a *lower leaf* for  $T$  if there exist  $k$  independent  $w - r$  paths  $\{P_1, \dots, P_k\} \subseteq G$  with no inner vertex in  $A$  and such that each of them contains (exactly) one edge in  $\{u_1r, \dots, u_kr\}$  (in other words,  $w$  is a lower leaf if there exists a  $w, U$ -fan in  $G - (A \setminus \{w\})$ ). In particular, no path contains two vertices in  $U$ . Otherwise, we call  $w$  an *upper leaf* for  $T$ .

We set  $L := \{u_1, \dots, u_k\}$  and we view  $L$  as the set of potential leaves for  $T$ . For any  $w \in A \setminus \{r\}$ , we do the following (see Figure 1):

- If  $w$  is an upper leaf, select a  $w - r$  path  $P \subseteq G_{wr}$  such that  $V(P) \cap U = \emptyset$ . Such a path exists: by Menger’s Theorem [4], there exist  $k$  independent  $w - r$  paths in  $G_{wr}$  and, if each of them contained a (different) vertex in  $U$ , we would obtain a  $w, U$ -fan in  $G - (A \setminus \{w\})$ . Finally, add  $w$  to  $L$  and remove cycles and appropriate edges so that the selected subgraph is a tree;
- If  $w$  is a lower leaf, select a  $w - r$  path  $P \subseteq G$  as in the definition of lower leaf and such that  $E(P) \cap \{u_1r, \dots, u_kr\} = \{u_1r\}$ , add  $w$  to  $L$  and remove  $u_1$  from  $L$ . Finally, remove edges from the newly added paths so that the selected subgraph is a tree.



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

In this way, we get a tree  $T$ , rooted at  $r$  and in which the elements of  $L$  are leaves. Moreover,  $|L| \geq |A| + k - 2$ . The construction works for any  $k$ , the case  $k = 1$  giving the upper bound in Theorem 2.

From now on, we assume  $k \geq 2$  and we prove Theorem 3 by contradiction. Therefore, let  $G$  be a counterexample and let  $A$  be a shattered set of maximum cardinality. Using the procedure described in the previous part, we can obtain a tree  $T \subseteq G$ , rooted at  $r$  and with at least  $|A| + k - 2$  leaves. Recall that  $r$  is chosen as a vertex of  $A$  having the minimum number of neighbours in  $V(G) \setminus A$ . In the following, we

deduce some structural properties of  $G$  and show how they lead to a contradiction. Each claim is followed by a short proof.

**Claim 4.** *Each leaf in  $T$  is adjacent to at most one vertex not in  $T$ .*

Indeed, if there exists a leaf in  $T$  adjacent to at least two vertices not in  $T$ , we immediately get a tree with at least  $|A| + k - 1$  leaves.  $\square$

**Claim 5.**  *$T$  contains at least one upper leaf.*

Indeed, suppose  $T$  has no upper leaves. For any  $2 \leq i \leq k$ , select a  $u_i - u_1$  path in  $G_r$  with no inner vertex in  $(U \setminus \{u_1, u_i\}) \cup \{r\}$ . Clearly, such a path has no inner vertex in  $A \cup U$ . But then we can obtain a tree, rooted at  $u_1$ , with at least  $|A| + k - 1$  leaves ( $r$  becomes an additional leaf).  $\square$

**Claim 6.** *There exists an upper leaf  $w$  with  $d_T(w, r) \geq 2$ .*

Indeed, suppose this is not the case. By Claim 5, the set  $W$  of upper leaves for  $T$  is non-empty and each of them is adjacent to  $r$ . Suppose that any  $w \in W$  has one neighbour (in  $G$ ) which is contained in  $T_{u_1} - A$ , where  $T_{u_1}$  denotes the subtree induced by  $u_1$  and its descendants. Then we select the edges joining each upper leaf to  $T_{u_1} - A$  and, for any  $2 \leq i \leq k$ , we select a  $u_i - u_1$  path in  $G_r$  with no inner vertex in  $A \cup U$ . In this way, we obtain a tree rooted at  $u_1$  and with at least  $|A| + k - 1$  leaves, a contradiction. Therefore, there exists  $w \in W$  such that  $N_G(w) \cap (V(T_{u_1}) \setminus A) = \emptyset$ . By Claim 4,  $w$  has at most one neighbour in  $V(G) \setminus V(T)$ . Therefore,  $N_{G_w}(w) \subseteq N_G(w) \setminus A \subseteq \{u_2, \dots, u_k\} \cup \{x\}$ , for some  $x \in V(G) \setminus V(T)$ . But since  $w$  has at least  $k$  neighbours in  $G_w$ , we have  $N_{G_w}(w) = N_G(w) \setminus A = \{u_2, \dots, u_k\} \cup \{x\}$ .

Consider now  $w' \in A$  such that  $ww' \notin E(G)$ . There exists a  $w - w'$  path  $P$  in  $G_{ww'}$  with no inner vertex in  $U \setminus \{u_1\}$ . Clearly,  $wx \in E(P)$ . Note that  $P$  does not contain any vertex in  $T_{u_1} - A$ , or else  $x$  could become an additional leaf for  $T$ . By selecting these paths, together with edges connecting vertices in  $A$  to  $w$ , it is easy to see we can get a new tree  $T'$  rooted at  $w$  and with at least  $|A| + k - 2$  leaves. Moreover, if an upper leaf for  $T$  has at least two neighbours in the subtree  $T_{u_1} - A$ , then we can get an additional leaf for  $T'$ . Therefore, each upper leaf for  $T$  has at most one neighbour in  $T_{u_1} - A$ .

Now let  $w' \neq w$  be an upper leaf for  $T$ . We claim that  $w'$  is adjacent to  $u_2$ . Indeed, suppose  $w'u_2 \notin E(G)$ . By Claim 4,  $w'$  has at most one neighbour in  $G - V(T)$  and by the paragraph above it has at most one neighbour in  $T_{u_1} - A$ . Since  $d_{G_{w'}}(w') \geq k$ , then  $N_{G_{w'}}(w') = (U \setminus \{u_1, u_2\}) \cup \{y, z\}$ , for some  $y \in V(G) \setminus V(T)$  and  $z \in V(T_{u_1}) \setminus A$ . Moreover,  $u_2 \notin V(G_{w'})$ . Otherwise, there exists a  $w' - u_2$  path  $P$  in  $G_{w'}$  with no inner vertex in  $(U \setminus \{u_1, u_2\}) \cup \{z\}$ . Clearly,  $w'y \in E(P)$  and  $P$  does not contain vertices in  $T_{u_1} - A$ , or else  $y$  would become an additional leaf for  $T$ . But then there exists a  $w', U$ -fan in  $G - (A \setminus \{w'\})$ , contradicting the fact that  $w'$  is an upper leaf. Therefore, since  $u_2 \notin V(G_{w'})$ , there exists a  $y - z$  path in  $G_{w'}$  with no inner vertex in  $A \cup (U \setminus \{u_1\})$ . Again, by adding this path to the initial tree  $T$ , we get a new tree with at least  $|A| + k - 1$  leaves, a contradiction. This means that  $w'u_2 \in E(G)$ , for any upper leaf  $w'$ . On the other hand, for any lower leaf  $w''$ , there exists a  $w'' - u_2$  path with no inner vertex in  $A \cup U$ . Finally, for any  $u_i \in U \setminus \{u_2\}$ , there exists a  $u_i - u_2$  path in  $G_r$  with no inner vertex in  $A \cup U$ . But then it is easy to construct a new tree rooted at  $u_2$  and with at least  $|A| + k - 1$  leaves.  $\square$

