SOME VC-DENSITY COMPUTATIONS IN SHELAH-SPENCER GRAPHS

ANTON BOBKOV

Abstract. We investigate vc-density in Shelah-Spencer graphs. We provide an upper bound on formula-by-formula basis and show that there isn't a uniform lower bound forcing the vc-function to be infinite.

VC-density was studied in [1] by Aschenbrenner, Dolich, Haskell, MacPherson, and Starchenko as a natural notion of dimension for NIP theories. In a complete NIP theory T we can define the vc-function

$$\mathrm{vc}^T = \mathrm{vc} : \mathbb{N} \longrightarrow \mathbb{R} \cup \{\infty\}$$

where $\mathrm{vc}(n)$ measures the worst-case complexity of families of definable sets in an n-fold Cartesian power of the underlying set of a model of T (see 1.13 below for a precise definition of vc^T). The simplest possible behavior is $\mathrm{vc}(n) = n$ for all n. Theories with the property that $\mathrm{vc}(1) = 1$ are known to be dp-minimal, i.e., having the smallest possible dp-rank. It is not known whether there can be a dp-minimal theory which doesn't satisfy $\mathrm{vc}(n) = n$ (see [1], diagram on pg. 41).

In this paper, we investigate vc-density of definable sets in Shelah-Spencer graphs. In our description of Shelah-Spencer graphs we follow closely the treatment in [2]. A Shelah-Spencer graph is a limit of random structures $G(n, n^{-\alpha})$ for an irrational $\alpha \in (0,1)$. $G(n, n^{-\alpha})$ is a random graph on n vertices with edge probability $n^{-\alpha}$.

Our first result is that in Shelah-Spencer graphs

$$\mathrm{vc}(n) = \infty$$

which implies that they are not dp-minimal. Our second result is providing an upper bound on a vc-density for a formula ϕ

$$\mathrm{vc}(\phi) \le K(\phi)\frac{Y(\phi)}{\epsilon(\phi)}$$

where $K(\phi), Y(\phi), \epsilon(\phi)$ are paramters easily computable from the quantifier free form of ϕ.

Chapter 1 introduces basic facts about VC-dimension and vc-density. More can be found in [1]. Chapter 2 summarizes notation and basic facts concerning Shelah-Spencer graphs. We direct the reader to [2] for a more in-depth treatment. In chapter 3 we introduce some measure of dimension for quantifier free formulas as well as proving some elementary facts about it. Chapter 4 computes a lower bound for vc-density to demonstrate that $\mathrm{vc}(n) = \infty$. Chapter 5 computes an upper bound for vc-density on a formula-by-formula basis.

1. VC-dimension and vc-density

Throughout this section we work with a collection $\mathcal{F}$ of subsets of an infinite set X. We call the pair $(X, \mathcal{F})$ a set system.

Definition 1.1.

- Given a subset A of X, we define the set system $(A, A \cap \mathcal{F})$ where $A \cap \mathcal{F} = \{A \cap F \mid F \in \mathcal{F}\}$.
- For $A \subseteq X$ we say that $\mathcal{F}$ shatters A if $A \cap \mathcal{F} = \mathcal{P}(A)$ (the power set of A).

Definition 1.2. We say $(X, \mathcal{F})$ has VC-dimension n if the largest subset of X shattered by $\mathcal{F}$ is of size n. If $\mathcal{F}$ shatters arbitrarily large subsets of X, we say that $(X, \mathcal{F})$ has infinite VC-dimension. We denote the VC-dimension of $(X, \mathcal{F})$ by $\mathrm{VC}(X, \mathcal{F})$.

Note 1.3. We may drop X from the notation $\mathrm{VC}(X, \mathcal{F})$, as the VC-dimension doesn't depend on the base set and is determined by $(\bigcup \mathcal{F}, \mathcal{F})$.

Set systems of finite VC-dimension tend to have good combinatorial properties, and we consider set systems with infinite VC-dimension to be poorly behaved.

