# Local Combinatorial Analogues for Bounded VC Dimension

**Olga Medrano Martín del Campo**
University of Chicago, IL, USA

## Abstract

Stable graphs, or equivalently Littlestone classes, were characterized by existence of linear-sized "good" sets, a kind of strongly homogeneous set, in work of Malliaris–Shelah and Malliaris–Moran. We prove a parallel result for VC classes, showing these are characterized by existence of linear-sized symmetric or asymmetric good pairs (which we define). We give several proofs, each drawing from methods and results from different areas, and resulting in different kinds of bounds. We finish with a few words on our learning theory motivation for these investigations and state some further research directions.

**2012 ACM Subject Classification** Theory of computation → Complexity theory and logic; Theory of computation → Finite Model Theory

**Keywords and phrases** Littlestone, Vapnik–Chervonenkis, good sets, Regularity Lemma, Haussler Packing Lemma, Fundamental Theorem of Statistical Learning

**Digital Object Identifier** 10.4230/LIPIcs.LICS.2026.72

**Funding** *Olga Medrano Martín del Campo*: Research partially supported by Fundación Alberto y Dolores de Andrade, I.A.P.

**Acknowledgements** Thanks to Maryanthe Malliaris for the invaluable mentorship as this paper was prepared; to her and to Leonardo Nagami Coregliano for their patient remarks and feedback; to Balder ten Cate, Steffen van Bergerem, Phokion Kolaitis, Caroline Terry, and Gabe Conant for insightful conversations; and to the anonymous referees for their thoughtful and detailed feedback.

## 1 Introduction

The context of studying classes $(X, \mathcal{H})$ as bigraphs has allowed for connections and applications among areas including learning theory, model theory, and extremal graph theory.

Important work has been done for the class of graphs that have bounded VC dimension, which are characterized by *PAC learnability* due to the Fundamental Theorem of Statistical Learning [26]. In extremal combinatorics, VC classes have a regularity property stronger than the one for the class of all graphs (see [13, 2, 18]).

Bigraphs with the property of *$k$-edge stability* correspond to Littlestone classes in learning theory [6], which are moreover *online learnable* [3] and *differentially private PAC learnable* [1]. In extremal combinatorics [20], the class of $k$-edge stable graphs has been proven to satisfy a stronger regularity condition that the class of all graphs. The work in [19] has found an application of the existing tools and techniques from learning theory into extremal combinatorics, proving that Littlestone classes can be shown to have $\epsilon$-excellent sets of large size, not only for $\epsilon < 1/2^{\ell}$, where $\ell$ is the Littlestone dimension of the class, but for all $\epsilon \in (0, 1/2)$, by using regret bound methods from online learning.

In the present work we are interested in the question of to what extent the strong results on good sets and Littlestone classes can be extended to the larger set of structures which are VC classes. Since the existence of linear-sized good sets characterized stability, there will need to be a distinct definition for VC classes. We propose and study a partite analogue and show that the existence of linear-sized symmetric good pairs, in the sense of Definition 13, characterizes VC classes.

We give three different proofs of the above. The first one is a combinatorial proof by hand. The second one is a proof appealing to Haussler's Packing Lemma [16]. The third one is a proof based on an application of the VC Regularity Lemma [13, 2, 18]. For our hope for applications in learning theory, there are different advantages to these proofs because the algorithmic and computational costs of the respective combinatorial tools used can be significant.

## Prior and Related Work

The work done used the following existing results: the VC Regularity Lemma [13, 2, 18]; Haussler's Packing Lemma [16]; the Stable Regularity Lemma [20]; the analysis of good and excellent sets in Littlestone classes [19].

The search for better Ramsey type results in VC classes has a long history of work. In particular, [23] recently established the Erdős-Hajnal property for VC classes. However, other work on the linear scale does not necessarily relate to this.

Let us give a few remarks on our motivation for considering analogues of good sets, which themselves can be considered as analogues of indiscernible sequences (Section 2 in [21]). Let us also point to the work from [4], [9], and [8].

The reader may ask why we didn't extend this analogue for Ramsey-type results (see [10]). Both notions are needed and different, as can be seen in the work of Malliaris and Shelah for the stable graph case [20]. In Section 3 of this paper, a regularity lemma is shown in which pieces are indiscernible to some extent, and thus need to be much smaller in size. Despite all the strong properties of this partition, the number of pieces grows with the number of vertices in the graph. By weakening the notion to that of so-called good and excellent sets, in Section 5, they recover a partition with linear-sized parts, which results in a partition of uniformly bounded size, as is the case in Szemerédi's Regularity Lemma [25]. The work here lays groundwork for future investigations for similar tradeoffs in the VC case, following this line of thought.

# 2 Paper Overview

Section 3 defines VC and Littlestone dimension, and defines good and antigood sets, provided a parameter $\epsilon \in (0, 1/2)$. Section 4 defines the main objects we study in their symmetric and asymmetric versions, and states our main result, Theorem 15.

We will prove this theorem, or analogues of it, in several ways. Section 4.1 gives a combinatorial proof, by hand. Section 9 compares Theorem 15 to the existing characterization for Littlestone classes. Section 7 expands on a key example, the half graph. Then, Section 8 gives an alternative proof of Theorem 15 using Haussler's Packing Lemma, and Section 9 gives another alternative proof, by using the VC (also known as ultra strong) Regularity Lemma. Section 10 ends with some open questions.

# 3 Background

## 3.1 Notation

The following are some objects that we will work with in this paper:

▶ **Definition 1.** *A* hypothesis class $(X, \mathcal{H})$*, also denoted* $\mathcal{H}$*, consists of a set of subsets of the set* $X$*. We may several times identify each* $h \in \mathcal{H}$ *with its characteristic function, which for each* $a \in h$ *gives* $h(a) = 1$*, and for each* $a \notin h$ *gives* $h(a) = 0$*.*

▶ **Definition 2.** *The* bipartite graph $G = (V_1 \cup V_2, E)$ *is given with a partition of the vertex set, and with an edge set* $E \subseteq V_1 \times V_2$ *which is neither symmetric nor irreflexive.*

Eventhough the following definition is not used in the proofs in this paper, it is needed to state Theorem 6.

▶ **Definition 3.** *Given a hypothesis class* $\mathcal{H}$*, we can regard it as a bipartite graph, or bigraph, denoted* $\mathrm{Bip}(X, \mathcal{H})$*. Here:*

1. $X$ *and* $\mathcal{H}$ *are the two sides of the vertex set partition, and an edge relation is given by* $E := \{(a, h) \ : \ h(a) = 1\}$*; and*
2. $\mathrm{Bip}(X', \mathcal{H}')$ *can be defined by letting* $X' = \mathcal{H}$*,* $\mathcal{H}' = X$*, and* $E := \{(h, a) \ : \ h(a) = 1\}$
3. *by letting* $\mathcal{H}' = \{\{h \in \mathcal{H} : \ h(x) = 1\} : \ x \in X\}$*, we obtain the dual class* $(\mathcal{H}', X')$*.*

## 3.2 Littlestone and VC dimensions

The following combinatorial structure was introduced by Littlestone [17].

▶ **Definition 4** (Littlestone Mistake tree)**.** *A* Littlestone mistake tree *is a binary decision tree whose nodes are* $x_i \in X$*, and whose branch edges are labeled respectively by* $y_i \in \mathcal{Y} = \{\pm 1\}$*, each node having its two outgoing edges with different labels. A hypothesis* $h \in \mathcal{H}$ realizes *a root-to-leaf path* $\{(x_i, y_i)\}_{i=1}^d$ *when* $h(x_i) = y_i$ *for every branch, and* $\mathcal{H}$ shatters *a tree when every root-to-leaf path is realized by some* $h \in \mathcal{H}$*.*

▶ **Definition 5** (Littlestone dimension)**.** *The maximum depth of a Littlestone mistake tree that is shattered by* $\mathcal{H}$ *is defined as the* Littlestone dimension *of the hypothesis class, and denoted by* $\mathrm{Ldim}(\mathcal{H})$*.*

Another complexity measure we work with is VC dimension:

▶ **Definition 6** (Vapnik–Chervonenkis (VC) dimension)**.** *We say that a class* $\mathcal{H}$ shatters *a set* $\{x_i\}_{i=1}^n \subseteq X$ *if for every possible labeling* $\{y_i\} \in \{0, 1\}^n$ *of the elements of* $X$*, there exists some* $h \in \mathcal{H}$ *such that* $h(x_i) = y_i$ *for all* $i \in [n]$*. We then define the* VC dimension*, denoted by* $\mathrm{VC}(\mathcal{H})$*, to be the supremum of the sizes of all finite sets in* $X$ *shattered by* $\mathcal{H}$ *this way.*

By using mistake trees, we can give an equivalent definition for VC dimension:

▶ **Definition 7** (VC mistake tree)**.** *A* VC mistake tree *is a binary decision tree with nodes* $x_i \in X$ *and branching edges labeled by* $y_i \in \{\pm 1\}$*, such that each node has two outgoing branches with different labels, and such that for every level* $0 \leq \ell \leq d$ *of the tree, the node labels of all elements of* $X$ *in level* $\ell$ *are the same. A hypothesis* $h \in \mathcal{H}$ realizes *a root-to-leaf path* $\{(x_i, y_i)\}_{i=1}^d$ *when* $h(x_i) = y_i$ *for every* $1 \leq i \leq d$*, and a class* $\mathcal{H}$ shatters *a tree when every root-to-leaf path is realized by some* $h \in \mathcal{H}$*.*

▶ **Definition 8** (Vapnik–Chervonenkis dimension)**.** *The maximum depth of a VC mistake tree with nodes from* $X$ *that is shattered by* $\mathcal{H}$ *is defined as the VC dimension of the hypothesis class,* $\mathrm{VC}(\mathcal{H})$*.*

## 3.3 Good sets

The notion of good and excellent sets from Malliaris and Shelah's work [20] is key here. In this paper, we will only focus on the notion of $\epsilon$-good sets.