**Claim 7.**  *$r$  has at most one neighbour in  $V(T) \setminus (A \cup U)$ .*

Indeed, suppose  $r$  has at least two neighbours in  $V(T) \setminus (A \cup U)$ . The subtree  $T - (\{u_2, \dots, u_k\} \cup (A \setminus \{r\}))$  contains a leaf  $q$  and let  $Q = N_T(q) \cap (A \setminus \{r\})$ . Let  $w \in Q$  and consider  $G_w$ . Clearly,  $w$  has at least  $k - 1 \geq 1$  neighbours in  $V(G) \setminus (A \cup \{q\})$ . If for any  $w \in Q$ , one of these neighbours is contained in  $V(T) \setminus (\{u_2, \dots, u_k\} \cup \{q\})$ , then we get a tree rooted at  $r$  and in which  $q$  is an additional leaf. Therefore, there exists  $w \in Q$  with no neighbours in  $V(T) \setminus (\{u_2, \dots, u_k\} \cup \{q\})$ . By Claim 4,  $w$  has at most one neighbour in  $V(G) \setminus V(T)$ . But then  $w$  has at most  $k + 1$  neighbours in  $V(G) \setminus A$ , contradicting the minimality of  $r$ .  $\square$

By Claim 7 and Claim 6,  $r$  has exactly  $k + 1$  neighbours in  $V(T) \setminus A$ . Therefore, let  $N_T(r) \setminus (A \cup U) = \{z\}$  and let  $T_z$  be the subtree induced by  $z$  and its descendants.

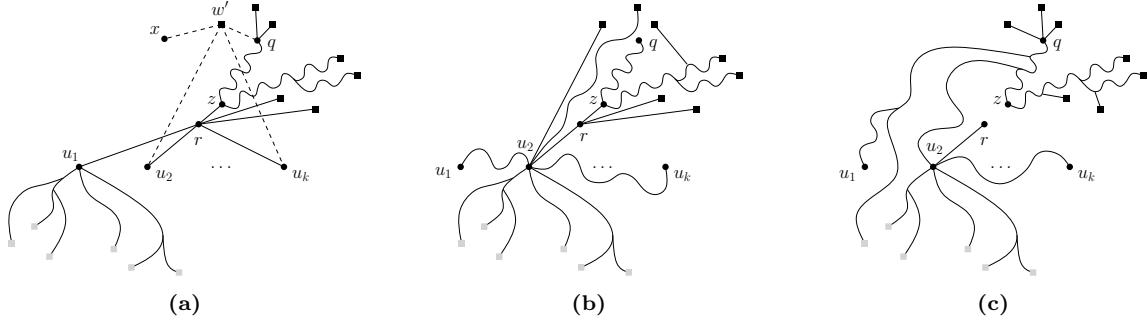
**Claim 8.**  $V(T_{u_1}) \cap V(T_z) = \emptyset$ .

Indeed, suppose there exists  $x \in V(T_{u_1}) \cap V(T_z)$  and let  $P(x)$  be the unique  $x - r$  path in  $T$ . By definition,  $P(x)$  contains both  $z$  and  $u_1$  and so a cycle arises in  $T$ .  $\square$

**Claim 9.** No leaf of  $T_z$  is a lower leaf for  $T$ .

Indeed, suppose there exists a lower leaf  $w$  for  $T$  which is a leaf of  $T_z$ . This means that the  $w - u_1$  subpath  $P$  in the definition of  $w$  intersects  $T_z$ . Then, select the  $u_1 - x$  subpath of  $P$ , where  $x$  is the first intersection of  $P$  with  $T_z$ , when traversed from  $u_1$ . Moreover, for any  $u_i \in U \setminus \{u_1\}$ , select a  $u_i - u_1$  path in  $G_r$  with no inner vertex in  $A \cup U$  and remove the edge set  $\{u_2r, \dots, u_kr, rz\}$ . After removing cycles and appropriate edges from the selected subgraph, we get a tree  $T'$  rooted at  $u_1$  and with at least  $|A| + k - 2$  leaves. Finally, by Claim 4 and minimality of  $r$ , any upper leaf for  $T$  adjacent to  $r$  has a neighbour in  $V(T') \setminus (U \setminus \{u_1\} \cup A)$  and so  $r$  could become an additional leaf for  $T'$ .  $\square$

Clearly,  $T_z - A$  contains a leaf  $q$  and let  $Q = N_T(q) \cap A$ . By Claim 9,  $Q$  is a set of upper leaves. By an argument similar to the one in the proof of Claim 7, there exists  $w' \in Q$  such that  $N_{G_{w'}}(w') \subseteq N_G(w') \setminus A \subseteq (U \setminus \{u_1\}) \cup \{q, x\}$ , for some  $x \in V(G) \setminus V(T)$ , and  $w'$  is adjacent to at least  $k - 2$  vertices in  $U \setminus \{u_1\}$  (see Figure 2(a)). Moreover,  $U \setminus \{u_1\} \subseteq V(G_{w'})$ , or else  $\{q, x\} \subseteq V(G_{w'})$  and there would exist an  $x - q$  path in  $G_{w'}$  with no inner vertex in  $A \cup (U \setminus \{u_1\})$ , thus turning  $x$  into an additional leaf for  $T$ .



**Figure 2:** (a) Neighbourhood of  $w'$ . (b) and (c) The different constructions of a tree  $T$  with at least  $|A| + k - 1$  leaves.

**Claim 10.** There is no path  $P$  from  $w'$  to  $v \in V(T_{u_1}) \setminus A$  with no inner vertex in  $A \cup (U \setminus \{u_1\})$ .

Suppose such a path  $P$  exists. Then  $w'x \notin E(P)$ , or else  $x$  becomes an additional leaf. Therefore,  $w'q \in E(P)$ . By the paragraph above,  $w'$  is adjacent to at least  $k - 2$  vertices in  $U \setminus \{u_1\}$ , say  $\{w'u_3, w'u_4, \dots, w'u_k\} \subseteq E(G)$ , and there exists a  $w' - u_2$  path  $P' \subseteq G_{w'}$  with no inner vertex in  $(U \setminus \{u_1, u_2\}) \cup \{q\}$ . If  $w'u_2 \in E(G)$ , then there exists a  $w', U$ -fan in  $G - (A \setminus \{w'\})$ , contradicting the fact that  $w'$  is an upper leaf. Therefore,  $w'u_2 \notin E(G)$  and  $w'x \in E(P')$ . But then  $P + P(v)$ , where  $P(v)$  is the unique  $v - u_1$  path in  $T$ , and  $P'$  do not intersect in an inner vertex, or else  $x$  becomes an additional leaf, and once again we obtain a  $w', U$ -fan in  $G - (A \setminus \{w'\})$ , a contradiction.  $\square$

