Identifying codes and VC-dimension

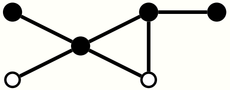
Aline Parreau
University of Liège, Belgium

Joint work with:
N. Bousquet, A. Lagoutte, Z. Li and S. Tomassé

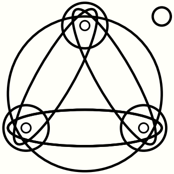
Combgraph Seminar - October 3, 2013

Contents

Identifying codes



VC-dimension

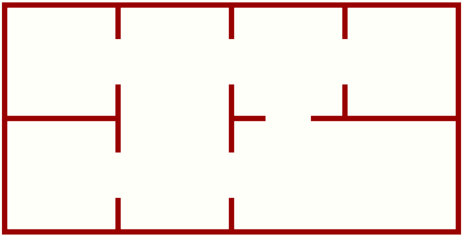


?

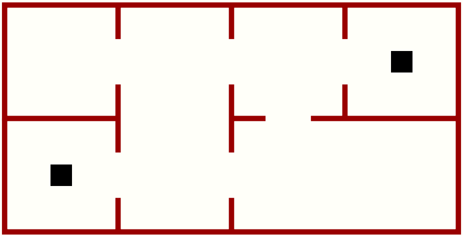
Part I

Identifying codes

Fire detection in a museum?

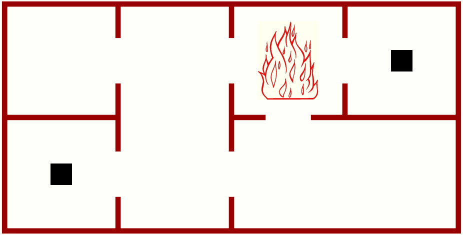


Fire detection in a museum?



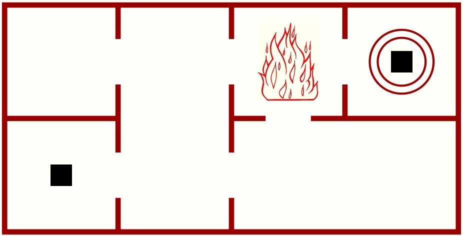
- Detector can detect fire in their room or in their neighborhood.

Fire detection in a museum?



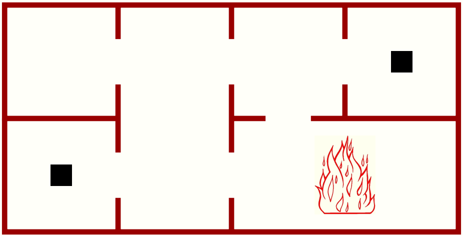
- Detector can detect fire in their room or in their neighborhood.

Fire detection in a museum?



- Detector can detect fire in their room or in their neighborhood.

Fire detection in a museum?

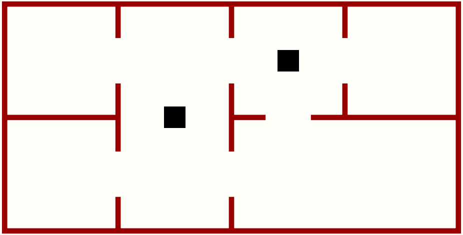


- Detector can detect fire in their room or in their neighborhood.

Fire detection in a museum?

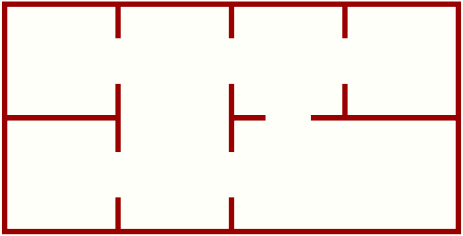
- Detector can detect fire in their room or in their neighborhood.
- Each room must contain a detector or have a detector in a neighboring room.

Fire detection in a museum?



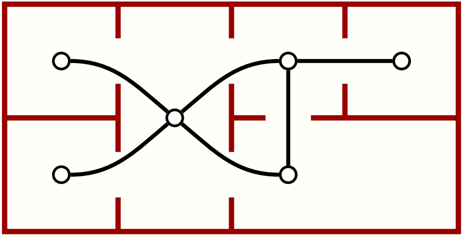
- Detector can detect fire in their room or in their neighborhood.
- Each room must contain a detector or have a detector in a neighboring room.

Modelization with a graph



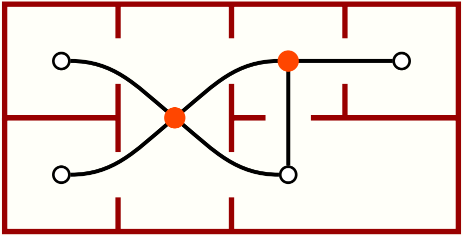
- Vertices V: rooms
- Edges E: between two neighboring rooms