▶ **Definition 9** (Good sets). *Let $G = (V, E)$ be a graph, and let $\epsilon \in (0, 1/2)$. An $\epsilon-$*good set *$S \subseteq V$ is one such that for every vertex $v \in V$, one of the following holds:*

**(i)** $|N(v, S)| < \epsilon|S|$*; in this case, we denote $\mathbf{t}(v, S) = 0$, and $v$* has a vote of 0 on *$S$;*

**(ii)** $|N(v, S)| > (1 - \epsilon)|S|$*; in this case, we denote $\mathbf{t}(v, S) = 1$, and $v$* has a vote of 1 on *$S$.*

Note that the property of being $\epsilon$-good is monotonic with respect to $\epsilon$: an $\epsilon_1-$good set is also $\epsilon_2-$good, if $\epsilon_1 < \epsilon_2$. Indeed, for each $v$, $|N(v, S)| < \epsilon_1|S|$ implies $|N(v, S)| < \epsilon_2|S|$, and $|N(v, S)| > (1 - \epsilon_1)|S|$ implies $|N(v, S)| > (1 - \epsilon_2)|S|$. The implication from $\epsilon_1$ to $\epsilon_2$ remains true if we quantify over all $v \in V$.

### 3.4 Anti-good sets

Given a set $S$ and a parameter $\epsilon$, we introduce the notion of its antigood set $\mathrm{Anti}(S, \epsilon)$, which denotes the set of vertices that prevent $S \subseteq V$ from being $\epsilon$-good.

▶ **Definition 10.** *Let $S \subseteq V$, and $\epsilon \in (0, 1/2)$. Its* antigood set*, $\mathrm{Anti}(S, \epsilon)$, is defined as follows*

$$\mathrm{Anti}(S, \epsilon) := \{v \in V : \epsilon|S| \leq |N(v, S)| \leq (1 - \epsilon)|S|\}$$

Note, that if a set $S \subseteq V$ is $\epsilon$-good, then $\mathrm{Anti}(S, \epsilon) = \varnothing$. Allowing for the antigood set to be nonempty, but requiring it to be small, we can still get relevant information, as will be elaborated below.

The following definition will be helpful later on. It distinguishes between the elements in $\mathrm{Anti}(S, \epsilon)$ and $V \setminus \mathrm{Anti}(S, \epsilon)$ in terms of how they partition the set $S$.

▶ **Definition 11.** *Let $\epsilon \in (0, 1/2)$ be a parameter, $S \subseteq V$ be a vertex set, and $v \in V$. We say that $v$ partitions $S$*

**(a)** *in an $\epsilon$-*balanced way*, if $|N(v, S)|/|S| \in [\epsilon, 1 - \epsilon]$; and*

**(b)** *in an $\epsilon$-*unbalanced way*, if $|N(v, S)|/|S| \in [0, \epsilon) \cup (1 - \epsilon, 1]$.*

## 4 Contributions of this work

### 4.1 Characterization of classes of bounded $VC$ dimension

We found a property equivalent to being of bounded VC dimension, and called this property $(c_1, c_2, \epsilon)$-good sampling. Before defining it, we state two preliminary notions of having a pair of sets be $\epsilon$-good. This could happen symmetrically or not.

▶ **Definition 12.** *[Asymmetric $\epsilon$-goodness] Let $G = (V, E)$ be a graph, $A, B \subset V$ be subsets of nodes in this graph, and $\epsilon \in (0, 1/2)$. We say that $A$ is $\epsilon$-*good for *$B$ if for every $b \in B$, the following holds:*

$$\frac{|\{a \in A : (b, a) \in E\}|}{|A|} \in [0, \epsilon) \cup (1 - \epsilon, 1].$$

▶ **Definition 13.** *[Symmetric $\epsilon$-goodness] Let $G = (V, E)$ be a graph, $A, B \subset V$ be subsets of nodes in this graph, and $\epsilon \in (0, 1/2)$. We say that $A$ and $B$ form a* symmetric $\epsilon$-good pair of sets *if $A$ is $\epsilon$-good for $B$ and $B$ is $\epsilon$-good for $A$.*

These definitions are different a priori. The relevance of this difference will become clearer in this section and Section 4.1.

Note that Definitions 12 and 13 differ: there exist pairs $A$ and $B$ of sets which are $\epsilon$-good, only in the asymmetric sense. Consider, for example, the bipartite graph with $V = A \cup B$, where half of $B$ is connected to all of $A$, whereas the other half is not connected to all of $A$, as Figure 1 below shows. Then, $A$ is $\epsilon$-good for $B$, but $B$ is not $\epsilon$-good for $A$, for any choice of parameter $\epsilon \in (0, 1/2)$.

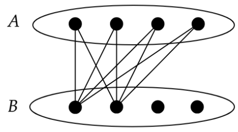


**Figure 1** An example of a pair that is $\epsilon$-good only in the asymmetric sense.

Motivated by the above, we give the following definition:

▶ **Definition 14.** *We say that a pair $(X, \mathcal{H})$ has $(c_1, c_2, \epsilon)$-good sampling if, for every $A \subseteq X$ and $B \subseteq \mathcal{H}$, there exist $A' \subseteq A$ and $B' \subseteq B$ such that $(A', B')$ form a symmetric $\epsilon$-good pair, and*

$$|A'| \geq c_1|A| \qquad \textit{and} \qquad |B'| \geq c_2|B|.$$

The following is the main contribution of this work, and states that the property of $(c_1, c_2, \epsilon)$-good sampling is characteristic in the classes that have bounded VC dimension:

▶ **Theorem 15.** *Let $\mathcal{H} \subseteq \{0, 1\}^X$ be a hypothesis class. The following are equivalent:*

**(a)** $\mathrm{VC}(\mathcal{H}) = d < \infty$*;*

**(b)** *for every $\epsilon \in (0, 1/2)$ there exist constants $c_1, c_2$ such that $(X, \mathcal{H})$ has $(c_1, c_2, \epsilon)$-good sampling;*

**(c)** *there exists $\epsilon_0 \in (0, 1/2)$ and constants $c_1, c_2$ such that $(X, \mathcal{H})$ has $(c_1, c_2, \epsilon_0)$-good sampling.*

As a remark, it will follow from the proof in Section 5 for Theorem 15 that there is a basic lower bound for each of the constants: $c_1 \geq \epsilon/2^{d+1}$ and $c_2 \geq 1/2^{d+1}$.

An example where we have the property of $(c_1, c_2, \epsilon)$-good sampling is the half graph, whose VC dimension is 1. This example is discussed more in depth in Section 7.

## 4.2 Existence of sets with small antigood sets

In the previous subsection, Theorem 15 shows that classes of bounded VC dimension have the property of $(c_1, c_2, \epsilon)$-good sampling, which in words means there exist linear-sized symmetric $\epsilon$-good pairs. This subsection studies a reinterpretation of this characterization, using the language of antigood sets introduced earlier.

Indeed, an immediate corollary of Theorem 15 states that given a VC class and a parameter $\epsilon > 0$, it is possible to find a set with lower bounded size, with an antigood set whose size is upper bounded:

▶ **Corollary 16.** *Let $\mathcal{H}$ be a hypothesis class with VC dimension $d$. Then, there exists a value $\epsilon_0 \in (0, 1/2)$ such that $(X, \mathcal{H})$ has $(c_1, c_2, \epsilon_0)$-good sampling.*
*That is, for every $A \subseteq X$ and $B \subseteq \mathcal{H}$, there exist subsets $A' \subseteq A, B' \subseteq B$ of sizes at least $c_1|A|$ and $c_2|B|$ respectively, such that $(A', B')$ is a symmetric $\epsilon_0$-good pair, or equivalently:*

$$|\mathrm{Anti}(A', \epsilon_0)| = |B| - |B'| \leq |B| \cdot (1 - c_2).$$

## 5 Proof of Theorem 15

First we will show that bounded VC dimension implies the existence of asymmetric $\epsilon$-good pairs:

▶ **Lemma 17.** *If* $\mathrm{VC}(\mathcal{H}) \leq d < \infty$ *then for every* $\epsilon \in (0, 1/2)$ *there exist constants* $c_1 := \epsilon^d$ *and* $c_2 := \frac{1}{2^d}$ *such that for every two finite sets* $A \subseteq \mathcal{H}$ *and* $B \subseteq X$*, there exist* $A' \subseteq A$ *and* $B' \subseteq B$ *with* $|A'| \geq c_1|A|$ *and* $|B'| \geq c_2|B|$*, such that* $A'$ *is* $\epsilon$*-good to* $B'$*.*

**Proof.** We can construct a VC mistake tree in the following way. We begin with $A_\varnothing := A$ at level 0. If there is some $b_0 \in B$ that partitions $A_\varnothing$ into $A_0 \cup A_1$ in an $\epsilon$-balanced way, then we place $A_0$ and $A_1$ at level 1.

Consider level $\ell$, with the nodes in it being a partition of $A_\varnothing$ into $\bigcup_{i \in 2^\ell} A_i$, with $\ell \in \mathbb{N}_{\geq 0}$. Constructing level $\ell + 1$ would require finding $b_\ell \in B$ such that $b_\ell$ partitions every $A_i$ in an $\epsilon$-balanced way, into $A_{i^\frown 0} := \{a \in A_i : a(b_\ell) = 0\}$ and $A_{i^\frown 1} := \{a \in A_i : a(b_\ell) = 1\}$. If all the sets $A_i$ for $i \in 2^\ell$ are nonempty, the VC tree can be obtained by extracting one $a_i \in A_i$ for $i \in \bigcup_{0 \leq m \leq \ell} 2^m$. This collection of hypotheses that shatter the set $\{b_k : 0 \leq k \leq \ell - 1\}$ would witness that $\mathrm{VC}(\mathcal{H}) \geq \ell$, a contradiction.

The upper bound $\mathrm{VC}(\mathcal{H}) \leq d$ implies that this VC tree construction can go only as far as $d$ levels or less. Thus, for some $\ell_0 \leq d$, no $b \in B$ partitions simultaneously all parts in that level in a balanced way. Then, for every $b \in B$, there exists at least one $A_i$ with $i \in 2^{\ell_0}$ that is partitioned in an $\epsilon$-unbalanced way by $b$.
By Pigeonhole Principle, at least a fraction of $1/2^{\ell_0} \geq 1/2^d = c_1$ of the elements $b$ of $B$ partition the same part $A_{i_0}$ in an $\epsilon$-unbalanced way; define such set $B'$, and define $A_{i_0} = A'$. We have that $|B'| \geq c_2|B|$, with $c_2 = 1/2^d$, and that $|A'| \geq \epsilon^{\ell_0}|A| \geq \epsilon^d|A| = c_2|A|$. Moreover, $A'$ is $\epsilon$-good to $B'$, since this set is partitioned in an $\epsilon$-unbalanced way by each $b \in B'$:

$$\forall b \in B', \qquad \frac{|N(b) \cap A'|}{|A'|} \in [0, \epsilon) \cup (1 - \epsilon, 1]. \qquad ◀$$

With the lemma above, we could find an asymmetric $\epsilon$-good pair, which is not symmetric a priori. Applying the following lemma will allow us to extract a symmetric $\epsilon$-good pair from it.