Another natural combinatorial notion is that of the dual system of a set system:

Definition 1.4. For $a \in X$ define $X_a = \{F \in \mathcal{F} \mid a \in F\}$. Let $\mathcal{F}^* = \{X_a \mid a \in X\}$. We call $(\mathcal{F}, \mathcal{F}^*)$ the dual system of $(X, \mathcal{F})$. The VC-dimension of the dual system of $(X, \mathcal{F})$ is referred to as the dual VC-dimension of $(X, \mathcal{F})$ and denoted by $\mathrm{VC}^*(\mathcal{F})$. (As before, this notion doesn't depend on X.)

Lemma 1.5 (see 2.13b in [3])**.** *A set system $(X, \mathcal{F})$ has finite VC-dimension if and only if its dual system has finite VC-dimension. More precisely*

$$\mathrm{VC}^*(\mathcal{F}) \leq 2^{1+\mathrm{VC}(\mathcal{F})}.$$

For a more refined notion of complexity of $(X, \mathcal{F})$ we look at the traces of our family on finite sets:

Definition 1.6. Define the shatter function $\pi_{\mathcal{F}} \colon \mathbb{N} \longrightarrow \mathbb{N}$ of $\mathcal{F}$ and the dual shatter function $\pi^*_{\mathcal{F}} \colon \mathbb{N} \longrightarrow \mathbb{N}$ of $\mathcal{F}$ by

$$\pi_{\mathcal{F}}(n) = \max\{|A \cap \mathcal{F}| \mid A \subseteq X \text{ and } |A| = n\}$$

$$\pi^*_{\mathcal{F}}(n) = \max\{\mathrm{atoms}(B) \mid B \subseteq \mathcal{F}, |B| = n\}$$

where $\mathrm{atoms}(B)$ = number of atoms in the boolean algebra of sets generated by B. Note that the dual shatter function is precisely the shatter function of the dual system: $\pi^*_{\mathcal{F}} = \pi_{\mathcal{F}^*}$.

A simple upper bound is $\pi_{\mathcal{F}}(n) \leq 2^n$ (same for the dual). If the VC-dimension of $\mathcal{F}$ is infinite then clearly $\pi_{\mathcal{F}}(n) = 2^n$ for all n. Conversely we have the following remarkable fact:

Theorem 1.7 (Sauer-Shelah '72, see [5], [6])**.** *If the set system $(X, \mathcal{F})$ has finite VC-dimension d then $\pi_{\mathcal{F}}(n) \leq \binom{n}{\leq d}$ for all n, where $\binom{n}{\leq d} = \binom{n}{d} + \binom{n}{d-1} + \ldots + \binom{n}{1}$.*

Thus the systems with a finite VC-dimension are precisely the systems where the shatter function grows polynomially. The vc-density of $\mathcal{F}$ quantifies the growth of the shatter function of $\mathcal{F}$:

Definition 1.8. Define the vc-density and dual vc-density of $\mathcal{F}$ as

$$\mathrm{vc}(\mathcal{F}) = \limsup_{n\to\infty} \frac{\log \pi_{\mathcal{F}}(n)}{\log n} \in \mathbb{R}^{\geq 0} \cup \{+\infty\},$$
$$\mathrm{vc}^*(\mathcal{F}) = \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}}(n)}{\log n} \in \mathbb{R}^{\geq 0} \cup \{+\infty\}.$$

Generally speaking a shatter function that is bounded by a polynomial doesn't itself have to be a polynomial. Proposition 4.12 in [1] gives an example of a shatter function that grows like $n \log n$ (so it has vc-density 1).

So far the notions that we have defined are purely combinatorial. We now adapt VC-dimension and vc-density to the model theoretic context.

Definition 1.9. Work in a first-order structure M. Fix a finite collection of formulas $\Phi(x, y)$ in the language $\mathcal{L}(M)$ of M.

- For $\phi(x, y) \in \mathcal{L}(M)$ and $b \in M^{|y|}$ let

$$\phi(M^{|x|}, b) = \{a \in M^{|x|} \mid \phi(a, b)\} \subseteq M^{|x|}.$$

- Let $\Phi(M^{|x|}, M^{|y|}) = \{\phi(M^{|x|}, b) \mid \phi_i \in \Phi, b \in M^{|y|}\} \subseteq \mathcal{P}(M^{|x|})$.
- Let $\mathcal{F}_\Phi = \Phi(M^{|x|}, M^{|y|})$, giving rise to a set system $(M^{|x|}, \mathcal{F}_\Phi)$.
- Define the VC-dimension $\mathrm{VC}(\Phi)$ of Φ to be the VC-dimension of $(M^{|x|}, \mathcal{F}_\Phi)$, similarly for the dual.
- Define the vc-density $\mathrm{vc}(\Phi)$ of Φ to be the vc-density of $(M^{|x|}, \mathcal{F}_\Phi)$, similarly for the dual.