Modelization with a graph



- Vertices V: rooms
- Edges E: between two neighboring rooms

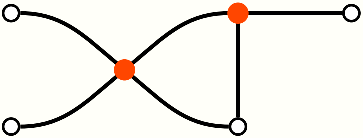
Modelization with a graph



- Vertices V: rooms
- Edges E: between two neighboring rooms
- Set of detectors = dominating set S:

$$\forall u \in V, N[u] \cap S \neq \emptyset$$

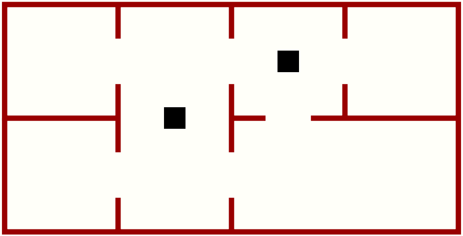
Modelization with a graph



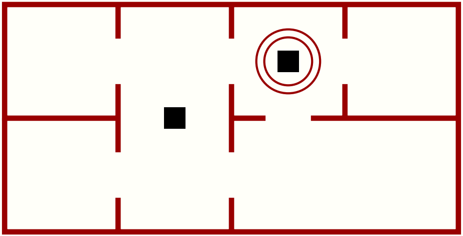
- Vertices V: rooms
- Edges E: between two neighboring rooms
- Set of detectors = dominating set S:

$$\forall u \in V, N[u] \cap S \neq \emptyset$$

Back to the museum

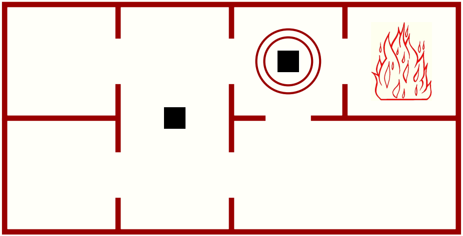


Back to the museum



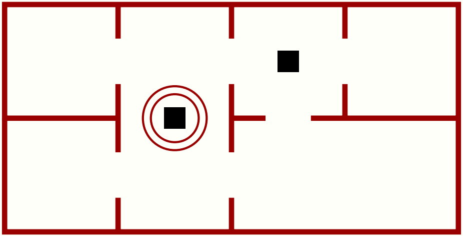
Where is the fire ?

Back to the museum



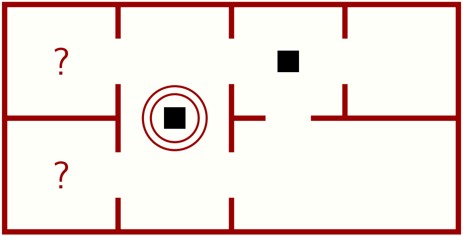
Where is the fire ?

Back to the museum



Where is the fire ?

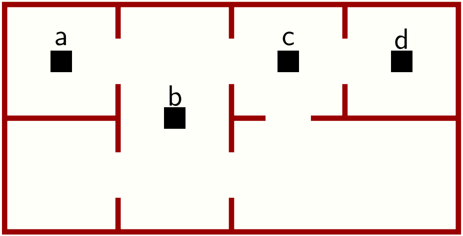
Back to the museum



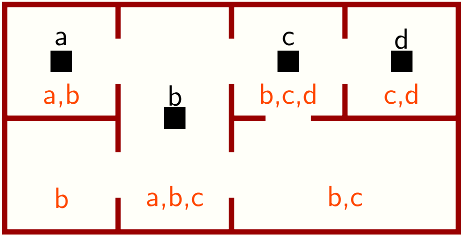
Where is the fire ?

To locate the fire, we need more detectors.

Identifying where is the fire



Identifying where is the fire

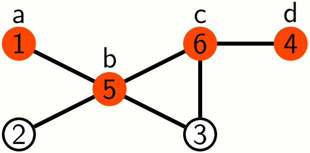


In each room, the set of detectors in the neighborhood is unique.

Modelization with a graph

Identifying code C = subset of vertices which is

- dominating : $\forall u \in V, N[u] \cap C \neq \emptyset$,
- separating : $\forall u, v \in V, N[u] \cap C \neq N[v] \cap C$.



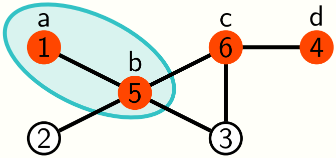
V \ C	a	b	c	d
1	•	•	-	-
2	-	•	-	-
3	-	•	•	-
4	-	-	•	•
5	•	•	•	-
6	-	•	•	•

Given a graph G, what is the size $\gamma^{ID}(G)$ of minimum identifying code ?

Modelization with a graph

Identifying code C = subset of vertices which is

- dominating : $\forall u \in V, N[u] \cap C \neq \emptyset$,
- separating : $\forall u, v \in V, N[u] \cap C \neq N[v] \cap C$.



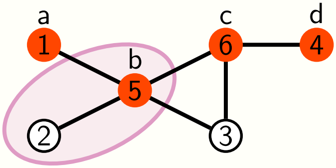
V \ C	a	b	c	d
1	•	•	-	-
2	-	•	-	-
3	-	•	•	-
4	-	-	•	•
5	•	•	•	-
6	-	•	•	•

Given a graph G, what is the size $\gamma^{ID}(G)$ of minimum identifying code ?

Modelization with a graph

Identifying code C = subset of vertices which is

- dominating : $\forall u \in V, N[u] \cap C \neq \emptyset$,
- separating : $\forall u, v \in V, N[u] \cap C \neq N[v] \cap C$.



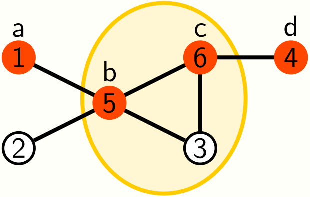
V \ C	a	b	c	d
1	•	•	-	-
2	-	•	-	-
3	-	•	•	-
4	-	-	•	•
5	•	•	•	-
6	-	•	•	•

Given a graph G, what is the size $\gamma^{ID}(G)$ of minimum identifying code ?