**Claim 11.** Each  $w \in A$  has at most one neighbour in  $T_{u_1} - A$ .

Consider  $w \in A \setminus \{w'\}$  such that  $w'w \notin E(G)$ . There exists a  $w' - w$  path  $P$  in  $G_{w'w}$  with no inner vertex in  $A \cup (U \setminus \{u_1\})$ . Moreover, by Claim 10,  $P$  does not contain any vertex of  $T_{u_1} - A$ . For any  $w \in A \setminus \{w'\}$  such that  $w'w \notin E(G)$ , we select these paths. For any  $w \in A \setminus \{w'\}$  such that  $w'w \in E(G)$ , we select the corresponding edges. Moreover, recall that  $w'$  is adjacent to at least  $k - 2$  vertices in  $U \setminus \{u_1\}$ , say  $\{u_3, \dots, u_k\}$ , and  $U \setminus \{u_1\} \subseteq V(G_{w'})$ . But then there exists a  $w' - u_2$  path in  $G_{w'}$  with no inner vertex in  $A \cup U$  and so, by Claim 10, with no inner vertex in  $T_{u_1} - A$  as well. Therefore, we can obtain a new tree  $T'$  rooted at  $w'$  and with at least  $|A| + k - 2$  leaves. But then each  $w \in A$  has at most one neighbour in  $T_{u_1} - A$ , or else we could get an additional leaf for  $T'$ .  $\square$

**Claim 12.** *If  $w$  is an upper leaf, then  $((U \setminus \{u_1\}) \cup \{v\}) \not\subseteq V(G_w)$ , for any  $v \in V(T_{u_1})$ .*

Indeed, suppose this is not the case and let  $v \in V(T_{u_1})$  be a vertex with minimum  $d_T(v, u_1)$  among those satisfying  $((U \setminus \{u_1\}) \cup \{v\}) \subseteq V(G_w)$ . Since the set  $U' = (U \setminus \{u_1\}) \cup \{v\}$  is contained in  $V(G_w)$  then, by the Fan Lemma [4], there exists a  $w, U'$ -fan in  $G_w$ . But then no path in the fan intersects the unique  $v - u_1$  path in  $T$  in an inner vertex and so we can obtain a  $w, U$ -fan in  $G - (A \setminus \{w\})$ , contradicting the fact that  $w$  is an upper leaf.  $\square$

**Claim 13.**  $V(G_w) \cap V(T_{u_1}) = \emptyset$ , for any  $w \in Q$ .

Indeed, suppose there exists  $w \in Q$  such that  $z \in V(G_w) \cap V(T_{u_1})$ . By Claim 12, we have  $U \setminus \{u_1\} \not\subseteq V(G_w)$ . By Claim 11,  $w$  has at most one neighbour in  $T_{u_1} - A$  and, by Claim 4,  $w$  has at most one neighbour in  $V(G) \setminus V(T)$ . But then there exists  $v \in V(G_w) \cap V(T_z)$ , or else  $N_{G_w}(w) = (U \setminus \{u_1, u_i\}) \cup \{x, y\}$ , for some  $x \in V(G) \setminus V(T)$ ,  $y \in V(T_{u_1}) \setminus A$  and  $2 \leq i \leq k$ , and we could find an  $x - y$  path in  $G_w$  with no inner vertex in  $A \cup U$ . Therefore, there exists a  $v - z$  path  $P \subseteq G_w$  with no inner vertex in  $A \cup (U \setminus \{u_1\})$ , contradicting Claim 10.  $\square$

We now show how to build a tree rooted at  $u_2$  and with at least  $|A| + k - 1$  leaves. This will provide a contradiction, thus concluding the proof.

Consider  $w \in Q$ . By Claim 13 and Claim 4, if  $V(G_w) \cap (V(T_z) \setminus \{q\}) = \emptyset$ , then  $N_{G_w}(w) \subseteq (U \setminus \{u_1\}) \cup \{q, x\}$ , for some  $x \in V(G) \setminus V(T)$ . Moreover, as we have seen before, we have that  $U \setminus \{u_1\} \subseteq V(G_w)$ , or else  $x$  could become an additional leaf for  $T$ .

We start by removing  $V(T_{u_1}) \cup U$  from  $T$ . Suppose now that, for every lower leaf  $w''$ , the  $w'' - u_2$  subpath  $P$  in the definition of  $w''$  contains no vertices in  $T_z$  and, for every  $u_i \in U \setminus \{u_2\}$ , each  $u_i - u_2$  path  $P'$  in  $G_r$  with no inner vertex in  $A \cup U$  contains no vertices in  $T_z$ . In particular, these paths do not contain  $q$  and we select them (see Figure 2(b)). Moreover, for any  $w \in Q$ , we do the following. If there exists  $v \in V(G_w) \cap (V(T_z) \setminus \{q\})$ , then we select a  $w - v$  path in  $G_w$  with no inner vertex in  $A \cup (U \setminus \{u_2\}) \cup \{q\}$  (such a path exists by Claim 13). Otherwise, by the paragraph above, we select a  $w - u_2$  path in  $G_w$  with no inner vertex in  $A \cup U \cup \{q\}$  (again, the existence follows from Claim 13). After removing cycles and appropriate edges we get a new tree  $T$ , rooted at  $u_2$  and with at least  $|A| + k - 1$  leaves ( $q$  becomes an additional leaf), a contradiction.

Therefore, there exists either a lower leaf  $\bar{w}$  such that the  $\bar{w} - u_2$  subpath  $P$  in the definition of  $\bar{w}$  contains vertices in  $T_z$ , or a  $u_i \in U \setminus \{u_2\}$  such that a  $u_i - u_2$  path  $P'$  in  $G_r$  with no inner vertex in  $A \cup U$  contains vertices in  $T_z$ . Then we select either the  $\bar{w} - x$  subpath and the  $y - u_2$  subpath of  $P$ , where  $x$  and  $y$  are, respectively, the first and the last intersection of  $P$  with  $T_z$  when traversed from  $\bar{w}$ , or the  $u_i - x'$  subpath and the  $y' - u_2$  subpath of  $P'$ , where  $x'$  and  $y'$  are, respectively, the first and the last intersection of  $P'$  with  $T_z$  when traversed from  $u_i$  (see Figure 2(c)). In this way, we get a new tree  $T'$  and we grow it as follows. Let  $w''$  be a lower leaf and  $P_{w''}$  be the  $w'' - u_2$  subpath in the definition of  $w''$ . For any lower leaf  $w''$ , we select the  $w'' - x_{w''}$  subpath of  $P_{w''}$ , where  $x_{w''}$  is the first intersection of  $P_{w''}$  with the tree constructed so far and in which  $w''$  becomes a leaf. Similarly, we add the remaining vertices of  $U$  as leaves. Finally, consider an upper leaf  $w''$  for the original  $T$  such that  $d_T(w'', r) = 1$ . By minimality of  $r$  and by an argument similar to Claim 4,  $w''$  is adjacent to a vertex in  $V(T') \setminus (A \cup U \setminus \{u_2\})$ . But then we can add these edges to the tree  $T'$  rooted at  $u_2$  in order to obtain a new tree with at least  $|A| + k - 1$  leaves ( $r$  becomes an additional leaf), a contradiction. This concludes the proof.  $\square$

Our bound is tight in the sense of the following.

**Proposition 14.** *For any  $k \geq 2$  and  $x \geq 2k$ , there exists a graph  $G$  with  $\ell(G) = x$  and  $\text{VC}_{k-\text{con}}(G) = \ell(G) - k + 1$ .*

For the proof we need the following elementary result which will be used in the upcoming sections as well.