We will also refer to the vc-density and VC-dimension of a single formula ϕ viewing it as a one element collection $\Phi = \{\phi\}$.

Counting atoms of a boolean algebra in a model theoretic setting corresponds to counting types, so it is instructive to rewrite the shatter function in terms of types.

Definition 1.10.

$$\pi^*_\Phi(n) = \max\left\{\text{number of } \Phi\text{-types over } B \mid B \subseteq M, |B| = n\right\}.$$

Here a Φ-type over B is a maximal consistent collection of formulas of the form $\phi(x, b)$ or $\neg\phi(x, b)$ where $\phi \in \Phi$ and $b \in B$.

The functions π^*_Φ and $\pi^*_{\mathcal{F}_\Phi}$ do not have to agree, as one fixes the number of generators of a boolean algebra of sets and the other fixes the size of the parameter set. However, as the following lemma demonstrates, they both give the same asymptotic definition of dual vc-density.

Lemma 1.11.

$$\mathrm{vc}^*(\Phi) = \textit{degree of polynomial growth of } \pi^*_\Phi(n) = \limsup_{n\to\infty} \frac{\log \pi^*_\Phi(n)}{\log n}.$$

Proof. With a parameter set B of size n, we get at most $|\Phi|n$ sets $\phi(M^{|x|}, b)$ with $\phi \in \Phi, b \in B$. We check that asymptotically it doesn't matter whether we look at growth of boolean algebra of sets generated by n or by $|\Phi|n$ many sets. We have:

$$\pi^*_{\mathcal{F}_\Phi}(n) \le \pi^*_\Phi(n) \le \pi^*_{\mathcal{F}_\Phi}(|\Phi|n).$$

Hence:

$$\begin{aligned}
\mathrm{vc}^*(\Phi) &\le \limsup_{n\to\infty} \frac{\log \pi^*_\Phi(n)}{\log n} \le \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}_\Phi}(|\Phi|n)}{\log n} = \\
&= \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}_\Phi}(|\Phi|n)}{\log |\Phi|n} \frac{\log |\Phi|n}{\log n} = \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}_\Phi}(|\Phi|n)}{\log |\Phi|n} \le \\
&\le \limsup_{n\to\infty} \frac{\log \pi^*_{\mathcal{F}_\Phi}(n)}{\log n} = \mathrm{vc}^*(\Phi).
\end{aligned}$$

$\square$

One can check that the shatter function and hence VC-dimension and vc-density of a formula are elementary notions, so they only depend on the first-order theory of the structure M.

NIP theories are a natural context for studying vc-density. In fact we can take the following as the definition of NIP:

Definition 1.12. Define ϕ to be NIP if it has finite VC-dimension in a theory T. A theory T is NIP if all the formulas in T are NIP.

In a general combinatorial context (for arbitrary set systems), vc-density can be any real number in $0 \cup [1, \infty)$ (see [4]). Less is known if we restrict our attention to NIP theories. Proposition 4.6 in [1] gives examples of formulas that have non-integer rational vc-density in an NIP theory, however it is open whether one can get an irrational vc-density in this model-theoretic setting.

Instead of working with a theory formula by formula, we can look for a uniform bound for all formulas:

Definition 1.13. For a given NIP structure M, define the vc-function

$$\begin{aligned} \mathrm{vc}^M(n) &= \sup\{\mathrm{vc}^*(\phi(x,y)) \mid \phi \in \mathcal{L}(M), |x| = n\} \\ &= \sup\{\mathrm{vc}(\phi(x,y)) \mid \phi \in \mathcal{L}(M), |y| = n\} \in \mathbb{R}^{\geq 0} \cup \{+\infty\}\,. \end{aligned}$$

As before this definition is elementary, so it only depends on the theory of M. We omit the superscript M if it is understood from the context. One can easily check the following bounds:

Lemma 1.14 (Lemma 3.22 in [1])**.** *We have* $\mathrm{vc}(1) \geq 1$ *and* $\mathrm{vc}(n) \geq n\,\mathrm{vc}(1)$.