Modelization with a graph

Identifying code C = subset of vertices which is

- dominating : $\forall u \in V, N[u] \cap C \neq \emptyset$,
- separating : $\forall u, v \in V, N[u] \cap C \neq N[v] \cap C$.

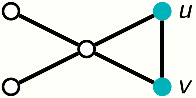


V \ C	a	b	c	d
1	•	•	-	-
2	-	•	-	-
3	-	•	•	-
4	-	-	•	•
5	•	•	•	-
6	-	•	•	•

Given a graph G, what is the size $\gamma^{ID}(G)$ of minimum identifying code ?

Some facts about identifying codes

- Introduced in 1998 by Karpvosky, Chakrabarty and Levitin
- Exists iff there is no twins



Twins: $N[u] = N[v]$

Some facts about identifying codes

- Introduced in 1998 by Karpvosky, Chakrabarty and Levitin
- Exists iff there is no twins
- NP-complete (Charon, Hudry, Lobstein, 2001)

Some facts about identifying codes

- Introduced in 1998 by Karpvosky, Chakrabarty and Levitin
- Exists iff there is no twins
- NP-complete (Charon, Hudry, Lobstein, 2001)
- Hard to approximate: best approximation factor $\log(|V|)$

Some facts about identifying codes

- Introduced in 1998 by Karpvosky, Chakrabarty and Levitin
- Exists iff there is no twins
- NP-complete (Charon, Hudry, Lobstein, 2001)
- Hard to approximate: best approximation factor $\log(|V|)$
- Lower bound:
  $\rightarrow$ A vertex is identified by a nonempty subset of $C \Rightarrow |V| \leq 2^{\gamma^{ID}(G)} - 1$

$$\gamma^{ID}(G) \geq \log(|V| + 1)$$

Tight example:

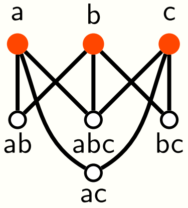
Some facts about identifying codes

- Introduced in 1998 by Karpvosky, Chakrabarty and Levitin
- Exists iff there is no twins
- NP-complete (Charon, Hudry, Lobstein, 2001)
- Hard to approximate: best approximation factor $\log(|V|)$
- Lower bound:

  $\rightarrow$ A vertex is identified by a nonempty subset of $C \Rightarrow |V| \leq 2^{\gamma^{ID}(G)} - 1$

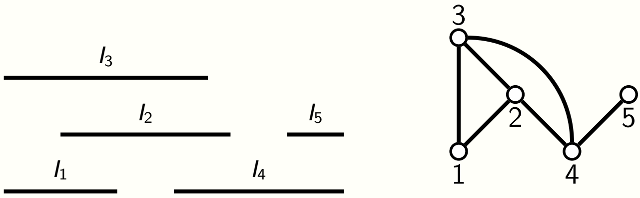
$$\gamma^{ID}(G) \geq \log(|V| + 1)$$

Tight example:



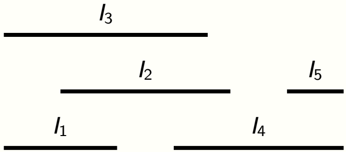
In restricted classes of graphs?

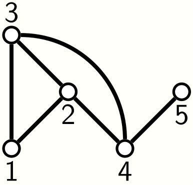
Example 1: Class of interval graphs



In restricted classes of graphs?

Example 1: Class of interval graphs



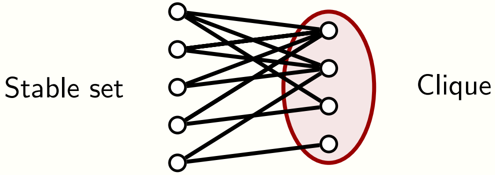


Proposition Foucaud, Naserasr, P., Valicov, 2012+

If G is an interval graph, $\gamma^{ID}(G) \geq \sqrt{2|V|}$.

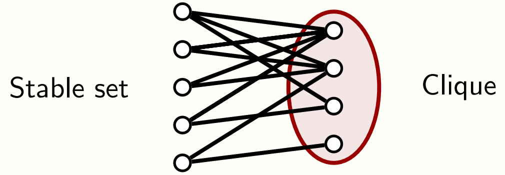
In restricted classes of graphs?

Example 2: Class of split graphs



In restricted classes of graphs?

Example 2: Class of split graphs



Proposition

For infinitely many split graphs G, $\gamma^{ID}(G) = \log(|V| + 1)$.

Proposition Foucaud, 2013

Min-Id-Code is log-APX-hard for split graphs.

Some known results in restricted classes of graphs

Restriction: classes of graphs closed by induced subgraphs

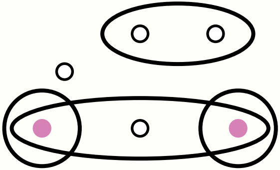
Graph class	lower bound (order)	Approximation
All	$\log n$	log APX-h
Split	$\log n$	log APX-h
Interval	$n^{1/2}$	open
Unit Interval	n	2
Bipartite	$\log n$	log APX-h
Line graphs	$n^{1/2}$	4
Chordal	$\log n$	log APX-h
Planar	n	7
Cograph	n	1

Part II

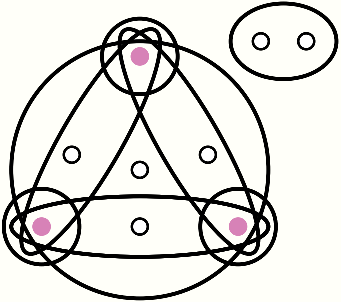
VC-dimension

Shattered set

- $\mathcal{H} = (V, \mathcal{E})$ an hypergraph
- A set $X \subseteq V$ is shattered if for all $Y \subseteq X$, there exists $e \in \mathcal{E}$, s.t $e \cap X = Y$.