**Lemma 15 (Expansion Lemma [4]).** *If  $G$  is a  $k$ -connected graph and  $G'$  is obtained from  $G$  by adding a new vertex with at least  $k$  neighbours in  $G$ , then  $G'$  is  $k$ -connected.*

*Proof of Proposition 14.* For a fixed  $k \geq 2$  and  $x = 2k$ , consider the graph  $G_k$  constructed as follows. Start with a clique  $H_k$  of size  $k+1$ . For each subset  $S \subset H_k$  of size  $k$ , add a vertex adjacent to precisely the vertices of  $S$ , and let  $A$  be the set of the added vertices. Clearly,  $|V(G_k)| = 2k+2$  and  $\ell(G_k) = 2k$ . Moreover, by the Expansion Lemma,  $A$  is shattered. For  $x \geq 2k$ , let  $G$  be the graph obtained from  $G_k$  by adding  $x - 2k$  vertices adjacent to exactly  $k$  vertices of  $H_k$ . It is easy to see that  $\ell(G) = x$  and  $\text{VC}_{k\text{-con}}(G) = \ell(G) - k + 1$ .  $\square$

As for a lower bound, we use the fact that having a sufficiently large complete subgraph guarantees shattering by  $k$ -connected subgraphs.

**Theorem 16.** *Let  $G$  be a connected graph of order  $n$ , size  $m$ , and maximum degree  $\Delta$ . For  $k \geq 2$ ,*

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

*Proof.* By Turán's Theorem [4], if  $m > \left(1 - \frac{1}{r}\right) \frac{n^2}{2}$ , then  $G$  contains  $K_{r+1}$  as a subgraph. Therefore, by the Expansion Lemma, 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

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

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

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

Using the two relations above, it is easy to see that

$$n - d_1 = \sum_{i=2}^{\Delta} d_i \geq \sum_{i=2}^{\Delta} \frac{i-1}{\Delta-1} d_i = \frac{1}{\Delta-1} \sum_{i=2}^{\Delta} (i-1) d_i = \frac{n-2}{\Delta-1}.$$

Since  $\ell(G)$  is the maximum of  $d_1$  taken over all spanning trees of  $G$ , then

$$\ell(G) \leq n - \left\lceil \frac{n-2}{\Delta-1} \right\rceil.$$

Summarizing, we get

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

Note that, in Theorem 16, equality is attained by complete graphs.

### 3. The decision problem

In this section we investigate the computational complexity of  $\text{GRAPH VC}_{k\text{-con}} \text{ DIMENSION}$ . Consider the following decision problem, usually called SET MULTICOVER:

SET MULTICOVER

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

Being a generalization of the well-known SET COVER (also known as MINIMUM COVER [9]), it is NP-complete. We use it in the proof of the following Theorem. Recall that a *split graph* is a graph in which the vertex set can be partitioned into a clique and an independent set. Each complete graph is clearly a split graph and the class of split graphs coincides with that of  $(2K_2, C_4, C_5)$ -free graphs (see [13]).

**Theorem 17.** GRAPH  $VC_{k-\text{con}}$  DIMENSION is NP-complete even for split graphs.

*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 18.** 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 of Lemma 18.* Let  $S \subset V(G \cup G')$  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.  $\square$

Let  $G = (V, E)$  and  $s \geq 1$  be an instance of GRAPH  $VC_{k-\text{con}}$  DIMENSION. We claim we can check in polynomial time whether a subset  $V' \subseteq V$  with  $|V'| \geq s$  is shattered. Indeed, by Lemma 18, it is enough to check all the  $\mathcal{O}(|V|^{k+1})$  subsets of  $V'$  of size at most  $k + 1$ . Recall that finding a minimum separating set of a graph  $G$  is polynomial in the order of  $G$ . Moreover, if  $S \subseteq V(G)$  is a minimum separating set and  $A \cup B = V(G - S)$  is a partition into two non-empty sets such that any path from a vertex in  $A$  to a vertex in  $B$  contains a vertex in  $S$  then, for  $k > |S|$ , the vertices of any  $k$ -connected subgraph of  $G$  are entirely contained in either  $A \cup S$  or  $B \cup S$ . These observations, as shown in [14, Theorem 1], allow to test whether  $G$  has a  $k$ -connected subgraph in polynomial time. Therefore, for any  $B \subseteq V'$  of size at most  $k + 1$ , we can check in polynomial time if there exists a  $k$ -connected subgraph contained in  $G - (V' \setminus B)$  and containing  $B$ .

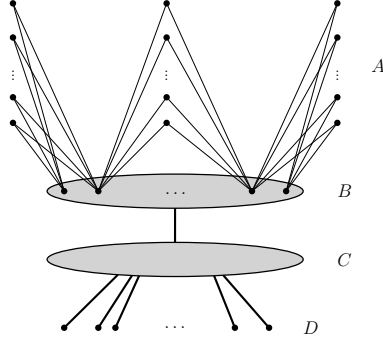
Now we prove NP-hardness by a reduction from SET MULTICOVER. Given an instance of SET MULTICOVER, we construct a graph  $G = (V, E)$  as follows (see Figure 3). The set of vertices  $V$  is formed by four pairwise disjoint sets  $A, B, C$  and  $D$ .  $A$  is an independent set of  $n \cdot (t + k + 1)$  vertices arranged in  $n$  columns of  $t + k + 1$  vertices each (every element in the  $j$ -th column corresponds to a copy of  $a_j$ ),  $B = \{v_1, \dots, v_m\}$  is a clique ( $v_i$  corresponds to the set  $S_i$ ),  $C$  is a clique of size  $k$  and  $D$  is an independent set of  $t + m + 1$  vertices. Each vertex in  $C$  is connected to all vertices in  $B$  (therefore,  $B \cup C$  is a clique of size  $m + k$ ) and  $D$ . Finally,  $v_i \in B$  is connected to every copy of  $a_j \in A$  if and only if  $a_j \in S_i$ .

Since  $B \cup C$  is a clique and  $A \cup D$  is an independent set, then  $G$  is a split graph. We claim that there is an index set  $I \subseteq \{1, \dots, m\}$  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  $VC_{k-\text{con}}(G) \geq |V| - (t + k)$ .

Suppose first 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, the subgraph  $G' = G[C \cup \{v_i \in B : i \in I\}]$  is a clique of size at least  $k + 1$  and each vertex in  $V'$  has at least  $k$  neighbours in  $G'$ . Therefore,  $V'$  is shattered by the Expansion Lemma. Finally,  $|I| \leq t$  implies that  $|V'| \geq |V| - t - k$ .



**Figure 3:** The graph  $G$  for the reduction. The gray ovals are cliques. A thick edge joining a vertex  $v \in D$  to  $C$  means that  $v$  is adjacent to all the vertices of  $C$ . Similarly, the thick edge between the ovals means that  $B \cup C$  is a clique.