▶ **Lemma 18.** *Let* $0 < \epsilon < 1/4$ *and let* $A, B \subset V$ *be such that* $A$ *is* $\epsilon$*-good to* $B$*. Then, there exist* $A' \subseteq A$ *and* $B' \subseteq B$ *such that* $(A', B')$ *is a symmetric* $2\epsilon$*-good pair and* $|A'| \geq |A|/2$*, and* $|B'| \geq |B|/2$*.*

**Proof.** Since $A$ is $\epsilon$-good to $B$, we can write $B = B_0 \cup B_1$, where $B_i := \{b \in B : \mathbf{t}(b, A) = i\}$. We have two cases, depending on which of the two subsets is larger.

If $|B_0| \geq |B_1|$, set $B' := B_0$. Each vertex in $B'$ has less than $\epsilon|A|$ neighbors in $A$, and edges are symmetric, thus

$$e(A, B') = e(B', A) < |B'| \cdot \epsilon|A|.$$

Let $A = A^- \cup A^+$, where $A^+ := \{a \in A : |N(a) \cap B'| \geq 2\epsilon|B'|\}$ and $A^- := A \setminus A^+$. The left-hand side is bounded:

$$e(A, B') = e(A^+, B') + e(A^-, B') \geq |A^+| \cdot 2\epsilon|B'| \ + \ |A^-| \cdot 0$$

And thus, $\epsilon|A| \cdot |B'| > 2\epsilon|A^+| \cdot |B'|$, which finally implies $\frac{1}{2}|A| > |A^+|$, and that $|A^-| \geq \frac{1}{2}|A|$. If we define $A' := A^-$, then $(A', B') = (A^-, B_0)$ is a symmetric $2\epsilon$-good pair. Indeed, every $a \in A^-$ has at most $2\epsilon|A'|$ neighbors in $B_0$, and moreover, every $b \in B_0$ satisfies the following:

$$|N(b) \cap A^-| = |N(b) \cap A| - |N(b) \cap A^+| < \epsilon|A| \leq 2\epsilon|A^-|.$$

If $|B_1| \geq |B_0|$, we set $B' := B_1$. Every vertex in $B'$ has more than $(1-\epsilon)|A|$ neighbors in $A$. Since edges are symmetric:

$$e(A, B') = e(B', A) > |B'| \cdot (1-\epsilon)|A|.$$

If we now let $A = A^- \cup A^+$, where $A^- := \{a \in A : |N(a) \cap B'| \leq (1-2\epsilon)|B'|\}$ and $A^+ := A \setminus A^-$, we then get the bound:

$$e(A, B') = e(A^-, B') + e(A^+, B') \leq |A^-| \cdot (1-2\epsilon)|B'| + |A^+| \cdot |B'|$$

Joining and simplifying the above two inequalities:

$$(1-\epsilon)|A| \cdot |B'| \;<\; e(A, B') \;<\; |A^-| \cdot (1-2\epsilon)|B'| + |A^+| \cdot |B'|$$

$$\Rightarrow \quad (1-\epsilon)|A| < (1-2\epsilon)|A^-| + |A^+|,$$

implying that $|A^+| \geq |A|/2 > |A^-|$. Then, the pair $(A', B') = (A^+, B_1)$ is symmetric $2\epsilon$-good: every $a \in A^+$ has at least $(1-2\epsilon)|B'|$ neighbors in $B'$, and every $b \in B'$ satisfies

$$|N(b)^c \cap A^+| = |N(b)^c \cap A| - |N(b)^c \cap A^-| \;<\; \epsilon|A| \leq 2\epsilon|A^+|$$

$$\Rightarrow |N(b) \cap A^+| \;=\; |A^+| - |N(b)^c \cap A^+| \;\geq\; (1-2\epsilon)|A^+|.$$

In either case, we obtained two subsets $A' \subseteq A$ and $B' \subseteq B$ of size at least half that of the original sets in the pair, that are symmetric $2\epsilon$-good. Note, $2\epsilon \in (0, 1/2)$ as long as $\epsilon \in (0, 1/4)$. ◀

Now we have all that we need in order to extract symmetric good pairs in the case that our hypothesis class is of finite VC dimension:

▶ **Corollary 19.** *If* $\mathrm{VC}(\mathcal{H}) \leq d < \infty$ *then for every* $0 < \epsilon < 1/2$ *there exist constants* $c_1 := \epsilon^d/2^{d+1}$, *and* $c_2 := 1/2^{d+1}$ *such that for every two finite sets* $A \subseteq \mathcal{H}$ *and* $B \subseteq X$, *there exist* $A' \subseteq A$ *and* $B' \subseteq B$ *with* $|A'| \geq c_1|A|$ *and* $|B'| \geq c_2|B|$, *such that* $(A', B')$ *is a symmetric* $\epsilon$*-good pair.*

**Proof.** By Lemma 17, but with parameter $\epsilon/2 > 0$, we obtain sets $A'', B''$ such that $A''$ is $(\epsilon/2)$-good to $B''$, with sizes at least

$$|A''| \geq \left(\frac{\epsilon}{2}\right)^d |A|, \qquad |B''| \geq \frac{1}{2^d}|B|.$$

Then, by Lemma 18 there exist sets $A' \subseteq A''$ and $B' \subseteq B''$, such that $(A', B')$ is a symmetric $\epsilon$-good pair (*since we have* $2 \cdot (\epsilon/2) = \epsilon$), and such that

$$|A'| \geq \frac{1}{2}|A''| \geq \frac{\epsilon^d}{2^{d+1}}|A|, \qquad |B'| \geq \frac{1}{2}|B''| \geq \frac{1}{2^{d+1}}|B|.$$ ◀

For further discussion on the comparative performance of the bounds obtained in Corollary 19 and those in further sections, please see Appendix C.

The next part of our characterization of bounded VC dimension will be to show that large VC dimension implies that there is no $(c_1, c_2, \epsilon)$-good sampling. We will first show that this happens in the case of infinite VC dimension.

▶ **Lemma 20.** *If* $\mathrm{VC}(\mathcal{H}) = \infty$, *then for any* $\epsilon \in (0, 1/2)$, *and for every constant values* $c_1, c_2 > 0$ *there exist two sets* $A \subseteq X$ *and* $B \subseteq \mathcal{H}$ *such that any* $A' \subseteq A$ *and* $B' \subseteq B$ *of sizes at least:*

$$|A'| \geq c_1|A| \quad \textit{and} \quad |B'| \geq c_2|B|,$$

*will not form an* $\epsilon$*-good pair. That is, a large subset* $A'$ *cannot be* $\epsilon$*-good to a large subset* $B'$*.*

**Proof.** Let $d \in \mathbb{N}$. Since $\mathrm{VC}(\mathcal{H}) \geq d$, there exist sets $A \subseteq X$ and $B \subseteq \mathcal{H}$, of sizes $d$ and $2^d$, respectively, such that every subset of $A$ is realized as a neighborhood of an element in $B$. We want to show the property stated above, namely that no sufficiently large subsets $A'$ and $B'$ can form an $\epsilon$-good pair.

For the sake of contradiction, suppose there exists such pair $(A', B')$. Then, for any subset of $A'$ of size $k' \geq c_1|A|$, every element of $B'$ must connect to any number of elements on $A'$ that lies in the range $[0, \epsilon k') \cup (k'(1-\epsilon), k']$. Thus,

$$|B'| \leq \sum_{s=0}^{\lfloor \epsilon k' \rfloor} \binom{k'}{s} \cdot 2^{d-k'} + \sum_{s=\lceil k'(1-\epsilon)\rceil}^{k'} \binom{k'}{s} \cdot 2^{d-k'} \leq \sum_{s=0}^{\epsilon k'} 2 \cdot \binom{k'}{s} \cdot 2^{d-k'}$$

We now argue that $|B'|/|B| = o(1)$ as $k \to \infty$:

$$\frac{|B'|}{|B|} \leq \sum_{s=0}^{\lfloor \epsilon k' \rfloor} 2 \cdot \binom{k'}{s} \cdot 2^{-k'} = 2^{1-k'} \cdot \sum_{s=0}^{\lfloor \epsilon k' \rfloor} \binom{k'}{s}$$

Since $\epsilon \in (0, 1/2)$, the above partial sum of binomial coefficients is upper bounded by

$$2^{H\left(\epsilon k'/k'\right)\cdot k'} = 2^{H(\epsilon)\cdot k'},$$

where $H$ is the binary entropy function (see e.g., [11]), defined as the entropy of a Bernoulli random trial with probabilities $\{p, 1-p\}$:

$$H(p) := \begin{cases} 0 & p = 0, 1 \\ -p\log_2(p) - (1-p)\log_2(1-p) & p \in (0,1) \end{cases}$$

We then have

$$\frac{|B'|}{|B|} \leq 2^{1-k'+k'\cdot H(\epsilon)} \tag{1}$$

We have that $k' \to \infty$ as $d \to \infty$, since $k' \geq c_1 \cdot d$. Moreover, $H(\epsilon) < 1$ because $\epsilon < 1/2$. Thus, the exponent in Equation 1 goes to $-\infty$ as $k' \to \infty$, and so $|B'|/|B|$ goes to zero. This would contradict that $|B'| \geq c_2|B|$ for any value $c_2$ that is constant with respect to $d$. ◀

The following presents a finite version of the above conclusion.