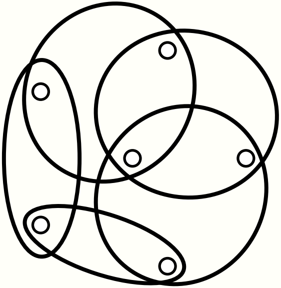
A 2-shattered set



A 3-shattered set

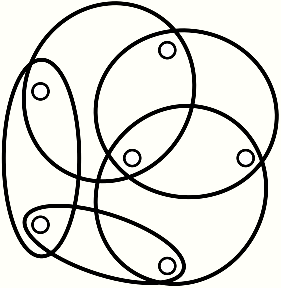
Vapnik Chervonenkis (VC) dimension of an hypergraph

- A set X is shattered if $\forall Y \subseteq X$, $\exists e \in \mathcal{E}$, s.t $e \cap X = Y$.
- VC-dimension of $\mathcal{H}$: largest size of a shattered set.



Vapnik Chervonenkis (VC) dimension of an hypergraph

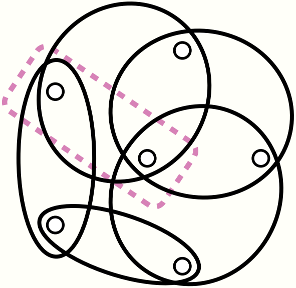
- A set X is shattered if $\forall Y \subseteq X$, $\exists e \in \mathcal{E}$, s.t $e \cap X = Y$.
- VC-dimension of $\mathcal{H}$: largest size of a shattered set.



No 3-shattered set $\Rightarrow$ VC-dim ≤ 2

Vapnik Chervonenkis (VC) dimension of an hypergraph

- A set X is shattered if $\forall Y \subseteq X$, $\exists e \in \mathcal{E}$, s.t $e \cap X = Y$.
- VC-dimension of $\mathcal{H}$: largest size of a shattered set.

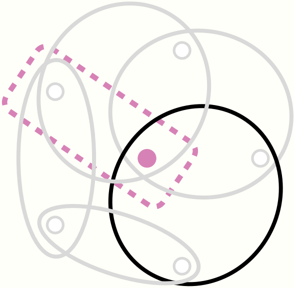


No 3-shattered set $\Rightarrow$ VC-dim ≤ 2

A 2-shattered set $\Rightarrow$ VC-dim $= 2$

Vapnik Chervonenkis (VC) dimension of an hypergraph

- A set X is shattered if $\forall Y \subseteq X$, $\exists e \in \mathcal{E}$, s.t $e \cap X = Y$.
- VC-dimension of $\mathcal{H}$: largest size of a shattered set.

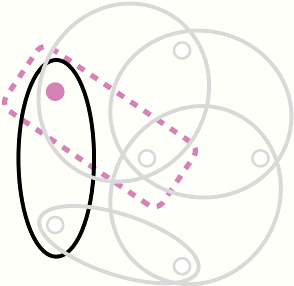


No 3-shattered set $\Rightarrow$ VC-dim ≤ 2

A 2-shattered set $\Rightarrow$ VC-dim $= 2$

Vapnik Chervonenkis (VC) dimension of an hypergraph

- A set X is shattered if $\forall Y \subseteq X$, $\exists e \in \mathcal{E}$, s.t $e \cap X = Y$.
- VC-dimension of $\mathcal{H}$: largest size of a shattered set.

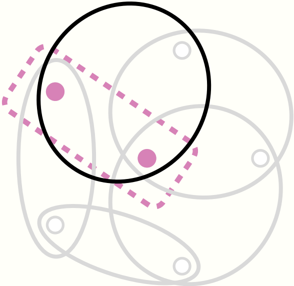


No 3-shattered set $\Rightarrow$ VC-dim ≤ 2

A 2-shattered set $\Rightarrow$ VC-dim $= 2$

Vapnik Chervonenkis (VC) dimension of an hypergraph

- A set X is shattered if $\forall Y \subseteq X$, $\exists e \in \mathcal{E}$, s.t $e \cap X = Y$.
- VC-dimension of $\mathcal{H}$: largest size of a shattered set.

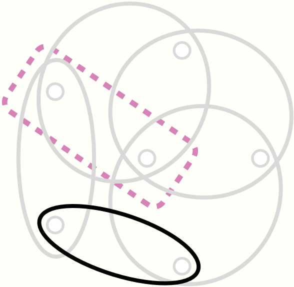


No 3-shattered set $\Rightarrow$ VC-dim ≤ 2

A 2-shattered set $\Rightarrow$ VC-dim $= 2$

Vapnik Chervonenkis (VC) dimension of an hypergraph

- A set X is shattered if $\forall Y \subseteq X$, $\exists e \in \mathcal{E}$, s.t $e \cap X = Y$.
- VC-dimension of $\mathcal{H}$: largest size of a shattered set.