Conversely, let  $V'$  be a shattered set of maximum cardinality. Then  $|V'| \geq |V| - (t + k)$ . Suppose there exists  $c \in V' \cap C$ . Then no vertex in  $D$  can be shattered, and so  $|V'| \leq |V| - (t + m + 1) < |V| - (t + k)$ . Therefore, no vertex of  $C$  is in  $V'$  and  $D \subseteq V'$ . Moreover, at least one vertex  $v \in A$  for each column is in  $V'$  and so  $v$  has at least  $k$  neighbours in the clique  $B \setminus V'$ . By the Expansion Lemma and since all the vertices in the column of  $v$  have identical neighbourhoods, then each vertex in the column of  $v$  belongs to  $V'$  and so  $A \subseteq V'$ . 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 SET MULTICOVER. Indeed, any vertex of  $A$  has at least  $k$  neighbours in  $B \setminus V'$ . In other words, each  $a_j \in S$  is contained in at least  $k$  of the subsets  $S_i$  with  $i \in I$ . Moreover,  $|I| \leq m - (m - t) = t$ .  $\square$

#### 4. Graphs of bounded clique-width

Graphs of bounded tree-width are particularly interesting from an algorithmic point of view: a lot of NP-complete problems can be solved in linear time for them. For example, all graph properties which are expressible in monadic second-order logic with edge-set quantification are decidable in linear time on graphs of bounded tree-width. This is the content of the celebrated algorithmic meta-theorem of Courcelle [7]. Let us briefly recall that monadic second-order logic is an extension of first-order logic by quantification over sets. The language of *monadic second-order logic of graphs* ( $\text{MSO}_1$  in short) contains the expressions built from the following elements:

- variables  $x, y, \dots$  for vertices and  $X, Y, \dots$  for sets of vertices;
- predicates  $x \in X$  and  $\text{adj}(x, y)$ ;
- equality for variables, standard Boolean connectives and the quantifiers  $\forall$  and  $\exists$ .

By considering edges and sets of edges as other sorts of variables and the incidence predicate  $\text{inc}(v, e)$ , we obtain *monadic second-order logic of graphs with edge-set quantification* ( $\text{MSO}_2$  in short).

If a graph property is expressible in the more restricted  $\text{MSO}_1$ , then Courcelle et al. [15] showed that it is decidable in linear time even on graphs of bounded clique-width, assuming a  $k$ -expression of the graph is explicitly given. The following observation shows that, for any graph, being shattered by its  $k$ -connected subgraphs is a property that makes no exception to this framework.

**Lemma 19.** *Being shattered by  $k$ -connected subgraphs is expressible in  $\text{MSO}_1$ .*

*Proof.* Let  $G = (V, E)$  be a graph. The following  $\text{MSO}_2$  sentence says that the subgraph induced by  $X \subseteq V$  is connected:

$$\text{conn}(X) = \forall Y \subseteq V [(\exists u \in X \ u \in Y \wedge \exists v \in X \ v \notin Y) \rightarrow (\exists e \in E \exists u \in X \exists v \in X \ \text{inc}(u, e) \wedge \text{inc}(v, e) \wedge u \in Y \wedge v \notin Y)].$$

It is easy to see that the quantification over single edges can be expressed by a  $\text{MSO}_1$  sentence as follows:

$$\exists a \in V \exists b \in V \exists u \in X \exists v \in X \text{ adj}(a, b) \wedge (u = a \vee u = b) \wedge (v = a \vee v = b) \wedge u \in Y \wedge v \notin Y.$$

The following  $\text{MSO}_1$  sentence says that the subgraph induced by  $X \subseteq V$  is  $k$ -connected:

$$\mathbf{k\text{-conn}}(X) = \exists v_1 \in X \dots \exists v_{k+1} \in X (\forall u_1 \in V \dots \forall u_{k-1} \in V \mathbf{conn}(X \setminus \{u_1, \dots, u_{k-1}\})).$$

Finally, the following  $\text{MSO}_1$  sentence says that the set  $A \subseteq V$  is shattered by  $k$ -connected subgraphs:

$$\mathbf{shatt}(A) = \forall B \subseteq A \exists X \subseteq V \mathbf{k\text{-conn}}(X) \wedge X \cap A = B. \quad \square$$

Therefore, as an immediate consequence of the meta-theorem stated above, we have the following.

**Corollary 20.**  $\text{GRAPH VC}_{k\text{-con}} \text{ DIMENSION}$  is decidable in linear time for graphs of bounded clique-width.

As shown by Courcelle and Olariu [5], every graph of bounded tree-width has bounded clique-width but there are graphs of bounded clique-width having unbounded treewidth (for example, complete graphs). Therefore, clique-width can be viewed as a more general concept than tree-width. Moreover,  $P_4$ -free graphs, also known as *cographs*, are exactly those graphs having clique-width at most 2 [5]. See [6] for other classes of graphs of bounded clique-width.

We have seen that  $\text{GRAPH VC}_{k\text{-con}} \text{ DIMENSION}$  is NP-hard even for split graphs. Are there some subclasses of split graphs on which the problem becomes easy? Recall that the *Dilworth number* of a graph  $G$  is the size of a largest antichain (or, equivalently, the size of a minimum chain partition) with respect to the quasi-order  $\preceq$  defined on the vertices of  $G$  as follows:  $x \preceq y$  if and only if  $N_G(x) \subseteq N_G(y)$ . Graphs with Dilworth number 1 are precisely the well-known *threshold* graphs, which are the  $P_4$ -free split graphs. Therefore, they have clique-width at most 2 and we have seen we can decide the VC-dimension in linear time. On the other hand, already a small jump for the Dilworth number of a split graph, from 1 to 2, changes the clique-width from bounded to unbounded [16]. Nevertheless, deciding the VC-dimension remains easy.

**Theorem 21.**  $\text{GRAPH VC}_{k\text{-con}} \text{ DIMENSION}$  is decidable in polynomial time for split graphs with Dilworth number at most 2.

*Proof.* Let  $G = (V, E)$  be the input graph. We assume the unique partition of  $V$  into a clique of size  $\omega(G)$  and an independent set  $I$  of size  $\alpha(G)$  is given. It is well-known that the problem of finding a minimum chain partition of a poset can be translated into a maximum bipartite matching problem [17]. Therefore, we can find in polynomial time a chain partition  $I_1 \cup I_2$  of  $I$ . For  $j \in \{1, 2\}$ , let  $I_{j, \geq k} = \{u \in I_j : d(u) \geq k\}$  and  $I_{j, \succeq u} = \{v \in I_j : v \succeq u\}$ . Note that, if  $\omega(G) \leq k$ , then  $G$  contains no  $k$ -connected subgraph. Therefore, we may assume  $\omega(G) > k$ . But then, a maximum size shattered set containing vertices from at most one between  $I_1$  and  $I_2$  has size

$$\max\{\omega(G) - (k + 1) + |I_{1, \geq k}|, \omega(G) - (k + 1) + |I_{2, \geq k}|\}.$$

On the other hand, it is not difficult to see that, for any pair  $x \in I_{1, \geq k}$  and  $y \in I_{2, \geq k}$ , a maximum size shattered set containing  $x$  as the minimal element from  $I_{1, \geq k}$  and  $y$  as the minimal element from  $I_{2, \geq k}$  has size

$$\omega(G) - \max\{k + 1, 2k - |N(x) \cap N(y)|\} + |I_{1, \succeq x}| + |I_{2, \succeq y}|. \quad \square$$

It is not difficult to see that outerplanar graphs have tree-width at most 2, and so bounded clique-width as well. On the other hand, planar graphs, in general, do not have bounded clique-width, even if restricted to be bipartite and with maximum degree 3 (see [6, Lemma 1]). In the next section we investigate the complexity of  $\text{GRAPH VC}_{k\text{-con}} \text{ DIMENSION}$  when restricted to planar bipartite graphs with maximum degree at most 4.