▶ **Lemma 21.** *For every $\epsilon \in (0, 1/2)$, for all constants $c_1, c_2 > 0$, and for every sufficiently large $d \in \mathbb{N}$, there exists a class $\mathcal{H}$ such that $\mathrm{VC}(\mathcal{H}) = d$, with the property that any two sets $A \subseteq X$ and $B \subseteq \mathcal{H}$ and any two subsets $A' \subseteq A$ and $B' \subseteq B$ with*

$$|A'| \geq c_1|A| \qquad \text{and} \qquad |B'| \geq c_2|B|$$

*cannot form an $\epsilon$-good pair $(A', B')$, not even in the asymmetric sense.*

**Proof.** Having $\mathrm{VC}(\mathcal{H}) = d$ gives two sets $A \subseteq X$ and $B \subseteq \mathcal{H}$ of sizes $d$ and $2^d$ respectively, $B$ realizing at least once every subset of $A$. In analogy to the work done above, a counting argument gives that the following bound must hold if the pair $(A', B')$ is (asymmetric) $\epsilon$-good:

$$c_2 \leq \frac{|B'|}{|B|} \leq 2^{1+k'\cdot(H(\epsilon)-1)}$$

Since $H(\epsilon) < 1$ and $k' \geq c_2 \cdot |B| = c_2 \cdot 2^d$, then for sufficiently large $d$ we have that the bound doesn't hold, or equivalently, the following is not true for all $d \in \mathbb{N}$.

$$\log_2(c_2/2) \leq k' \cdot (H(\epsilon) - 1) \leq d \cdot (H(\epsilon) - 1)$$

$$\Leftrightarrow \frac{\log(2/c_2)}{1 - H(\epsilon)} \geq d. \qquad \blacktriangleleft$$

We can now complete the proof of Theorem 15, as follows:

**Proof of Theorem 15.** The implication $(a) \Rightarrow (b)$ is given in the statement of Lemma 17. The implication $(b) \Rightarrow (c)$ only consists of the specification of the parameter $\epsilon_0$, and finding the constants $c_1, c_2$ corresponding to that $\epsilon_0$ that give the property of $(c_1, c_2, \epsilon_0)$-good sampling. Finally, the implication $(c) \Rightarrow (a)$ is precisely the contrapositive of Lemma 21. ◀

## 6 Comparison with Littlestone characterization

Let us restate in the language of this paper an existing characterization of Littlestone classes, from Claim 2.11 in the work of [19]. First we define an important object that appears both in this characterization and in Section 7.

▶ **Definition 22.** *The* $k$-half graph $H_k$ *is given by two disjoint sets of vertices* $\{a_1, \dots, a_k\}$, $\{b_1, \dots, b_k\}$ *so that*

$$(a_i, b_j) \in E \text{ if } i < j \text{ and } (a_i, b_j) \notin E \text{ if } j \geq i.$$

*In a bigraph $G$, we say that it contains a* $k$-half graph *if there exist a pair of sets of vertices on each part of the vertex set partition, that follows the above configuration. Also, $G$ is called $k$-edge stable, if it does not contain $k$-half graphs.*

Above, note that not all edges or nonedges are specified between the pair of sets of vertices; therefore, the property of being $k$-edge stable forbids a family of configurations ocurring in $G$, rather than a single one.

Having defined the $k$-half graph and $k$-edge stability, we state the following:

▶ **Theorem 23** ([19, Claim 2.11])**.** *Consider a hypothesis class $(X, \mathcal{H})$. The following properties are equivalent:*

**(a)** *For all $\epsilon \in (0, 1/2)$, there is a constant $c(\epsilon)$ such that every finite $A \subseteq \mathcal{H}$ has $B \subseteq A$, with $|B| \geq c|A|$, that is $\epsilon$-good.*
**(b)** *For some $\epsilon < 1/2$ and some constant $c > 0$, every finite $A \subseteq \mathcal{H}$ has $B \subseteq A$ such that $|B| \geq c|A|$, that is $\epsilon$-good.*
**(c)** $\mathrm{Ldim}(\mathcal{H}) < \infty$*; or, $\mathcal{H}$ is a Littlestone class.*
**(d)** *The bigraph $(X, \mathcal{H})$ is $k$-edge stable for some finite $k > 0$.*
**(e)** *The bigraph* $\mathrm{Bip}(\mathcal{H}, X)$ *is $\ell$-edge stable for some finite $\ell > 0$.*
**(f)** *The dual class $(\mathcal{H}, X)$ has finite Littlestone dimension.*
**(g)** *For all $\epsilon < 1/2$, there is a constant $c(\epsilon)$ such that every finite $A \subseteq X$ has $B \subseteq A$, such that $|B| \geq c|A|$, that is $\epsilon$-good.*
**(h)** *For some $\epsilon < 1/2$ and some constant $c > 0$, every finite $A \subseteq X$ has $B \subseteq A$ such that $|B| \geq c|A|$, that is $\epsilon$-good.*

Thus, Theorem 15 provides a characterization similar to some of the equivalent conditions in the cited result above. The properties in Theorem 15 are weaker, since there exist hypothesis classes $\mathcal{H}$ where $\mathrm{VC}(\mathcal{H})$ is bounded and $\mathrm{Ldim}(\mathcal{H})$ is not.

# 7 Analysis of the half graph

To further explore the distinction between Theorem 15 and Theorem 6, we consider the important basic example of the half graph, which has bounded VC dimension. Further discussion on the half graph and a few other examples of hypothesis classes of bounded VC dimension can be found in Appendix B.2.

## 7.1 Countable half graph

We begin with the following definition, which is the countable version of Definition 22:

▶ **Definition 24.** *The half graph $\mathcal{H}_\omega$ of countable size is given by two disjoint sets $\{a_i : i < \omega\}$, $\{b_i : i < \omega\}$ so that*

$$(a_i, b_j) \in E \text{ if } i < j \text{ and } (a_i, b_j) \notin E \text{ if } i \geq j.$$

Note that the VC dimension of $\mathcal{H}_\omega$ is 1, and its Littlestone dimension is infinite. The observations we show still hold true if we just work with a finite $k$-half graph, which we denote $\mathcal{H}_k$. Its VC dimension is 1 and its Littlestone dimension is $\lfloor \log(k) \rfloor$.

## 7.2 Good sampling

The property of good sampling (cf. Definition 14) holds, as a corollary of Theorem 15.

▶ **Example 25.** The half graph $\mathcal{H}_\omega$ has $(c_1, c_2, \epsilon)$-good sampling for any $\epsilon \in (0, 1/2)$, where $c_1 = \epsilon/4$ and $c_2 = 1/4$.

**Proof.** Consider any two finite sets $A \subseteq X$ and $B \subseteq \mathcal{H}$. Either $(A, B)$ form an asymmetric $(\epsilon/2)$-good pair and we are done, or there exists $b_1 \in B$ which partitions $A = A_0 \sqcup A_1$ into vertices that connect and don't connect to it, and such that $|A_0| \geq (\epsilon/2) \cdot |A|$ and $|A_1| \geq (\epsilon/2) \cdot |A|$.

In the second case, the structure of $\mathcal{H}_\omega$ implies that all vertices in $A_0$ must have a smaller index than those of $A_1$. Hence, there does not exist a second element $h_2$ of $B$ that simultaneously induces a balanced partition in both $A_0$ and $A_1$. Then, by Pigeonhole Principle, there is a set $B' \subseteq B$ of size at least $|B|/2$ with defined true value towards the same $A_i$; let us take that set to be $A_0$, without loss of generality. Then, $(A_0, B')$ forms an asymmetric $(\epsilon/2)$-good pair such that $A_0$ is $(\epsilon/2)$-good to $B'$, $|A_0| \geq (\epsilon/2)|A|$, and $|B'| \geq |B|/2$.

Then, using the symmetrization argument from Lemma 18, we can extract two subsets $A'' \subseteq A_0$ and $B'' \subseteq B'$ such that $(A'', B'')$ is symmetric $\epsilon$-good, $|A''| \geq |A_0|/2 \geq \epsilon|A|/4$, and $|B''| \geq |B|/4$. ◀

## 7.3 Interaction of $B$, $\epsilon$, and $|\mathrm{Anti}(B, \epsilon)|$

Next, we see that, given $(A, B)$ a pair of finite subsets of $X$ and $\mathcal{H}$, it is possible to have small subsets of $B$ which have both *small* and *large* antigood sets *in* $A$, depending on the choice of $\epsilon$ and the *sparsity of* $B$.

One general observation is that there is a tradeoff between the value of $\epsilon$ and the size of antigood set of a set $B$ of given size $|B|$. In particular, we consider the nested sequence of antigood sets determined by the following choices of parameters $\epsilon_i$, for $i = 0, \ldots, \lceil |B|/2 \rceil$:

$$\begin{aligned}
\epsilon_0 &\in \left(0, \frac{1}{|B|}\right), \\
\epsilon_1 &\in \left[\frac{1}{|B|}, \frac{2}{|B|}\right), \\
\epsilon_{\lceil |B|/2\rceil -1} &\in \left[\frac{\lceil |B|/2\rceil - 1}{|B|}, \frac{\lceil |B|/2\rceil}{|B|}\right), \\
&\vdots \\
\epsilon_{\lceil |B|/2\rceil} &\in \left[\frac{\lceil |B|/2\rceil}{|B|}, \frac{1}{2}\right).
\end{aligned}$$

In particular, $\mathrm{Anti}(B, \epsilon_{\lceil |B|/2\rceil}) = \varnothing$ since no vertex can give an $\epsilon_{\lceil |B|/2\rceil}$-unbalanced partition of a set of size $|B|$.

The following gives an expression for the size of the antigood set of any subset of $\mathcal{H}_k$:

▶ **Proposition 26.** *Let $(X, \mathcal{H})$ be a class forming the bipartite half graph $H_\omega$. Let $B \subseteq \mathcal{H}$ defined as $B = \{h_m, \dots, h_M\}$, where $m$ and $M$ are the smallest and largest indexes. If $0 < \epsilon < 1/|B|$, then*

$$\mathrm{Anti}(B, \epsilon) = \{x_i : m \le i \le M - 1\} \subseteq A \text{ , and } |\mathrm{Anti}(B, \epsilon)| = M - m.$$

*In particular, we can let $B$ be so that the size of the antigood set is large, i.e., $M - m = \Theta(n)$ yet $|B| = o(n)$, by letting such set be sparse. We can also let $B$ so that the size of the antigood set is small, i.e., $M - m = |B|$.*

**Proof.** Any vertex set whose neighborhood is a proper subset of $B$ is in the antigood set. By the half graph's structure, such set will contain precisely the vertices $x_i$, where the index $i$ is at least $m$, and at most $M - 1$. ◀

▶ **Example 27.** Let $A = \{x_1, \dots, x_{15}\}$, $B = \{h_3, h_4, h_5, h_6, h_7\}$, and $\epsilon \in (0, 1/5)$. In this case, $\mathrm{Anti}(B, \epsilon) = \{x_3, x_4, x_5, x_6\}$, which is the smallest possible nonzero size that the antigood set of a set in $B$ of size 5 can have here.