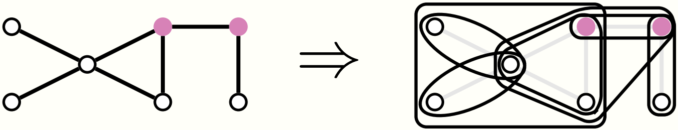


No 3-shattered set $\Rightarrow$ VC-dim ≤ 2

A 2-shattered set $\Rightarrow$ VC-dim $= 2$

VC dimension of a graph / of a class of graph

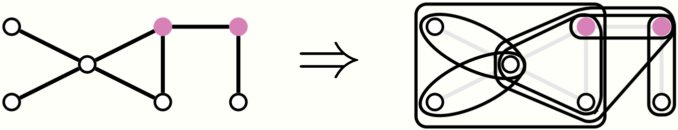
- VC-dimension of G: VC-dim of the hypergraph of closed neighborhoods



$$\text{VC-dim}(G) = 2$$

VC dimension of a graph / of a class of graph

- VC-dimension of G: VC-dim of the hypergraph of closed neighborhoods

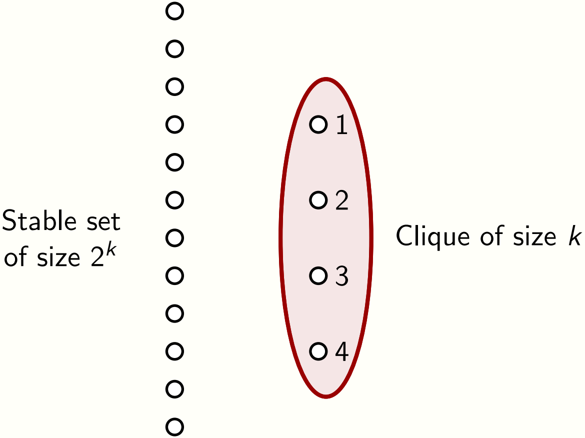


$$\text{VC-dim}(G) = 2$$

- VC-dimension of a class $\mathcal{C}$: maximal VC-dimension over $\mathcal{C}$
  - Class of interval graphs has VC-dimension 2.
  - Class of split graphs has infinite VC-dimension.

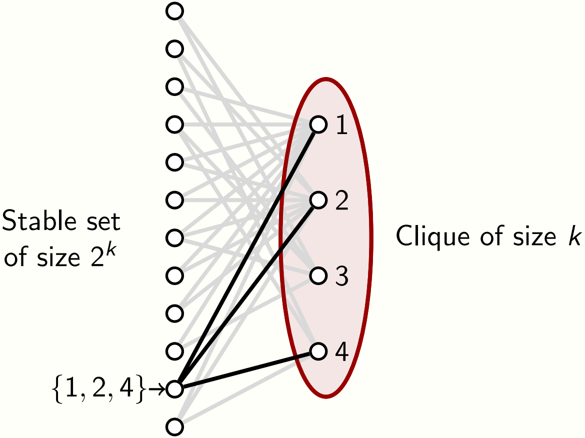
Split graphs have infinite VC-dimension

For any k, there is a split graph with VC-dimension k.



Split graphs have infinite VC-dimension

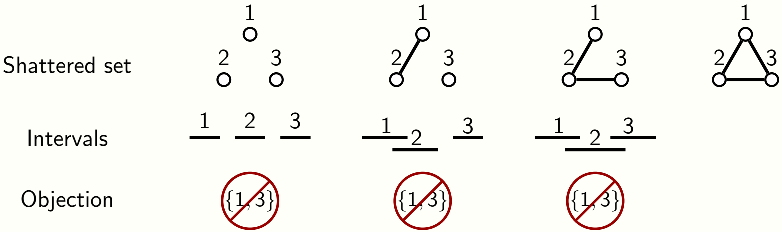
For any k, there is a split graph with VC-dimension k.



Intervals have finite VC-dimension

There is no interval graph with VC-dimension 3.

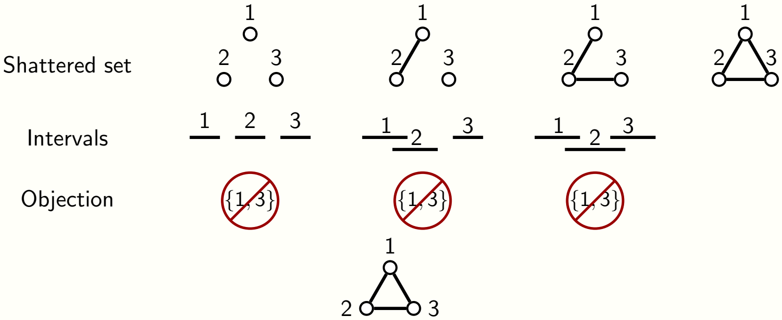
Assume there is a shattered set $\{1, 2, 3\}$.



Intervals have finite VC-dimension

> There is no interval graph with VC-dimension 3.

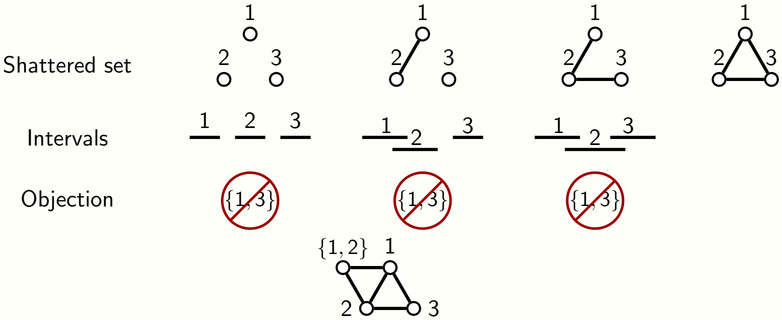
Assume there is a shattered set $\{1, 2, 3\}$.