However, it is not known whether the second inequality can be strict or even whether $\mathrm{vc}(1) < \infty$ implies $\mathrm{vc}(n) < \infty$.

2. Graph Combinatorics

Throughout this paper A, B, C, M will denote finite graphs, and $\mathbb{D}$ will be used to denote potentially infinite graphs. For a graph A the set of its vertices is denoted by $v(A)$, and the set of its edges by $e(A)$. Number of vertices of A will be denoted as $|A|$. Subgraph always means induced subgraph and $A \subseteq B$ means that A is a subgraph of B. For two subgraphs A, B of a larger graph, the union $A \cup B$ denotes the graph induced by $v(A) \cup v(B)$. Similarly, $A - B$ means a subgraph of A induced by the vertices of $v(A) - v(B)$. For $A \subseteq B \subseteq D$ and $A \subseteq C \subseteq D$, graphs B, C are said to be disjoint over A if $v(B) - v(A)$ is disjoint from $v(C) - v(A)$ and there are no edges from $v(B) - v(A)$ to $v(C) - v(A)$ in D.

For the remainder of the paper fix $\alpha \in (0, 1)$, irrational.

Definition 2.1.

- For a graph A let $\dim(A) = |A| - \alpha|e(A)|$.
- For A, B with $A \subseteq B$ define $\dim(B/A) = \dim(B) - \dim(A)$.
- We say that $A \leq B$ if $A \subseteq B$ and $\dim(A'/A) > 0$ for all $A \subsetneq A' \subseteq B$.
- Define A to be positive if for all $A' \subseteq A$ we have $\dim(A') \geq 0$.
- We work in theory S_α in the language of graphs axiomatized by:
  - Every finite substructure is positive.
  - Given a model $\mathbb{G}$ and graphs $A \leq B$, every embedding $f : A \longrightarrow \mathbb{G}$ extends to an embedding $g : B \longrightarrow \mathbb{G}$.

  (Here an embedding maps edges to edges and nonedges to nonedges.) This theory is complete and stable (see 5.7 and 7.1 in [2]). From now on fix an ambient model $\mathbb{G} \models S_\alpha$. This will be the only infinite graph we work with.
- For A, B positive, (A, B) is called a minimal pair if $A \subseteq B$, $\dim(B/A) < 0$ but $\dim(A'/A) \geq 0$ for all proper $A \subseteq A' \subsetneq B$. We call B a minimal extension of A. The dimension of a minimal pair is defined as $|\dim(B/A)|$.
- A sequence $\langle M_i \rangle_{0 \leq i \leq n}$ is called a minimal chain if (M_i, M_{i+1}) is a minimal pair for all $0 \leq i < n$.
- For a graph A with the tuple of vertices x let $\mathrm{diag}_A(x)$ be the atomic diagram of A, i.e. the first-order formula recording whether there is an edge between every pair of vertices.
- Given $A \subseteq B$ let

$$\phi_{A,B}(x) = \mathrm{diag}_A(x) \wedge \exists z \ \mathrm{diag}_B(x, z).$$

Any graph isomorphic to B is called a witness of $\phi_{A,B}$.

- A formula $\phi_{A,B}$ is called a basic formula if there is a minimal chain $\langle M_i \rangle_{0 \leq i \leq n}$ such that $A = M_0$ and $B = M_n$.

Theorem 2.2 (Quantifier elimination, 5.6 in [2]). *In theory S_α every formula is equivalent to a boolean combination of basic formulas.*

Definition 2.3. A graph $S \subseteq \mathbb{D}$ is called N-strong if for any $S \subseteq T \subseteq D$ with $|T| - |S| \leq N$ we have $S \leq T$.

3. Basic Definitions and Lemmas

Definition 3.1. Suppose $\phi(x, y)$ is a basic formula. Define X to be the graph on vertices x with edges defined by ϕ. Similarly define Y. Note that X, Y are positive. Additionally, let Y' be a subgraph of Y induced by vertices of Y that are connected to $W - (X \cup Y)$, where W is a witness of ϕ.