## 5. Subclasses of planar graphs

For our reductions we need a series of auxiliary results. The following variant of HAMILTONIAN CIRCUIT THROUGH SPECIFIED EDGE was shown to be NP-complete in a comment on [18].

HAMILTONIAN CIRCUIT THROUGH SPECIFIED EDGE

**Instance:** A planar cubic bipartite graph  $G = (V, E)$  and  $e \in E$ .  
**Question:** Does  $G$  contain a Hamiltonian circuit through  $e$ ?

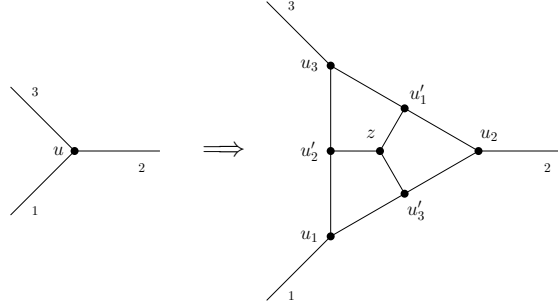
For completeness, we give the details of the reduction from the following variant of HAMILTONIAN CIRCUIT, shown to be NP-complete by Akiyama et al. [19].

HAMILTONIAN CIRCUIT

**Instance:** A planar cubic bipartite graph  $G = (V, E)$ .  
**Question:** Does  $G$  contain a Hamiltonian circuit?

**Theorem 22 (Labarre [18]).** HAMILTONIAN CIRCUIT THROUGH SPECIFIED EDGE is NP-complete even for planar cubic bipartite graphs.

*Proof.* Given a planar cubic bipartite graph  $G = (V, E)$ , we construct a graph  $G' = (V', E')$  by replacing a vertex  $u$  with the gadget depicted in Figure 4 and we set  $e = u'_2 z$ . Clearly,  $G'$  is a planar cubic bipartite graph.



**Figure 4:** Construction of the graph  $G'$ . The vertex  $u$  is replaced by a gadget containing 7 vertices.

It is easy to see that  $G'$  contains a Hamiltonian circuit through  $e$  if and only if  $G$  contains a Hamiltonian circuit.  $\square$

Now we strengthen some hardness results obtained by Douglas [20] and related to the problems HAMILTONIAN PATH and CONNECTED DOMINATING SET. We think they might be of independent interest.

**Theorem 23.** HAMILTONIAN PATH is NP-complete even for planar bipartite graphs where all the vertices have degree 3, except two, which have degree 1.

*Proof.* Let  $G = (V, E)$  and  $uv \in E$  be an instance of HAMILTONIAN CIRCUIT THROUGH SPECIFIED EDGE, where  $G$  is a planar cubic bipartite graph. Our reduction constructs a graph  $G' = (V', E')$  as follows. Let  $V' = V \cup \{a, b\}$ , where  $a, b$  are new vertices, and  $E' = (E \setminus \{uv\}) \cup \{au, bv\}$ . Clearly,  $G'$  is a planar bipartite graph where all the vertices have degree 3, except  $a$  and  $b$  of degree 1. It is easy to see that  $G$  has a Hamiltonian circuit through  $uv$  if and only if  $G'$  has a Hamiltonian path (between  $a$  and  $b$ ).  $\square$

By Theorem 23 and using the same reduction as in [20, Theorem 1] (if the graph  $G$  in [20, Theorem 1] is bipartite, then  $G'$  is bipartite too), we have the following.

**Theorem 24.** *Given a planar bipartite graph  $G = (V, E)$  with  $\Delta(G) = 3$ , it is NP-complete to decide if there exists a spanning tree  $T$  for  $G$  such that  $d_T(v)$  is either 1 or 3, for any  $v \in V$ .*

Moreover, by the Proofs of [20, Corollary 1 and Corollary 2], we have that the following variant of CONNECTED DOMINATING SET is NP-complete.

CONNECTED DOMINATING SET

**Instance:** A planar bipartite graph  $G = (V, E)$  with maximum degree 3.  
**Question:** Does there exist  $D \subseteq V$  such that  $G[D]$  is connected, every vertex in  $V \setminus D$  is adjacent to at least one vertex in  $D$  and  $|D| \leq \frac{|V|}{2} - 1$ ?

Now we are ready to prove the NP-hardness of GRAPH  $VC_{\text{con}}$  DIMENSION for the class of planar bipartite graphs with maximum degree 3. Given the connection between  $VC_{\text{con}}$  and the connected domination number, hidden in Theorem 2, it is no surprise we are going to reduce from the problem stated above.

**Theorem 25.** *GRAPH  $VC_{\text{con}}$  DIMENSION is NP-complete even for planar bipartite graphs with maximum degree 3.*

*Proof.* We prove NP-hardness by a reduction from the variant of CONNECTED DOMINATING SET introduced above. Let  $G = (V, E)$  be an instance of CONNECTED DOMINATING SET where  $G$  is a planar bipartite graph,  $\Delta(G) = 3$  and  $|V| = n$ . We claim  $G$  has a connected dominating set  $D$  with  $|D| \leq \frac{n}{2} - 1$  if and only if  $VC_{\text{con}}(G) \geq \frac{n}{2} + 1$ . Clearly, we may assume  $n \geq 46$ .

Suppose first  $G$  has a connected dominating set  $D$  with  $|D| \leq \frac{n}{2} - 1$ . Then each vertex in  $V \setminus D$  can be joined to  $G[D]$  independently of one another, and so  $V \setminus D$  can be shattered by connected subgraphs. Therefore,  $VC_{\text{con}}(G) \geq |V \setminus D| \geq \frac{n}{2} + 1$ .

Conversely, suppose  $VC_{\text{con}}(G) \geq \frac{n}{2} + 1$  and let  $A$  be a shattered set of maximum size. Recall that, by assumption, we have  $|A| \geq 24$ . The idea is to show there exists a connected component  $C$  of  $G - A$  such that each vertex of  $A$  is joined to  $C$ . Claim 26 would then imply that  $|C| \geq |A| - 2 \geq \frac{n}{2} - 1$  and so  $C$  would be a connected dominating set of size  $\frac{n}{2} - 1$ . We now elaborate on this idea in a series of claims, each one followed by a short proof.

**Claim 26.** *Each connected component  $C$  of  $G - A$  has at most  $|C| + 2$  neighbours in  $A$ .*

Indeed, if  $C$  contains at most two 1-vertices, then the claim clearly holds. Therefore, let  $\{u_1, \dots, u_k\}$ , with  $k \geq 3$ , be the set of 1-vertices in  $C$ . Since  $C$  is connected, there exists a  $u_1 - u_2$  path  $P$  in  $C$ . Moreover, any  $u_3 - u_1$  path intersects  $P$  in a 3-vertex. Applying this reasoning again, we have that  $d_1(C) - 2 \leq d_3(C)$ , where  $d_i = |\{v \in C : d_C(v) = i\}|$ . But then  $C$  has at most  $2d_1(C) + d_2(C) = |C| + d_1(C) - d_3(C) \leq |C| + 2$  neighbours in  $A$ .  $\square$

Since  $\Delta(G) = 3$ , then each vertex  $u \in A$  has at least one neighbour in  $G - A$ , otherwise it would not be possible to shatter  $u$  and a vertex in  $A \setminus N(u)$ . Let  $C_1, \dots, C_k$  be the connected components of  $G - A$ .