Intervals have finite VC-dimension

There is no interval graph with VC-dimension 3.

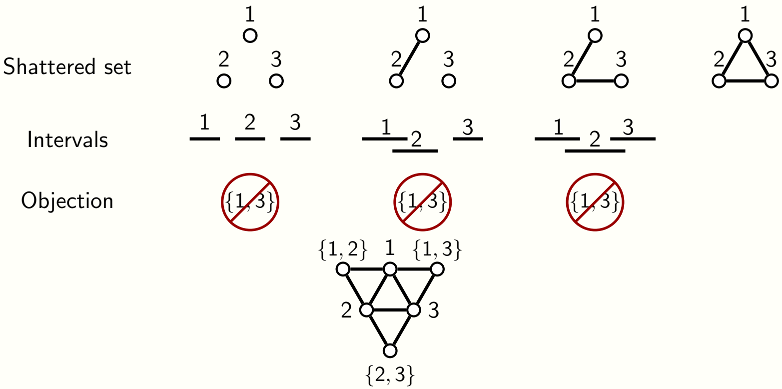
Assume there is a shattered set $\{1, 2, 3\}$.



Intervals have finite VC-dimension

There is no interval graph with VC-dimension 3.

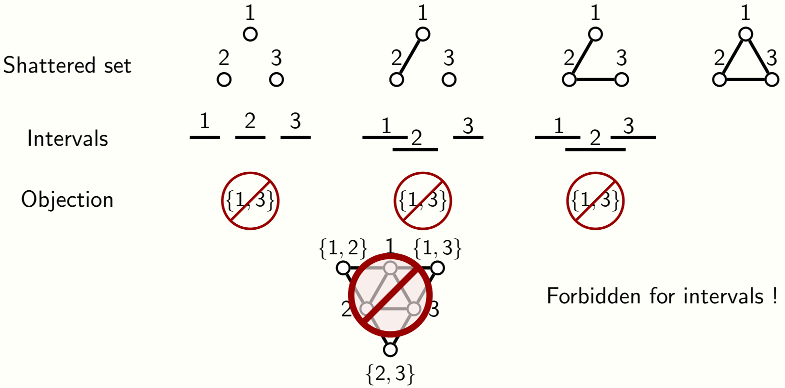
Assume there is a shattered set $\{1, 2, 3\}$.



Intervals have finite VC-dimension

There is no interval graph with VC-dimension 3.

Assume there is a shattered set $\{1, 2, 3\}$.



Interval graphs have VC-dimension at most 2.

Part III

Identifying codes and VC-dimension

Back to identifying codes

Graph class	Lower bound (order)	Approx
All	$\log n$	log APX-h
Split	$\log n$	log APX-h
Interval	$n^{1/2}$	open
Unit Interval	n	2
Bipartite	$\log n$	log APX-h
Line graphs	$n^{1/2}$	4
Chordal	$\log n$	log APX-h
Planar	n	7
Cograph	n	1

Back to identifying codes

Graph class	Lower bound (order)	Approx	VC dim
All	$\log n$	log APX-h	∞
Split	$\log n$	log APX-h	∞
Interval	$n^{1/2}$	open	2
Unit Interval	n	2	2
Bipartite	$\log n$	log APX-h	∞
Line graphs	$n^{1/2}$	4	4
Chordal	$\log n$	log APX-h	∞
Planar	n	7	4
Cograph	n	1	2

Back to identifying codes

Graph class	Lower bound (order)	Approx	VC dim
All	$\log n$	log APX-h	∞
Split	$\log n$	log APX-h	∞
Interval	$n^{1/2}$	open	2
Unit Interval	n	2	2
Bipartite	$\log n$	log APX-h	∞
Line graphs	$n^{1/2}$	4	4
Chordal	$\log n$	log APX-h	∞
Planar	n	7	4
Cograph	n	1	2

Back to identifying codes

Graph class	Lower bound (order)	Approx	VC dim
All	$\log n$	log APX-h	∞
Split	$\log n$	log APX-h	∞
Bipartite	$\log n$	log APX-h	∞
Chordal	$\log n$	log APX-h	∞
Interval	$n^{1/2}$	open	2
Unit Interval	n	2	2
Line graphs	$n^{1/2}$	4	4
Planar	n	7	4
Cograph	n	1	2

A dichotomy result

Theorem

VC-dim of $\mathcal{C}$?

Infinite	Finite d
$\Downarrow$	$\Downarrow$
There are infinetly many G,	For all G,
$\gamma^{ID}(G) \approx \log \lvert V \rvert$	$\gamma^{ID}(G) \geq c \lvert V \rvert^{1/d}$

Proof - Case with infinite VC dimension

Proposition

If $\mathcal{C}$ has infinite VC-dimension, for any integer k, $\mathcal{C}$ contains a graph G with at least 2^k vertices and an identifying code $\leq 2k$.

Proof - Case with infinite VC dimension

Proposition

If $\mathcal{C}$ has infinite VC-dimension, for any integer k, $\mathcal{C}$ contains a graph G with at least 2^k vertices and an identifying code $\leq 2k$.

Proof:

- Let $G \in \mathcal{C}$ of VC-dim k and X be a shattered set of size k.
- Let Y be a set shattering X.
- $C = X$ identifies all vertices of Y.
- Add to C at most k vertices of Y to identify vertices of X.