We will require the following lemmas from [2]:

Lemma 3.2. *[See 2.3 in [2]] Let $A, B \subseteq \mathbb{D}$. Then*

$$\dim(A \cup B/A) \leq \dim(B/A \cap B).$$

Moreover,

$$\dim(A \cup B/A) = \dim(B/A \cap B) - \alpha E,$$

where E is the number of edges connecting the vertices of $A \cup B - A$ to the vertices of $A - A \cap B$.

Lemma 3.3. *[See 4.1 in [2]] Suppose A is a positive graph, with at least $1/\alpha + 2$ vertices. Then for any $\epsilon > 0$ there exists a graph B such that (A, B) is a minimal pair with dimension $\leq \epsilon$. Moreover, every vertex in A is connected to a vertex in $B - A$.*

Lemma 3.4. *[See 4.4 in [2]] Suppose A is a positive graph, and $\mathcal{G}$ a model of S_α. Then for any integer S there exists an embedding $f\colon A \longrightarrow \mathcal{G}$ such that $f(A)$ is S-strong in $\mathcal{G}$.*

Lemma 3.5. *[See 3.8 in [2]] For all $S > 0$ there exists $M = M(S, \alpha) \in \mathbb{N}$ with the following property. Suppose $A \subseteq \mathcal{G}$ where $\mathcal{G}$ is a model of S_α. Then there exists B with $A \subseteq B \subseteq \mathcal{G}$ such that B is S-strong in $\mathbb{G}$ and $|B| \leq M|A|$.*

We conclude this section by stating a couple of technical lemmas that will be useful in our proofs later.

Lemma 3.6. *Work in an ambient graph $\mathbb{D}$. Suppose we have a set B and a minimal pair (A, M) with $A \subseteq B$ and $\dim(M/A) = -\epsilon$. Then either $M \subseteq B$ or $\dim(M \cup B/B) < -\epsilon$.*

Proof. By Lemma 3.2

$$\dim(M \cup B/B) \leq \dim(M/M \cap B),$$

and as $A \subseteq M \cap B \subseteq M$

$$\dim(M/A) = \dim(M/M \cap B) + \dim(M \cap B/A).$$

In addition we are given $\dim(M/A) = -\epsilon$. If $M \not\subseteq B$ then $A \subseteq M \cap B \subsetneq M$ and by minimality $\dim(M \cap B/A) > 0$. Combining the inequalities above we obtain the desired result:

$$\dim(M \cup B/B) \leq \dim(M/M \cap B) = \dim(M/A) - \dim(M \cap B/A) < -\epsilon.$$

□

Lemma 3.7. *Work in an ambient graph $\mathbb{D}$. Suppose we have a set B and a minimal chain $\langle M_i \rangle_{0 \leq i \leq n}$ with dimensions*

$$\dim(M_{i+1}/M_i) = -\epsilon_i.$$

Let $\epsilon = \min_{0 \le i \le n} \epsilon_i$*. Then either* $M_n \subseteq B$ *or* $\dim((M_n \cup B)/B) < -\epsilon$*.*

Proof. Let $\bar{M}_i = M_i \cup B$. Then:

$$\dim(\bar{M}_n/B) = \dim(\bar{M}_n/\bar{M}_{n-1}) + \ldots + \dim(\bar{M}_2/\bar{M}_1) + \dim(\bar{M}_1/B).$$

Either $M_n \subseteq B$ or at least one of the summands above is nonzero. Apply previous lemma. □

Lemma 3.8. *Suppose we have a minimal pair* (A, M) *with dimension* ϵ*. Suppose we have some* $B \subseteq M$*. Then* $\dim B/(A \cap B) \ge -\epsilon$*. Moreover if* $B \cup A \ne M$ *then* $\dim B/(A \cap B) \ge 0$*.*

Proof. We have $\dim(B \cup A/A) \le \dim B/(A \cap B)$ by Lemma 3.2. As $A \subseteq B \cup A \subseteq M$ we have $\dim(B \cup A/A) \ge -\epsilon$ by minimality. Moreover, minimality implies that it is positive if $B \cup A \ne M$. □

Lemma 3.9. *Suppose we have a minimal chain* $\langle M_i \rangle_{0 \le i \le n}$ *with dimensions*

$$\dim(M_{i+1}/M_i) = -\epsilon_i.$$

Let ϵ *be the sum of all* ϵ_i*. Suppose we have a graph* B *with* $B \subseteq M_n$*. Then* $\dim B/(M_0 \cap B) \ge -\epsilon$*.*