**Claim 27.** *There exists a vertex in  $A$  joined to less than three connected components.*

Indeed, suppose each vertex in  $A$  is joined to exactly three connected components. By Claim 26 and double counting the size of the edge cut  $[A, \overline{A}]$ , we have  $3|A| \leq \sum (|C_i| + 2)$ . Therefore,

$$k \geq \frac{3|A| - \sum |C_i|}{2} = \frac{3|A| - (n - |A|)}{2} \geq \frac{n}{2} + 2,$$

contradicting the fact that  $|V \setminus A| \leq \frac{n}{2} - 1$ .  $\square$

**Claim 28.** *No vertex in  $A$  is joined to exactly two connected components.*

Suppose, to the contrary, there exists  $u \in A$  joined to exactly two connected components, say  $C_1$  and  $C_2$ . Then, every vertex in  $A \setminus N(u)$  is joined to either  $C_1$  or  $C_2$ . By Claim 26, we have  $|C_1| + |C_2| + 3 \geq |A| - 1$ , from which  $|C_1| + |C_2| \geq \frac{n}{2} - 3$  and so  $\sum_{i=3}^k |C_i| \leq 2$ .

We now claim that the (at least  $|A| - 1$ ) vertices in  $A \setminus N(u)$  are all joined to  $C_1$  or all joined to  $C_2$ . Suppose, to the contrary, there exist  $v$  and  $w$  in  $A \setminus N[u]$  such that  $v$  is joined to  $C_1$ ,  $w$  is joined to  $C_2$  but none of them is joined to both  $C_1$  and  $C_2$ . Let  $B \subseteq A$  be the subset of vertices which are joined to both  $C_1$  and  $C_2$ . Since  $|C_1| + 2 \geq |B|$  and  $|C_2| + 2 \geq |B|$ , then  $|B| \leq \frac{|C_1| + |C_2|}{2} + 2 \leq \frac{n+6}{4}$ . Therefore, at least  $\frac{n}{2} + 1 - \frac{n+6}{4} = \frac{n-2}{4}$  vertices of  $A$  are not joined to both  $C_1$  and  $C_2$ . But then the set  $B'$  of vertices of  $A \setminus (N[v] \cup N[w])$  which are not joined to both  $C_1$  and  $C_2$  has size at least  $\frac{n-2}{4} - 6 \geq 5$ . Since  $\{x, v\}$  and  $\{x, w\}$  are shattered, for any  $x \in B'$ , then each vertex in  $B'$  is joined to some connected component different from  $C_1$  and  $C_2$ , contradicting the fact that the remaining connected components can be joined to at most four vertices in  $A$ .

Therefore, the (at least  $|A| - 1$ ) vertices in  $A \setminus N(u)$  are all joined to the same connected component, say  $C_1$ . Suppose now there exists  $u' \in N(u)$  not joined to  $C_1$ . By Claim 26, we have  $|C_1| \geq \frac{n}{2} - 2$ . Moreover, the set  $\{u', w\}$  is shattered, for any  $w \in A \setminus \{u, u'\}$ , contradicting the fact that the remaining component has at most 3 neighbours in  $A$ . Therefore, each  $v \in A$  is joined to  $C_1$ . Since  $|C_1| \geq |A| - 2 \geq \frac{n}{2} - 1$ , then  $C_2 = \emptyset$ , a contradiction.  $\square$

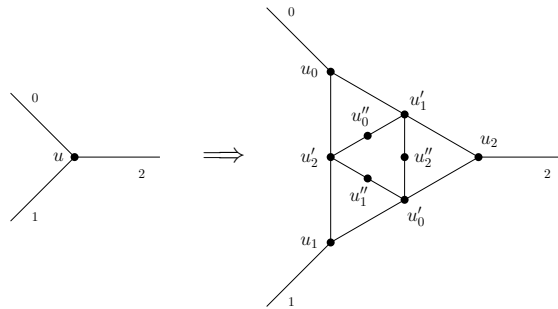
By Claim 27 and Claim 28, there exists  $v \in A$  joined to exactly one connected component  $C$  of  $G - A$ . Then  $v$  has at most two neighbours in  $A$ . Moreover, each of the (at least  $|A| - 3$ ) nonneighbours of  $v$  in  $A$  is joined to  $C$ , otherwise it would not be possible to shatter the set  $\{v, w\}$ , for some  $w \in A \setminus N[v]$ . Suppose there exists  $v' \in N(v) \cap A$  not joined to  $C$ . By Claim 26, we have  $|C| + 2 \geq |A| - 2$ , from which  $|C| \geq \frac{n}{2} - 3$  and so the remaining connected components contain at most two vertices. Since the set  $\{v', w\}$  is shattered, for any  $w \in A \setminus N[v']$ , and there are at least  $|A| - 3 \geq 21$  such nonneighbours, we get a contradiction to the fact that a component in  $G - A - C$  can have at most four neighbours in  $A$ . Therefore, each  $v \in A$  is joined to  $C$  and  $|C| \geq |A| - 2 \geq \frac{n}{2} - 1$ . But then  $|C| = \frac{n}{2} - 1$  and  $C$  is a connected dominating set for  $G$ .  $\square$

Clearly, GRAPH  $\text{VC}_{\text{con}}$  DIMENSION becomes easy for graphs  $G$  with  $\Delta(G) \leq 2$ . Indeed,  $\text{VC}_{\text{con}}(P_n) = 2$ , for  $n \geq 3$ , and  $\text{VC}_{\text{con}}(C_3) = 2$  and  $\text{VC}_{\text{con}}(C_n) = 3$ , for  $n \geq 4$ . Note that, if  $\Delta(G) \leq 2$ , then  $G$  has tree-width at most 2. Therefore, the conclusion follows from Corollary 20 as well.

We conclude by showing the NP-hardness of GRAPH  $\text{VC}_{2-\text{con}}$  DIMENSION for planar bipartite graphs with  $\Delta(G) = 4$ . We leave as an open problem to determine what happens if we further impose  $\Delta(G) = 3$ .

**Theorem 29.** GRAPH  $\text{VC}_{2-\text{con}}$  DIMENSION is NP-complete even for planar bipartite graphs with maximum degree 4.

*Proof.* Our reduction is from HAMILTONIAN CIRCUIT, which remains NP-complete even for planar cubic bipartite graphs [19]. Given an instance  $G = (V, E)$  of HAMILTONIAN CIRCUIT, where  $G$  is a planar cubic bipartite graph with  $|V| = n$ , we construct a graph  $G'$  by replacing each vertex  $u$  of  $G$  with the gadget depicted in Figure 5.



**Figure 5:** Construction of the graph  $G'$ . The vertex  $u$  is replaced by a gadget containing 9 vertices.

For  $0 \leq i \leq 2$ , the vertices  $u_i$  are the *gates* and the vertices  $u'_i$  are the *connectors*. Finally, each pair  $u'_i u'_{i+1}$  (indices modulo 3) of connectors in the gadget is connected by a path of length 2 with inner vertex  $u''_{i+2}$ , called a *crossing vertex*. Clearly, the construction can be done in polynomial time and the resulting graph  $G' = (V', E')$  is planar, bipartite and  $\Delta(G') = 4$ . We claim that  $\text{VC}_{2\text{-con}}(G') \geq |V'| - 5n$  if and only if  $G$  contains a Hamiltonian circuit.