$\rightarrow$ $G' = G[X \cup Y]$ satisfies the claim.

Proof - Case with finite VC dimension

Proposition

If $\mathcal{C}$ has finite VC-dimension d, $\forall G \in \mathcal{C}$, $\gamma^{ID}(G) \geq c|V|^{1/d}$.

Proof - Case with finite VC dimension

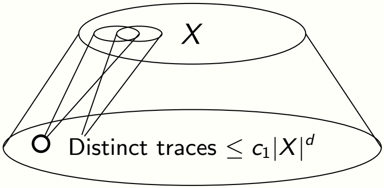
Proposition

If $\mathcal{C}$ has finite VC-dimension d, $\forall G \in \mathcal{C}$, $\gamma^{ID}(G) \geq c|V|^{1/d}$.

Proof: direct consequence of:

Sauer's Lemma

Let X be a subset of vertices of graph G of VC-dimension d. The number of distinct traces on X is at most $\sum_{i=1}^{d} \binom{|X|}{i} \leq c_1|X|^d$.



Proof - Case with finite VC dimension

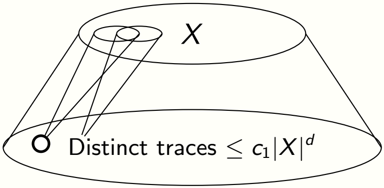
Proposition

If $\mathcal{C}$ has finite VC-dimension d, $\forall G \in \mathcal{C}$, $\gamma^{ID}(G) \geq c|V|^{1/d}$.

Proof: direct consequence of:

Sauer's Lemma

Let X be a subset of vertices of graph G of VC-dimension d. The number of distinct traces on X is at most $\sum_{i=1}^{d} \binom{|X|}{i} \leq c_1|X|^d$.



Identifying code $\gamma^{ID}(G)$

All vertices $|V|$ $\leq c_1\gamma^{ID}(G)^d$

Back to the table

Graph class	Lower bound	Approx	VC-dim
All	$\log n$	log APX-h	∞
Split	$\log n$	log APX-h	∞
Bipartite	$\log n$	log APX-h	∞
Chordal	$\log n$	log APX-h	∞
Interval	$n^{1/2}$	open	2
Unit Interval	n	2	2
Line graphs	$n^{1/2}$	4	4
Planar	n	7	4
Cograph	n	1	2
Permutation	$n^{1/3}$	open	3
Unit disk graphs	$n^{1/3}$	open	3

- Lower bound not optimal (ex: Line graphs)

Back to the table

Graph class	Lower bound ✓	Approx ?	VC-dim
All	$\log n$	log APX-h	∞
Split	$\log n$	log APX-h	∞
Bipartite	$\log n$	log APX-h	∞
Chordal	$\log n$	log APX-h	∞
Interval	$n^{1/2}$	open	2
Unit Interval	n	2	2
Line graphs	$n^{1/2}$	4	4
Planar	n	7	4
Cograph	n	1	2
Permutation	$n^{1/3}$	open	3
Unit disk graphs	$n^{1/3}$	open	3

- Lower bound not optimal (ex: Line graphs)
- What about approximation ?

Inapproximability in infinite VC dimension

Theorem

If $\mathcal{C}$ has ∞ VC-dimension, $\textsc{Min-Id-Code}$ is log-APX-hard on $\mathcal{C}$.

Inapproximability in infinite VC dimension

Theorem

If $\mathcal{C}$ has ∞ VC-dimension, $\textsc{Min-Id-Code}$ is log-APX-hard on $\mathcal{C}$.

Consequence of:

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains:

- all bipartite graphs, or
- all split graphs, or
- all cobipartite graphs.

and

Theorem Foucaud, 2013

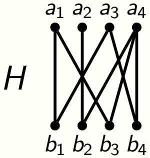
$\textsc{Min-Id-Code}$ is log-APX-hard on bipartite, split and cobipartite graphs.

Sketch of the proof

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.

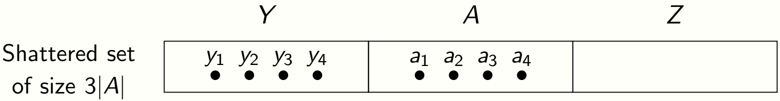


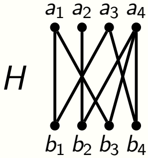
Sketch of the proof

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.





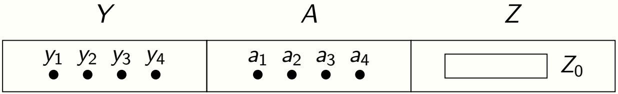
Sketch of the proof

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

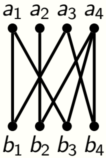
1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.

Shattered set of size $3|A|$



$\forall i, N[a_i] \cap Z \neq Z_0$

H

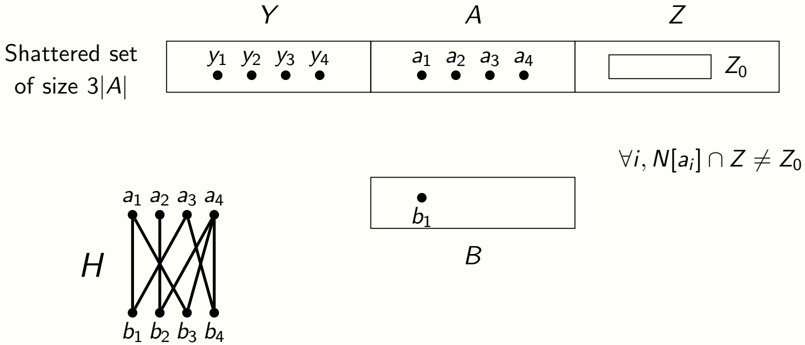


Sketch of the proof

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.