If instead we had chosen $B = \{x_1, x_5, x_6, x_{11}, x_{15}\}$ and let $\epsilon \in (0, 1/5)$, then $\mathrm{Anti}(B, \epsilon) = \{x_1, \dots, x_{14}\}$, which is the largest possible size that the antigood set of a set in $B$ can have.

A constructive proof of the existence of the $\epsilon$-good symmetric pair $(A', B')$ can be found in Appendix B.1.

One of the observations about the example above is that sparse sets have large antigood sets. In the following, we will show a stronger property than the existence of sets with large antigood sets. Moreover, it becomes highly likely that a subset of $X$ in a half graph will have a large antigood set, due to its sparseness.

▶ **Proposition 28.** *Let $\epsilon > 0$ and consider the half graph $(X, \mathcal{H}_n)$. There exists $M = M(\epsilon) = O(1/\epsilon^2)$ such that the following holds. Let $H \subseteq \mathcal{H}_n$ be a subset of size at least $M$, sampled with respect to the uniform probability distribution $\mathcal{D}$ over $[n]$. With high probability of at least $(1 - \epsilon)$, $H$ has a large antigood set, and more concretely:*

$$\mathrm{Pr}_{H \sim \mathcal{D}^M}[\, |\mathrm{Anti}(H, \epsilon)| \ \ge n - 2\epsilon n] \ge 1 - \epsilon.$$

**Proof.** Consider the set system $\mathcal{S} = \{S_1, \ldots, S_{\lfloor 1/\epsilon \rfloor}\}$ on the ground set $[n]$, where each set $S_i$ is one of $\lfloor 1/\epsilon \rfloor$ nearly equally sized parts $S_i := \{s_{i-1}+1, \ldots, s_i\}$, where the values $s_i$ are given by:

$$s_0 = 0, \text{ and for all } i \geq 1,\ s_i = s_{i-1} + \left\lfloor \frac{n}{\lfloor 1/\epsilon \rfloor} \right\rfloor + \mathbb{1}[i \leq n \mod \lfloor 1/\epsilon \rfloor].$$

Since $\mathrm{VC}(\mathcal{S}) = 1$, the Fundamental Theorem of Statistical Learning (see Appendix A.2) implies that this set system has the uniform convergence property. Namely, for any $\epsilon > 0$, there exists $M = M(\epsilon) = O(d/\epsilon^2)$ (which is $O(1/\epsilon^2)$ since $d = 1$), such that:

$$\Pr_{H \sim \mathcal{D}^M} \left[ \sup_{i=1,\ldots,n/\epsilon} |\mu_H(S_i) - \mu(S_i)| < \epsilon \right] \geq 1 - \epsilon,$$

where $\mu_H$ is the empirical counting measure on $[n]$ and $\mu$ is the counting measure on $[n]$. In particular, the above implies that for all $i = 1, \ldots, \lfloor 1/\epsilon \rfloor$, it must happen that $H \cap S_i \neq \varnothing$. Otherwise:

$$\mu_H(S_i) = \frac{|H \cap S_i|}{n} = 0 \ \text{ whereas } \ \mu(S_i) = \frac{\epsilon n}{n} = \epsilon,$$

a contradiction. In particular, $H \cap S_1 \neq \varnothing$ and $H \cap S_{\lfloor 1/\epsilon \rfloor} \neq \varnothing$, and by Proposition 26,

$$|\mathrm{Anti}(H, \epsilon)| \geq (n - \lfloor n\epsilon \rfloor) - \lfloor n\epsilon \rfloor \geq n - 2\epsilon n. \qquad ◀$$

In particular, in examples where the graph has finite VC dimension and infinite Littlestone dimension, it is possible to find a large half graph, in which many instances of sets with large antigood sets can be found, as Proposition 28 shows.

## 8 Comparison with proof via Haussler Packing

In this section we present an alternative way which allows us to find a large set with a small antigood set by applying directly Haussler's Packing Lemma [16]. We compare the outcome of this method if applied in a setup with the same initial conditions. Please see the cited results and lemmas in Appendix A.3.
In the setting we work on, we are given $A \subseteq X$ and $B \subseteq \mathcal{H}$ finite sets, and $\mathrm{VC}(\mathcal{H}) = d$. Our goal for this section will be to show that a lower bounded sized portion of the elements in $B$ lies within some ball of radius $\delta$, which then has a small antigood set. In order to do so, we can apply the following implication of Haussler's Packing Lemma:

▶ **Lemma 29.** *If $(A, B)$ is a finite pair in $(X, \mathcal{H})$ whose* VC *dimension is $d$, then there exists some element $b \in B$ whose $\delta$-neighborhood $B_\delta(b)$ is of size at least $|B| \cdot (e(d+1))^{-1} \cdot (\delta/2e)^d$.*

**Proof.** Let a set $C$ be constructed by adding elements of $B$ to it so that the elements added so far are always pairwise $\delta$-separated. By Haussler's Packing Lemma, this process must stop after no more than $e(d+1) \cdot (2e/\delta)^d$ many steps. After doing this, no elements can be added, hence any other element in $B$ has to be $\delta$-close to at least one of the elements one element of $C$. By Pigeonhole Principle, at least $|B| / \left(e(d+1) \cdot (2e/\delta)^d\right) = |B| \cdot (e(d+1))^{-1} \cdot (\delta/2e)^d$ elements of $B$ are $\delta$-close to the same element $b \in B$. Thus, $|B_\delta(b)| \geq |B| \cdot (e(d+1))^{-1} \cdot (\delta/2e)^d$. ◀

▶ **Proposition 30.** *For any two $h_1, h_2$ in $B_\delta(h_0) \subseteq \mathcal{H}$, the distance between these two functions is $d(h_1, h_2)$ less than $2\delta$.*

**Proof.** This fact is a direct application of the triangle inequality for the error distance:

$$d(h_1, h_2) = \frac{|S_{h_1} \Delta S_{h_2}|}{|X|} \leq \frac{|S_{h_1} \Delta S_{h_0}|}{|X|} + \frac{|S_{h_0} \Delta S_{h_2}|}{|X|} < \delta + \delta. \qquad \blacktriangleleft$$

▶ **Proposition 31.** *Let $\epsilon \in (0, 1/2)$, and let $B = B_\epsilon(h_0)$ be any $\delta$-ball centered at a hypothesis $h_0$. Then the size of the set $\mathrm{Anti}(B, \epsilon)$ is no larger than:*

$$\min\left\{ \frac{|B| - 1}{|B|} \cdot \frac{\delta}{\epsilon(1-\epsilon)} \cdot |A| \,,\quad |A| \right\}.$$

*Note that $|A|$ is always larger than the antigood set's size: this statement is always true regardless of the choice of $\delta > 0$. As we will see later, certain particular choices of $\delta$ will be so that $|A|$ is significantly larger than $|\mathrm{Anti}(B, \epsilon)|$.*

**Proof.** In this proof, we will double count the size of the following set of triples

$$T := \{(h_1, h_2, x) \in B \times B \times A \ : \ h_1(x) \neq h_2(x)\}.$$

Consider any $x \in \mathrm{Anti}(B, \epsilon) \subseteq A$. Since this element is in the antigood set, the size of both its neighborhood and its complement in $B$ is between $\epsilon|B|$ and $(1-\epsilon)|B|$. Since the function $x(1-x)$ is decreasing with respect to $|x - 0.5|$, we have that:

$$|\{(h_1, h_2) \in B \times B \ : \ h_1(x) \neq h_2(x)\}| \geq 2\epsilon|B| \cdot (1-\epsilon)|B|.$$

If we add together over all elements $x$ in the antigood set, we get

$$|T| \geq \epsilon|B| \cdot (1-\epsilon)|B| \cdot |\mathrm{Anti}(B, \epsilon)|.$$

Now, consider any pair of distinct elements $(h_1, h_2) \in B \times B$. By Proposition 30, any such tuple is $2\delta$-close and therefore

$$|\{x \in A \ : \ h_1(x) \neq h_2(x)\}| \leq 2\delta \cdot |A|.$$

If we add together over all the possible ordered pairs $(h_1, h_2)$ in $B \times B$, we get that

$$|T| \leq 2\delta \cdot |A| \cdot |B| \cdot (|B| - 1).$$

The above inequalities are lower and upper bounds for the same count for the size of the set $T$. Threading the two together, and then simplifying, we obtain the claim, which is an upper bound on the antigood set size. ◀

Note that the above upper bound is stronger than $|A|$ only if $\delta$ is not much larger than $\epsilon$. In particular, $\delta \leq 2\epsilon$ since we need:

$$\frac{\delta}{\epsilon} < (1-\epsilon) \cdot \frac{|B|}{|B| - 1} < 2.$$

One implication of the work done above is that we can now reprove Corollary 19, now using Proposition 31 and therefore obtaining distinct constant values. In particular, if we set $\delta = \epsilon/2$, we get $(\epsilon, c_1, c_2)$-good sampling with:

$$c_1 = \left(\frac{\epsilon}{60}\right)^d, c_2 = \frac{1}{2}.$$

The reader may wonder how this result compares with the bounds from Corollary 19. We can point out in the case that $\delta = \frac{\epsilon}{2}$, that even if the bound on the size of the antigood set is better (or the symmetric good pair is larger in size):

$$|\mathrm{Anti}(B,\epsilon)| \leq \frac{|B|-1}{|B|} \cdot \frac{1}{2(1-\epsilon)} \cdot |A|$$

Still, the lower bound on the resulting set $B = B_\delta(h_0)$ is much smaller, as its size is lower bounded only by $|B| \cdot (\epsilon/60)^d$.

## 9 Comparison with proof via VC Regularity Lemma

Following up on the proof of Theorem 19, we can also find an alternative way to obtain good pairs by usage of the VC Regularity Lemma [2, 18, 13]. The statement of this regularity lemma can be found in Appendix A.1.