Suppose first  $G$  contains a Hamiltonian circuit  $C$ . Without loss of generality,  $u \in V$  is incident to the edges 1 and 2 in  $C$ . Then we augment the subgraph induced by  $E(C)$  in  $G'$  with the path  $u_1 u'_0 u_2$ . Repeating this procedure for every vertex of  $G$ , we get a circuit in  $G'$  containing three vertices from every gadget. The vertices in  $\{u_0, u''_0, u'_1, u''_1\}$  can be joined to this circuit, independently of one another, via paths through the connectors  $u'_0, u'_1$  and  $u'_2$ . In all cases, the resulting subgraph is clearly 2-connected. Repeating this process for every gadget, we have that a set of size  $|V'| - 5n$  can be shattered.

Suppose now  $\text{VC}_{2\text{-con}}(G') \geq |V'| - 5n$  and let  $A$  be a shattered set of maximum cardinality.

**Claim 30.** *For any gadget  $H \subseteq G'$ , exactly one gate and the three crossing vertices are in  $A$ .*

We show first that at most four vertices of  $H$  are in  $A$ . Suppose, to the contrary,  $H$  contains at least five vertices of  $A$ . Then at least one crossing vertex is in  $A$ , otherwise at least two gates would be in  $A$ , contradicting the fact that the set consisting of a connector in  $H$  and a vertex not in  $H$  is shattered. Therefore, at least one crossing vertex is in  $A$ , say without loss of generality  $u''_1$ . Then the connectors  $u'_0$  and  $u'_2$  are not in  $A$ . If another crossing vertex is in  $A$ , then  $u'_1 \notin A$  and at least two gates are in  $A$ , a contradiction. Therefore,  $u''_1$  is the only crossing vertex in  $A$ . But then all the gates are in  $A$ , a contradiction again.

Since  $\text{VC}_{2\text{-con}}(G') \geq |V'| - 5n$ , then exactly four vertices per gadget are in  $A$ . We have seen that at most one gate per gadget is in  $A$ . Moreover, exactly one gate per gadget is in  $A$ , otherwise a crossing vertex and one of its neighbouring connectors would both be in  $A$ . Let  $u_0$  be the gate of gadget  $H$  in  $A$ . Suppose one of its neighbouring connectors is in  $A$  (clearly, there exists at most one such connector). But then, again, a crossing vertex and at least one of its neighbouring connectors would both be in  $A$ , a contradiction. Therefore, both  $u'_1$  and  $u'_2$  are not in  $A$  and it is easy to see that it must be  $A \cap V(H) = \{u_0, u''_0, u'_1, u'_2\}$ .  $\square$

By Claim 30, there exists a 2-connected subgraph of  $G'$  containing crossing vertices in every gadget and avoiding exactly one gate per gadget. Therefore, for any gadget, this subgraph contains exactly two of the edges incident to its gates and originally in  $G$ . This means that, contracting each gadget to a single vertex, we obtain a 2-regular connected spanning subgraph. Therefore,  $G$  contains a Hamiltonian circuit.  $\square$

## 6. Conclusion

This work represents a continuation of the systematic study of the VC-dimensions of graphs initiated by Kranakis et al. [3]. We have concentrated on the VC-dimension with respect to  $k$ -connected subgraphs. In particular, we have proved the NP-completeness of the associated decision problem and observed its decidability in linear time for graphs of bounded clique-width. In this context, we believe two interesting open problems arise: determine the complexity for unit interval graphs and for planar graphs in the remaining cases  $3 \leq k \leq 5$ . We conjecture that the first problem is in P. Finally, it would be interesting to focus on the study of the VC-dimension with respect to other classes of subgraphs.

## Acknowledgement

The author would like to thank the anonymous referee for useful comments improving the readability of the paper.

## References

- [1] V. N. Vapnik, A. Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, *Theory of Probability and Its Applications* 16 (2) (1971) 264–280.
- [2] D. Haussler, E. Welzl,  $\varepsilon$ -Nets and Simplex Range Queries, *Discrete & Computational Geometry* 2 (1987) 127–151.
- [3] E. Kranakis, D. Krizanc, B. Ruf, J. Urrutia, G. J. Woeginger, The VC-dimension of set systems defined by graphs, *Discrete Applied Mathematics* 77 (3) (1997) 237–257.
- [4] D. B. West, *Introduction to Graph Theory*, Prentice Hall, 2 edn., 2001.
- [5] B. Courcelle, S. Olariu, Upper bounds to the clique width of graphs, *Discrete Applied Mathematics* 101 (1–3) (2000) 77–114.
- [6] M. Kamiński, V. V. Lozin, M. Milanič, Recent developments on graphs of bounded clique-width, *Discrete Applied Mathematics* 157 (12) (2009) 2747–2761.
- [7] B. Courcelle, The Monadic Second-Order Logic of Graphs. I. Recognizable Sets of Finite Graphs, *Information and Computation* 85 (1) (1990) 12–75.
- [8] M. Anthony, G. Brightwell, C. Cooper, The Vapnik-Chervonenkis Dimension of a Random Graph, *Discrete Mathematics* 138 (1) (1995) 43–56.
- [9] M. R. Garey, D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.
- [10] C. H. Papadimitriou, M. Yannakakis, On Limited Nondeterminism and the Complexity of the V-C Dimension, *Journal of Computer and System Sciences* 53 (2) (1996) 161–170.
- [11] S. Arora, B. Barak, *Computational Complexity - A Modern Approach*, Cambridge University Press, 2009.
- [12] M. Schaefer, Deciding the VC-dimension is  $\Sigma_3^P$ -complete, II, Tech. Rep. TR00-006, DePaul University, 2000.
- [13] M. C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, vol. 57 of *Annals of Discrete Mathematics*, Elsevier, 2 edn., 2004.
- [14] L. M. Kirousis, M. J. Serna, P. G. Spirakis, Parallel Complexity of the Connected Subgraph Problem, *SIAM J. Comput.* 22 (3) (1993) 573–586.
- [15] B. Courcelle, J. A. Makowsky, U. Rotics, Linear Time Solvable Optimization Problems on Graphs of Bounded Clique-Width, *Theory of Computing Systems* 33 (2) (2000) 125–150.
- [16] N. Korpelainen, V. V. Lozin, C. Mayhill, Split Permutation Graphs, *Graphs and Combinatorics* 30 (3) (2014) 633–646.
- [17] L. Lovász, M. D. Plummer, *Matching Theory*, North-Holland, 1986.
- [18] A. Labarre, Comment on "Complexity of finding 2 vertex-disjoint ( $|V|/2$ )-cycles in cubic graphs?" URL <http://cstheory.stackexchange.com/questions/6107/complexity-of-finding-2-vertex-disjoint-v-2-cycles-in-cubic-graphs>.
- [19] T. Akiyama, T. Nishizeki, N. Saito, NP-Completeness of the Hamiltonian Cycle Problem for Bipartite Graphs, *Journal of Information Processing* 3 (1980) 73–76.
- [20] R. J. Douglas, NP-completeness and degree restricted spanning trees, *Discrete Mathematics* 105 (1–3) (1992) 41–47.