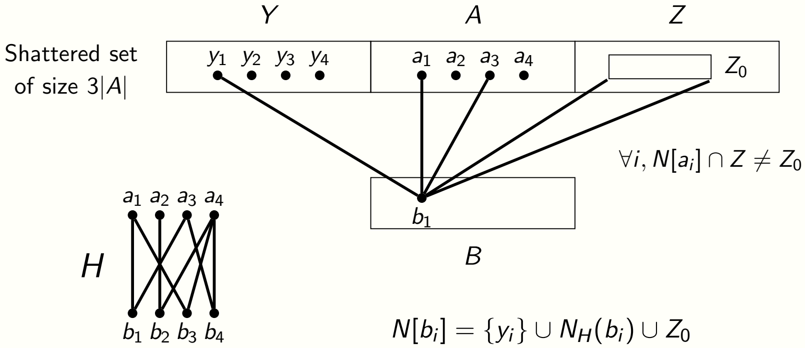


Sketch of the proof

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.



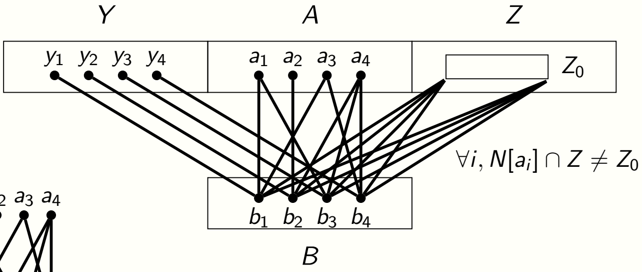
Sketch of the proof

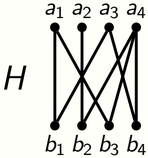
Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.

Shattered set of size $3|A|$





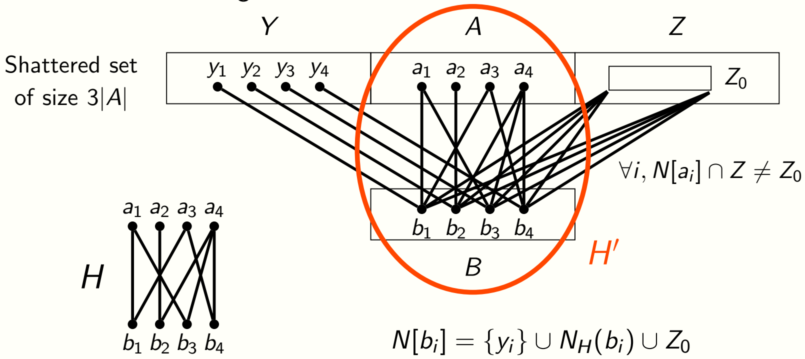
$$N[b_i] = \{y_i\} \cup N_H(b_i) \cup Z_0$$

Sketch of the proof

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.



Sketch of the proof

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

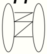
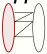
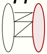
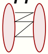
1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.
2. H' induces cliques or stable sets on A and B.

Sketch of the proof

Proposition

If $\mathcal{C}$ has infinite VC-dimension, $\mathcal{C}$ contains all bipartite graphs, or all split graphs, or all cobipartite graphs.

1. $\mathcal{C}$ is full-crossing bipartite:
   For any bipartite graph $H = (A \cup B, E)$, by adding edges on A or on B to H, we can get an element H' of $\mathcal{C}$.
2. H' induces cliques or stable sets on A and B.
3. Conclusion :

For any bipartite H, H^0  or H^1  or H^2  or H^3  $\in \mathcal{C}$

(H_n) : sequence of universal bipartite graphs.
$\Rightarrow \exists i \in \{0, 1, 2, 3\}, H_n^i \in \mathcal{C}$ for infinitely many n.
$\Rightarrow$ All bipartites ($i = 0$) or all splits ($i = 1, 2$) or all cobipartites ($i = 3$) are in $\mathcal{C}$.

In the finite case ?

Graph class	Lower bound	Approx	VC-dim
All	$\log n$	log APX-h	∞
Split	$\log n$	log APX-h	∞
Bipartite	$\log n$	log APX-h	∞
Chordal	$\log n$	log APX-h	∞
Interval	$n^{1/2}$	open	2
Unit Interval	n	2	2
Line graphs	$n^{1/2}$	4	4
Planar	n	7	4
Cograph	n	1	2
Permutation	$n^{1/3}$	open	3
Unit disk graphs	$n^{1/3}$	open	3

Is there a constant approximation in finite VC-dimension?

A class of finite VC-dimension with no good approximation

Theorem

MIN-ID-CODE cannot be approximated within a $o(\log |V|)$ factor in polynomial time for the class of bipartite C_4-free graphs.

- Class of VC-dimension 2
- Reduction from SET COVERING WITH INTERSECTION 1.

A class of finite VC-dimension with no good approximation

Theorem

MIN-ID-CODE cannot be approximated within a $o(\log |V|)$ factor in polynomial time for the class of bipartite C_4-free graphs.

- Class of VC-dimension 2
- Reduction from SET COVERING WITH INTERSECTION 1.

Gracias !