In this setting, we are again given $A \subseteq X$ and $B \subseteq \mathcal{H}$ finite sets, and $\mathrm{VC}(\mathcal{H}) = d$. We can then directly use the equipartition obtained from applying this regularity lemma in order to obtain the following bounds:

▶ **Proposition 32.** *Consider a class $(X, \mathcal{H})$ such that $\mathrm{VC}(\mathcal{H}) = d$, finite subsets $A \subseteq X$, $B \subseteq \mathcal{H}$, and a parameter $\epsilon \in (0, 1/204)$. It is possible to find a symmetric $(100^2\epsilon/99)$-good pair $(A', B')$ of subsets of $A$ and $B$, respectively, such that, for some $m = O(1/\epsilon^{2d+1})$:*

$$|A'| \geq (99 \cdot |A|)/(100 \cdot m),$$
$$|B'| \geq (99 \cdot |B|)/(100 \cdot m).$$

**Proof.** Using the VC Regularity Lemma, let us take first an equipartition of the bigraph $(A, B)$ such that each of $A, B$ are partitioned into $m = m(\epsilon)$ nearly-equal sized parts, with the bound $m = O(1/\epsilon^{2d+1})$, and such that all but at most $\epsilon \cdot m^2$ of them are $\epsilon$-homogeneous, i.e., they have edge density in $[0, \epsilon) \cup (1-\epsilon, 1]$. In particular, let $(A, B)$ be one of these $\epsilon$-homogeneous pairs.

Without loss of generality, $e(A, B) < \epsilon|A| \cdot |B|$. We define the sets

$$S_A := \{a \in A : |N(a) \cap B| \geq 100\epsilon \cdot |B|\}, \text{ and}$$

$$S_B := \{b \in B : |N(b) \cap A| \geq 100\epsilon \cdot |A|\}.$$

For $e(A, B) < \epsilon|A| \cdot |B|$ to hold, it must be the case that $|S_A| \leq |A|/100$ and $|S_B| \leq |B|/100$. Defining $A' := A \setminus S_A$ and $B' = B \setminus S_B$, we have that

$$|A'| \geq |A| \cdot \frac{99}{100} \quad \text{and} \quad |B'| \geq |B| \cdot \frac{99}{100},$$

$$\Longrightarrow \forall a \in A', e(a, B') \leq e(a, B) \leq 100\epsilon \cdot |B| \leq \frac{100^2\epsilon}{99} \cdot |B'|.$$

Similarly, for every $b \in B'$, the number of edges from $b$ to $A'$ is at most $(100^2\epsilon/99) \cdot |A'|$, which implies that the pair $(A', B')$ is symmetric $(100^2\epsilon/99)$-good (and thus, $102\epsilon$-good, with $102\epsilon \in (0, 1/2)$). This finishes the proof. ◀

Let us put into perspective the bounds on the size of the symmetric pair obtained from Proposition 32 with those of the pair we previously obtained. Since the number of parts was $m = O(1/\epsilon^{2d+1})$, then the $|A'| \geq (99 \cdot |A|)/(100m) = \Omega(|A| \cdot \epsilon^{2d+1})$ (and $|B'| = \Omega(|B| \cdot \epsilon^{2d+1})$).

This gives a worse bound at least in the sense, that the constant values $c_1 = \epsilon^d/2^{d+1}$ and $c_2 = 1/2^{d+1}$ from Theorem 15 result in the existence of a pair with a higher lower bound in size.

The following is one more application of the VC Regularity Lemma towards showing the existence of good pairs or sets with bounded antigood sets in VC classes:

▶ **Corollary 33.** *Let $n > 0$ be a positive integer, and $\epsilon > 0$ be a positive real such that $\epsilon < 1/n$. If $k \in \mathbb{N}$, then there exists a constant $c = c(k)$ such that: for every class $(X, \mathcal{H})$ with $\mathrm{VC}(\mathcal{H}) = d < \infty$ and every finite subsets $A \subseteq \mathcal{H}$ and $B \subseteq X$, there is some set $A' \subseteq A$ of size $|A'| \geq c^{-1} \cdot |A| \cdot \epsilon^{2d+1}$ such that*

$$|\mathrm{Anti}(A', \epsilon n) \cap B| \leq |B| \cdot \frac{2}{n}.$$

**Proof.** By the VC Regularity Lemma, there exists a near equipartition of each of $A$ and $B$ into $m \leq c \cdot (1/\epsilon)^{2d+1}$ parts, for some constant $c = c(d)$, such that all but at most $\epsilon \cdot m^2$ of the pairs of parts obtained are $\epsilon$-homogeneous. Suppose that $A$ is partitioned as $\bigcup_{i \in [m]} S_i$, and that $B$ is partitioned as $\bigcup_{j \in [m]} T_j$. By Pigeonhole Principle, at least one of the resulting parts, say $A' := S_1$ without loss of generality, doesn't form an $\epsilon$-homogeneous pair with $\leq \epsilon \cdot m$ pairs, which must be among the $T_j$.

For those $T_j$ such that $(A', T_j)$ is $\epsilon$-homogeneous, either $e(A', T_j) < \epsilon \cdot |T_j| \cdot |A'|$, or $ne(A', T_j) < \epsilon \cdot |T_j| \cdot |A'|$ (here, *ne* stands for *non-edges*). In the first case, there are less than $|T_j|/n$ vertices $v \in T_j$ such that $e(v, A') \geq \epsilon n |A'|$. In the second case, there are less than $|T_j|/n$ vertices $v \in T_j$ such that $ne(v, A') \geq n\epsilon|A'|$. In either case, then, $|\mathrm{Anti}(A', \epsilon n) \cap T_j| < |T_j|/n$.

We can get the following upper bound by aggregating:

$$\begin{aligned}
|\mathrm{Anti}(A', \epsilon n)| &= \sum_{i=1}^{m-1} |\mathrm{Anti}(A', \epsilon n) \cap T_j| \\
&\leq \sum_{i:(T_j, A')\ \epsilon\text{-hom.}} \frac{|T_j|}{n} + \sum_{i:(T_j, A')\text{not}\ \epsilon\text{-hom.}} |T_j| \\
&\leq \frac{|B|}{n} + \left(1 - \frac{1}{n}\right) \cdot \sum_{i:(T_j, A')\text{not}\ \epsilon\text{-hom.}} |T_j| \\
&= |B|(\epsilon - \epsilon/n + 1/n)
\end{aligned}$$

By the condition on $\epsilon$, the above is less than $|B| \cdot (2/n)$. ◀

The above shows that for finite bigraphs $(A, B)$ embedded in a class of bounded VC dimension, there exists a large set $A' \subseteq A$ such that $|\mathrm{Anti}(A', \epsilon n)| \leq 2|B|/n$. Namely, its antigood set with error parameter $n\epsilon$, has an upper bounded size.

It remains to explore how large $|\mathrm{Anti}(A', \epsilon n)|$ is. In fact, the upper bound for the ratio is not small a priori, especially if $\epsilon$ is close to zero.

$$\frac{|\mathrm{Anti}(A', n\epsilon)|}{|B|} \leq \frac{|B| \cdot 2/n}{|B| \cdot c^{-1} \cdot \epsilon^{2d+1}} = \frac{2c}{n \cdot \epsilon^{2d+1}}.$$

The above uses of the VC Regularity Lemma suggest that it is worth clarifying the interactions between the pair property of being $\epsilon$-homogeneous and that of being symmetric $\epsilon$-good.

## 10 Concluding remarks

This work provides a characterization of classes with bounded VC dimension. A few open questions include:

1. What results can be shown in the case that we work with graphs, instead of bigraphs?
2. How does the existence of $\epsilon$-good pairs interact with learning theoretical questions and existing results? In particular, part of the Fundamental Theorem of Statistical Learning may have interactions with this characteristic of VC classes.
3. Let us denote this class $(X, \mathcal{H}_{H,n})$, and define it as the class on the input domain $X = \mathbb{R}^n$ of all closed halfspace memberships, namely:

$$\mathcal{H}_{H,n} = \left\{ \{ x \in \mathbb{R}^n \middle| \langle \vec{x}, \vec{a} \rangle \geq 0 \} : \vec{a} \in \mathbb{R}^n \setminus \{\vec{0}\} \right\}$$

   Knowing that $\mathrm{VC}(\mathcal{H}_{H,n}) = n + 1 < \infty$, a possible subject of further study is the good sampling property for this class.
4. Is there anything from our analysis that can allow us to find half-graphs by using the $\epsilon$-good pairs shown in Corollary 19, or in any of the other analogous results that use different technologies?
5. Is there any other technology that could be used in order to improve the bounds obtained by using Haussler's Packing Lemma as done in Section 8?

Another more generally phrased follow up is to look for specific tradeoffs between the bounds obtained in each of the results, and to point to which bound is more useful in the aforementioned (or other) applications.

### References


1. Noga Alon, Mark Bun, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private and online learnability are equivalent. In *Journal of the ACM (Association for Computing Machinery)*, volume 69, No. 4, Article 28. ACM, 2022.
2. Noga Alon, Eldar Fischer, and Ilan Newman. Efficient testing of bipartite graphs for forbidden induced subgraphs. *SIAM Journal on Computing Vol. 37, Issue. 3*, 2007.
3. Shai Ben-David, Dávid Pál, and Shai Shalev-Shwartz. Agnostic online learning. In *22nd Annual Conference Computational Learning Theory*, 2009. URL: `https://api.semanticscholar.org/CorpusID:9403043`.
4. Achim Blumensath. Simple monadic theories and indiscernibles. *Mathematical Logic Quarterly 57 (1)*, pages 65–86, 2011. `doi:10.1002/MALQ.200910121`.
5. Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred Warmuth. Learnability and the Vapnik–Chervonenkis dimension. *Journal of the Association of Computing Machinery*, 36.4:929–965, 1989. `doi:10.1145/76359.76371`.
6. Hunter Chase and James Freitag. Model theory and machine learning. *The Bulletin of Symbolic Logic*, 25:319–332, 2019. `doi:10.1017/BSL.2018.71`.
7. David Conlon and Jacob Fox. Bounds for graph regularity and removal lemmas. *Geom. Funct. Anal. 22:5*, pages 1191–1256. MR Zbl, 2012.
8. Jan Dreier, Nikolas Mählmann Ioannis Eleftheriadis, Rose McCarty, Michał Pilipczuk, and Szymon Toruńczyk. First-order model checking on monadically stable graph classes. *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, 2024.
9. Jan Dreier, Nikolas Mählmann, Sebastian Siebertz, and Szymon Toruńczyk. Indiscernibles and flatness in monadically stable and monadically NIP classes. *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, 2023.
10. P. Erdős and A. Hajnal. Ramsey-type theorems. *Discrete Applied Mathematics*, 1989.
11. Jörg Flum and Martin Grohe. *Parameterized Complexity Theory*. Springer, 2006. `doi:10.1007/3-540-29953-X`.