Proof. Let $B_i = B \cap M_i$. We have $\dim B_{i+1}/B_i \ge \dim M_{i+1}/M_i$ by the previous lemma. Thus

$$\dim B/(M_0 \cap B) = \dim B_n/B_0 = \sum \dim B_{i+1}/B_i \ge -\epsilon.$$

□

4. Lower bound

In this section we restrict our attention to the following family of basic formulas $\phi(x, y)$:

- All formulas have $Y' = Y$ (see Definition 3.1).

- All formulas define no edges between X and Y.
- Minimal chain of $\phi(x, y)$ consists of one step, that is we only have one minimal extension as opposed to a chain of minimal extensions.
- The dimension of that minimal extension is smaller than α.

We obtain a lower bound for the formulas that are boolean combinations of basic formulas written in the disjunctive-conjunctive form. First, define $\epsilon_L(\phi)$.

Definition 4.1. For a basic formula $\phi = \phi_{\langle M_i \rangle_{0 \leq i \leq n}}(x, y)$ let

- $\epsilon_i(\phi) = -\dim(M_i/M_{i-1})$.
- $\epsilon_L(\phi) = \sum_1^n \epsilon_i(\phi)$.

Definition 4.2 (Negation)**.** If ϕ is a basic formula, then define

$$\epsilon_L(\neg\phi) = \epsilon_L(\phi).$$

Definition 4.3 (Conjunction)**.** Take a collection of formulas $\phi_i(x, y)$ where each ϕ_i is a positive or a negative basic formula. If both positive and negative formulas are present then $\epsilon_L(\phi) = \infty$. We don't have a lower bound for that case. If different formulas define X or Y differently then $\epsilon_L(\phi) = \infty$. In the case of conflicting definitions the formula would have no realizations. Otherwise let

$$\epsilon_L\left(\bigwedge \phi_i\right) = \sum \epsilon_L(\phi_i).$$

Definition 4.4 (Disjunction)**.** Take a collection of formulas ψ_i where each instance is a conjunction as above all agreing on X and Y. Then

$$\epsilon_L\left(\bigvee \psi_i\right) = \min \epsilon_L(\psi_i).$$

Theorem 4.5. *For a formula ψ as above we have*

$$\operatorname{vc}\psi \geq \left\lfloor \frac{Y(\psi)}{\epsilon_L(\psi)} \right\rfloor,$$

where $Y(\psi)$ is $\dim(Y)$ (as all basic componenets agree on Y).

Proof. First, work with a formula that is a conjunction of positive basic formulas $\psi = \bigwedge_{i \in I} \phi_i$. Then as we have defined above

$$\epsilon_L(\psi) = \sum_{i \in I} \epsilon_L(\phi_i).$$

If W_i is a witness of ϕ_i, let $S_i = |W_i|$. Let n_1 be the largest natural number such that

$$n_1 \epsilon_L(\psi) < Y(\psi).$$

Let ϵ' be the smallest value among $\epsilon_L(\phi_i)$. Suppose it corresponds to the formula ϕ'. Let n_2 be the largest natural number such that

$$n_1 \epsilon_L(\psi) + n_2 \epsilon' < Y(\psi).$$

Fix some $N > n_1 + n_2$. Let

$$J = \{0 \leq j < N\} \subseteq \mathbb{N}.$$

Let a_j be a graph isomorphic to X for each $j \in J$, pairwise disjoint. Let $A = \bigcup_{1 \leq j \leq N} a_j$. Let

$$S = |Y| + (n_1 + n_2 + 1) \sum_{i \in I} S_i.$$

By Lemma 3.4 the graph A can be embedded into $\mathbb{G}$ as an S-strong graph. Abusing notation, we identify A with this embedding. Thus we have $A \subseteq \mathbb{G}$, S-strong.