12 Jacob Fox and László Lovász. A tight lower bound for Szemerédi's Regularity Lemma". *Combinatorica 37:5*, pages 911–951. MR Zbl, 2017.

13 Jacob Fox, János Pach, and Andrew Suk. Erdős–Hajnal conjecture for graphs with bounded VC dimension. *Discrete Computational Geometry*, 61(4):809–829, 2019. `doi:10.1007/S00454-018-0046-5`.

14 W.T. Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *GAFA, vol. 7*, pages 322–337, 1997.

15 W.T. Gowers. Lower bounds of tower type for Szemerédi's Uniformity Lemma". *Geom. Funct. Anal., 7:2*, pages 322–337, MR Zbl,, 1997.

16 David Haussler. Sphere packing numbers for subsets of the boolean $n$-cube with bounded Vapnik–Chervonenkis dimension. *Journal of Combinatorial Theory Ser. A*, 69:217–232. 5, 1995.

17 Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2:285–318, 1987. `doi:10.1007/BF00116827`.

18 László Lovász and Balázs Szegedy. Regularity partitions and the topology of graphons. *Bárány, I., Solymosi, J., Sági, G. (eds) An Irregular Mind.*, 21:415–446, 2010.

19 Maryanthe Malliaris and Shay Moran. Agnostic online learning and excellent sets. *Transactions of the American Mathematical Society*, 2024. `arXiv:2108.05569`.

20 Maryanthe Malliaris and Saharon Shelah. Regularity lemmas for stable graphs. *Transactions of the American Mathematical Society*, 366(3):1551–1585, 2014. `doi:10.1090/S0002-9947-2013-05820-5`.

21 Maryanthe Malliaris and Saharon Shelah. Notes on the stable regularity lemma. *The Bulletin of Symbolic Logic*, 27, No.4:415–425, 2021.

22 Guy Moshkovitz and Asaf Shapira. A short proof of Gowers' lower bound for the Regularity Lemma". *Combinatorica 36:2*, pages 187–194. MR Zbl, 2016.

23 Tung Nguyen, Alex Scott, and Paul Seymour. Induced subgraph density. VI. Bounded VC-dimension. *Advances in Mathematics*, 2025.

24 Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms.* Cambridge University Press, 2014.

25 Endre Szemerédi. Regular partitions of graphs. *Colloques Internationaux du CNRS*, 260:399–401, 1978.

26 Vladimir Vapnik and Alexey Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, XVI, no.2:264–280, 1971.


# A Appendix: referenced and used results

## A.1 Regularity Lemma

A large collection of work in extremal combinatorics concerns the large scale properties of certain classes of graphs. In particular, some classes with forbidden configurations are found to satisfy stronger versions of regularity lemmas. Szemerédi's Regularity Lemma is a result in extremal combinatorics that allows us to approximate large graphs with smaller ones of fixed size. Below, an $\epsilon$-regular pair $(A, B)$ is one such that for every $A' \subseteq A$ and $B' \subseteq B$, such that $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$, the difference between the edge densities $d(A, B) := e(A, B)/|A| \cdot |B|$ and $d(A', B') := e(A', B')/(|A'| \cdot |B'|)$ is less than $\epsilon$.

▶ **Theorem 34** (Szemerédi's Regularity Lemma, [25]). *For every $\epsilon > 0$ there is $N(\epsilon) \in \mathbb{N}$ so that every finite graph $G$ may be partitioned into $m$ classes $V_1 \cup \cdots \cup V_m$ where $m \leq N$ and*

**(i)** *all of the pairs $V_i, V_j$ satisfy $||V_i| - |V_j|| \leq 1$;*

**(ii)** *all but at most $\epsilon m^2$ of the pairs $(V_i, V_j)$ are $\epsilon$-regular.*

**Table 1** Three different Regularity Lemmas.

| Assumption | No. parts | Guarantees |
| --- | --- | --- |
| General | $N(\epsilon)$ | all but $m\epsilon^2$ pairs $\epsilon$-regular |
| $VC(\mathcal{H}) = d$ | $N = O(1/\epsilon^{2d+1})$ | all but $k\epsilon^2$ pairs $\epsilon$-homogeneous |
| $k$-stable | $N = O\left(1/\epsilon^{2^{k+3}-7}\right)$ | all pairs $\epsilon$-regular |

One caveat of the above is that $N(\epsilon)$ may be a tower of exponents, as shown by Gowers in [14], and in the lower bound construction in [15]; such tower dependence was also shown in [12, 7, 22] and was shown to be characterized by infinite VC dimension. Another caveat is the existence of irregular pairs, which is characterized by the existence of half graphs.

Before we state the VC Regularity Lemma, we define VC dimension for graphs:

▶ **Definition 35.** *Let $G = (V, E)$ be a graph, and define the set system of $G$ as the bigraph where both sides have $V$'s elements as vertices, and whose edges are given by pairs $(v, w)$ such that the vertices $v, w \in V$ are connected by an edge $e \in E$. In the sense of Definition 6, we define* VC$(G)$ *as the* VC *dimension of this bigraph.*

For graphs $G$ such that their $VC$-dimension is bounded, it is possible to find a smaller $N(\epsilon)$ than originally, allowing to find $N(\epsilon)$-sized equipartitions with almost all of the pairs being $\epsilon$-homogeneous. Here, an $\epsilon$-homogeneous pair $(A, B)$, is one such that $d(A, B) \in [0, \epsilon) \cup (1-\epsilon, 1]$.

▶ **Theorem 36** (Bounded-VC Regularity Lemma [18, 2, 13])**.** *For any $\epsilon > 0$, there is a $N = N(\epsilon) \in \mathbb{N}$ such that the vertex set $V$ of any finite graph $G$ with $VC$-dimension $d < \infty$ has a partition into at most $m \leq N$ parts. Here, $N \in [8/\epsilon, c \cdot \epsilon^{-2d+1}]$, where $c = c(d)$ is a constant only depending on $d$, and*

**(i)** *all of the pairs $V_i, V_j$ satisfy $||V_i| - |V_j|| \leq 1$;*

**(ii)** *all but at most $\epsilon \cdot N^2$ pairs are $\epsilon$-homogeneous*

An even stronger property holds if we focus on the class of $k$-edge stable graphs. Namely, there is a regularity lemma that gives stronger bounds and properties, as stated below.

▶ **Theorem 37** (Stable Regularity Lemma [20])**.** *For each $\epsilon > 0$ and $k \in \mathbb{N}$ there is $N = N(\epsilon, k)$ such that for a sufficiently large finite $k$-edge stable graph $G = (V, E)$, there exists $\ell \leq N$ such that $V$ can be partitioned into disjoint pieces $A_1, \ldots, A_\ell$ such that*

**(i)** *$||A_i| - |A_j|| \leq 1$ for all $i, j$,*

**(ii)** *$(A_i, A_j)$ is $\epsilon$-regular for all $i, j$,*

**(iii)** *$A_i$ is $\epsilon$-excellent for all $i$,*

**(iv)** *$N < (4/\epsilon)^{2^{k+3}-7}$.*

Above, the $poly(\epsilon^{-1})$ bound on the number of parts and the guarantee that all pairs will be $\epsilon$-regular show that the class of stable graphs enjoy stronger regularity properties than the class of all graphs.

## A.2 Fundamental Theorem of Statistical Learning

The work of [5] provides a characterization of PAC-learnability by finiteness of VC dimension:

▶ **Theorem 38.** *Let $\mathcal{H} \subseteq \{\pm 1\}^X$ be a hypothesis class on the input domain $X$. The following properties are equivalent:*

1. *$\mathcal{H}$ is (agnostic) PAC-learnable;*
2. *$\mathcal{H}$ is (realizable) PAC-learnable;*

3. $\mathrm{VC}(\mathcal{H}) = d$ *is finite;*
4. $\mathcal{H}$ *has the uniform convergence property;*
5. *there exists an ERM algorithm that learns* $\mathcal{H}$.

In particular, the quantitative statement (see e.g., [24]) states that the uniform convergence property of $\mathcal{H}$ is satisfied with sample complexity $m_{\mathcal{H}}^{\mathrm{UC}}(\epsilon, \delta)$:

$$C_1 \cdot \frac{d + \log(1/\delta)}{\epsilon^2} \leq m_{\mathcal{H}}^{\mathrm{UC}}(\epsilon, \delta) \leq C_2 \cdot \frac{d + \log(1/\delta)}{\epsilon^2},$$

where $C_1$ and $C_2$ are constants. In other words, the uniform convergence property is satisfied as long as $m_{\mathcal{H}}^{\mathrm{UC}}(\epsilon, \delta)$ many samples were taken. This is $O(d/\epsilon^2)$ when the value of $\delta$ has been fixed.

## A.3 Haussler's Packing Lemma

We cite the results and notions we used from the work in [16], beginning with the following distance notion among elements of $\mathcal{H}$ in the pair $(X, \mathcal{H})$. Here, we assume $|X|$ to be finite.

▶ **Definition 39.** *For any two hypotheses $h_1, h_2 \in \mathcal{H}$, their distance is given by the counting measure*

$$d(h_1, h_2) := \frac{|S_{h_1} \Delta S_{h_2}|}{|X|},$$

*where $S_h := \{x \in X : h(x) = 1\}$. Then, an $\epsilon$-ball around an element $h_0$ is defined as follows:*

$$B_\epsilon(h_0) := \{h \in \mathcal{H} \ : \ d(h, h_0) < \epsilon\}.$$

In other words, a pair of functions is $\epsilon$-close is they roughly look the same to the sample $X$, and we can also refer to this pair of functions to be $\epsilon$-close in the error measure.

In the remaining part of this section, we will consider mostly interchangeably the notions of the pair $(X, \mathcal{H})$ as a set system in the sense that $\mathcal{H} \subseteq 2^X$, and as a hypothesis class in the sense that each element $h \in \mathcal{H}$ is a function $h : X \to \{0, 1\}$, since they will allow us to work with the same combinatorial structure.

A next definition that will be needed to understand the statement and use of Haussler's Packing Lemma is the following:

▶ **Definition 40.** *A set system $(X, \mathcal{H})$ on $X$ is* $\delta$-separated *when $|h_1 \Delta h_2| \geq \delta|X|$, for all $h_1, h_2 \in \mathcal{H}$.*

An alternative way of stating the above is that if a set system is $\delta$-separated, then for every pair of elements $h_1, h_2$ in it, it has to happen that $h_1 \notin B_\epsilon(h_2)$, and vice versa. Haussler's Packing Lemma notes that a general set system of bounded VC dimension cannot be both large and $\delta$-separated.

▶ **Theorem 41** ([16])**.** *Let $d, \delta > 0$. If $(X, \mathcal{H})$ are such that $\mathrm{VC}(\mathcal{H}) \leq d$, there are at most*

$$e(d+1) \cdot \left(\frac{2e}{\delta}\right)^d$$

*elements in $\mathcal{H}$ if it is a $\delta$-separated system of $X$.*

# B Follow-ups for Section 7

## B.1 Constructive proof of good sampling property for the half graph

In Section 7, the $(c_1, c_2, \epsilon)$ good sampling property was shown to hold for the half graph $\mathcal{H}_\omega$, where $c_1 = \epsilon/4$, $c_2 = 1/4$, and $\epsilon$ is any value in $(0, 1/2)$. We show this constructively, now indicating two possible $\epsilon$-symmetric good pairs for this example.

▶ **Proposition 42.** *Consider $(X, \mathcal{H})$ with half graph structure $\mathcal{H}_\omega$ and two finite subsets $A \subset X$ and $B \subset \mathcal{H}$. Let $A_1$ be the set given by the leftmost $c_1|A|$ vertices of $A$ according to indexing. Let $A_2$ be the set given by the rightmost $c_2|A|$ vertices of $A$, correspondingly. Then, define $B_1$ and $B_2$ analogously.Then, at least one of the pairs $(A_i, B_j)$ with $i, j \in \{1, 2\}$ is symmetric $\epsilon$-good.*

**Proof.** Let $\ell(A_1)$ and $\ell(B_1)$ be the last indexes of elements of $A_1$ and $B_1$, respectively. Also let $f(A_2)$ and $f(B_2)$ be the first indexes of the elements of $A_2$ and $B_2$.

We claim that, either the sets of indexes of $A_1$ and $B_2$ are disjoint, or the sets of indexes of $A_2$ and $B_1$ are disjoint. By definition of the half graph structure, $(A_1, B_2)$ is either a complete bigraph, or $(A_2, B_1)$ is an empty bigraph, both of which are $\epsilon$-symmetric good pairs. Suppose this is not true for the sake of contradiction. Then, the following must hold:

$$\ell(A_1) > f(B_2),$$
$$\ell(B_1) > f(A_2).$$

By the assumptions on the sizes of the subsets $A_i, B_j$, we have

$$f(B_2) \geq \ell(B_1) + (1 - 2c_2)|B|,$$
$$f(A_2) \geq \ell(A_1) + (1 - 2c_1)|A|.$$

Threading all these inequalities together, we obtain the absurd:

$$\ell(A_1) > \ell(A_1) + (1 - 2c_1)|A| + (1 - 2c_2)|B|.$$

Therefore it must be that one of $(A_1, B_2)$ or $(A_2, B_1)$ forms a complete or empty bipartite graph, hence symmetric $\epsilon$-good pair. ◀

## B.2 Other examples of VC classes

Also following up from Section 7, we visit two more examples of VC classes in which, under an appropriate choice of $\epsilon \in (0, 1/2)$, antigood sets may be large.

The first example consists of the class of *intervals* in $\mathbb{R}$. Let us denote this class $(X, \mathcal{H}_I)$, and define it to have the input domain $X = \mathbb{R}$ and to consist of all closed interval memberships, namely:

$$\mathcal{H}_I = \{h_{a,b} := \mathbb{1}_{[a,b]} \ : \ a, b \in \mathbb{R}\}$$

We know that $\mathrm{VC}(\mathcal{H}_I) = 2$, and have the following.

▶ **Example 43.** Consider $B$ to be the set of all functions $h_{a,b}$ given by the intervals $[a, b]$, where $a$ and $b$ take integer values from 0 to $N$. Consider $A$ to be the set of real numbers $\{-0.5, 0.5, \ldots, N + 0.5\}$. Then, if $\epsilon < \frac{1}{|N|}$, $\mathrm{Anti}(B, \epsilon) = \{0.5, \ldots, N - 0.5\}$, which is large in $A$, since

$$\frac{|\mathrm{Anti}(B, \epsilon)|}{|A|} = \frac{N - 1}{N + 1}$$

The second example consists of the class of *rectangles* in $\mathbb{R}^2$. Let us denote this class $(X, \mathcal{H}_R)$, and define it as the class on the input domain $X = \mathbb{R}^2$ of all closed rectangle memberships, namely:

$$\mathcal{H}_R = \{h_{a,b,c,d} := \mathbb{1}_{[a,b]\times[c,d]} \ : \ a,b,c,d \in \mathbb{R}\}$$

We know that $\mathrm{VC}(\mathcal{H}_R) = 4$, and have the following.

▶ **Example 44.** Consider $B$ to be the set of all functions $h_{a,b,c,d}$ given by the rectangles $[a,b]\times[c,d]$, where $a,b,c$ and $d$ take integer values from 0 to $N$. Consider $A$ to be the set of points $\{(a-0.5, b-0.5) \ : \ a,b \in \{0,\dots,N\}\}$. Then, if $\epsilon < N/\binom{N}{2}^2 = O(1/N^3)$, we have that $\mathrm{Anti}(B,\epsilon) = \{(a-0.5, b-0.5) \ : \ 1 \le a,b \le N\}$, which is large in $A$, since

$$\frac{|\mathrm{Anti}(B,\epsilon)|}{|A|} = \frac{N^2}{(N+1)^2}.$$

## C Summary of results

Table 2 summarizes the results of the these main proofs (see Theorem 15, Corollaries 19 and 31, Proposition 32, and Corollary 33).

**Table 2** A table summarizing the results on symmetric good pairs and large sets with small antigood sets which can be found characteristically in classes of VC dimension bounded by some fixed $d \in \mathbb{N}$. In the table below, $(c_1, c_2, \epsilon)$-good sampling stands for the following property: *For any $A \subseteq X, B \subseteq \mathcal{H}$ finite, $\exists A' \subseteq A, B' \subseteq B$, such that $|A'| \ge c_1|A|$, $|B'| \ge c_2|B|$, and $(A', B')$ forms a symmetric $\epsilon$ good pair.* The antigood set bound stands for the following: *For any $B \subseteq \mathcal{H}$, there is $B' \subseteq B$ with small $\epsilon'$-antigood set: concretely, $|B'| \ge c_2' \cdot |B|$, $|\mathrm{Anti}(B',\epsilon)| \le c_1'|A|$.*

| | $(c_1, c_2, \epsilon')$-good sampling | Antigood set bound |
|---|---|---|
| Combinatorial | $\epsilon' = \epsilon$<br>$c_1 = \epsilon^d/2^{d+1}$<br>$c_2 = 1/2^{d+1}$ | $\epsilon' = \epsilon$<br>$c_1' = 1-(\epsilon/2)^d$<br>$c_2' = 1/2^d$ |
| Haussler Packing | $\epsilon' = 2\epsilon$<br>$c_1 = (1/2)\cdot\left(1 - \frac{\lvert B\rvert-1}{2\lvert B\rvert(1-\epsilon)}\right)$<br>$c_2 = (1/2)\cdot(\epsilon/60)^d$ | $\epsilon' = 2\epsilon$<br>$c_1' = \max\left\{1, \frac{(\lvert B\rvert-1)\delta}{\lvert B\rvert\epsilon(1-\epsilon)}\right\}$<br>$c_2' = (\delta/30)^d$, $\delta > 0$ |
| VC Regularity | $\epsilon' = (100^2/99)\epsilon < 102\epsilon$<br>$c_1 = (99/100)\cdot O(\epsilon^{-(2d+1)})$<br>$c_2 = (99/100)\cdot O(\epsilon^{-(2d+1)})$ | $\epsilon' = 102\epsilon$<br>$c_1' = 1/51$<br>$c_2' = \epsilon^{2d+1}/C(d)$ |

Adjusting the parameters $\epsilon > 0$ to be the same accross the three methods, we obtain the Table 3, which can be helpful to compare the results of the three proof methods:

**Table 3** A table summarizing the results on symmetric good pairs and large sets with small antigood sets characteristic of classes of VC dimension bounded by some fixed $d \in \mathbb{N}$. In this table, the parameter $\epsilon \in (0, 1/2)$ is scaled for. In the table below, $(c_1, c_2, \epsilon)$-good sampling stands for the following property: *For any $A \subseteq X, B \subseteq \mathcal{H}$ finite, $\exists A' \subseteq A, B' \subseteq B$, such that $|A'| \geq c_1|A|$, $|B'| \geq c_2|B|$, and $(A', B')$ forms a symmetric $\epsilon$ good pair.* The antigood set bound stands for the following: *For any $B \subseteq \mathcal{H}$, there is $B' \subseteq B$ with small $\epsilon'$-antigood set: concretely, $|B'| \geq c_2' \cdot |B|$,* $|\mathrm{Anti}(B', \epsilon)| \leq c_1'|A|$.

| | $(c_1, c_2, \epsilon)$-good sampling | Antigood set bound |
|---|---|---|
| Combinatorial | $c_1 = \epsilon^d / 2^{d+1}$<br><br>$c_2 = 1/2^{d+1}$ | $c_1' = 1 - (\frac{\epsilon}{2})^d$<br><br>$c_2' = 1/2^d$ |
| Haussler Packing | $c_1 = (1/2) \cdot \left(1 - \frac{\lvert B\rvert-1}{2\lvert B\rvert(1-\epsilon/2)}\right)$<br><br>$c_2 = (1/2) \cdot (\epsilon/120)^d$ | $c_1' = \max\left\{1, \frac{(\lvert B\rvert-1)\delta \cdot 2}{\lvert B\rvert\epsilon(1-\epsilon/2)}\right\}$<br><br>$c_2' = (\delta/30)^d$, $\delta > 0$ |
| VC Regularity | $c_1 = \frac{99^{-2d} \cdot O(\epsilon^{-(2d+1)})}{100^{-4d+1}}$<br><br>$c_2 = \frac{99^{-2d} \cdot O(\epsilon^{-(2d+1)})}{100^{-4d+1}}$ | $c_1' = 1/51$<br><br>$c_2' = \epsilon^{2d+1}/(102^{2d+1}C(d))$ |