Let J_1, J_2 be disjoint subsets of J, of sizes n_1, n_2 respectively. Let b be a graph isomorphic to Y. For each $i \in I, j \in J_1$ let W_{ij} be a witness of $\phi_i(a_j, b)$. (Note that then $(a_j \cup b, W_{ij})$ is a minimal pair.) For each $j \in J_1$ let W_j be a union of $\{W_{ij}\}_{i \in I}$ disjoint over $a_j \cup b$. For each $j \in J_2$ let W_j be a witness of $\phi'(a_j, b)$. Let W' be a union of $\{W_j\}_{j \in J_1 \cup J_2}$ disjoint over b. Let W be a union of W' and A disjoint over $\{a_j\}_{j \in J_1 \cup J_2}$.

Claim 4.6. *We have* $A \leq W$.

Proof. Consider some $A \subsetneq B \subseteq W$. We need to show $\dim(B/A) > 0$. Let $\bar{A} = A \cup b$. We have

$$\dim(B/A) = \dim(B/B \cap \bar{A}) + \dim(B \cap \bar{A}/A).$$

Let $B_{ij} = B \cap W_{ij}$. Let $B_j = B \cap W_j$. To unify indices, relabel all the graphs above as $\{B_k\}_{k \in K}$ for some index set K. By the construction of W we have

$$\dim(B/B \cap \bar{A}) = \sum_{k \in K} \dim(B_k/B_k \cap \bar{A}).$$

Fix k. We have $B_k \subseteq W_k$, where W_k is a minimal extension of $M_0^k = a \cup b$ for some $a \in A$. Let ϵ_k be the dimension of this minimal extension. We have $\dim(B_k/B_k \cap \bar{A}) = \dim(B_k/a \cup (B \cap b))$.

Case 1: $B \cap b = b$. Then $M_0^k \subseteq B_k \subseteq W_k$ and

$$\dim(B_k/a \cup (B \cap b)) = \dim(B_k/M_0^k).$$

By minimality of (M_0^k, B_k) we have $\dim(B_k/M_0^k) \geq -\epsilon_k$. Thus

$$\dim(B/B \cap \bar{A}) \geq -\sum_{k \in K} \epsilon_k = -\left(n_1 \epsilon_L(\psi) + n_2 \epsilon'\right).$$

In addition

$$\dim(B \cap \bar{A}/A) = \dim(b) = Y(\psi).$$

Combining the two, we get

$$\dim(B/A) \geq Y(\psi) - \left(n_1 \epsilon_L(\psi) + n_2 \epsilon'\right),$$

which is positive by the construction of n_1, n_2 as needed.

Case 2: $B \cap b \subsetneq b$.

Claim 4.7. *We have* $\dim(B_k/B_k \cap \bar{A}) > 0$.

Proof. Recall that $\dim(B_k/B_k \cap \bar{A}) = \dim(B_k/a \cup (B \cap b))$. First, suppose that $B_k \cup M_0^k \neq W_k$. Then by Lemma 3.8 we get the required inequality. Thus we may assume that $B_k \cup M_0^k = W_k$. By Lemma 3.2 we have

$$\dim(B_k \cup M_0^k/M_0^k) = \dim(B_k/B_k \cap M_0^k) - \alpha E,$$

where E is the number of edges connecting the vertices of $B_k \cup M_0^k - M_0^k$ to the vertices of $M_0^k - B_k \cap M_0^k$. Noting that $B_k \cup M_0^k = W_k$, $\dim W_k/M_0^k = -\epsilon_k$, and $B_k \cap M_0^k = a \cup (B \cap b)$ we may rewrite the equality above as

$$\dim(B_k/a \cup (B \cap b)) = \alpha E - \epsilon,$$

and E is the number of edges connecting the vertices of $W_k - M_0^k$ to the vertices of $M_0^k - a \cup (B \cap b)$. As $Y = Y'$ and $B \cap b \subsetneq b$ we must have $E \geq 1$. But then as $\alpha > \epsilon$ we have $\dim(B_k/a \cup (B \cap b)) > 0$ as needed. $\square$

Now, recall that

$$\dim(B/A) = \dim(B \cap \bar{A}/A) + \sum_{k \in K} \dim(B_k/B_k \cap \bar{A}).$$

By the claim above each of $\dim(B_k/B_k \cap \bar{A}) > 0$, thus

$$\dim(B/A) > \dim(B \cap \bar{A}/A).$$

In addition

$$\dim(B \cap \bar{A}/A) = \dim(B \cap b) \geq 0,$$

as b is postive. Thus $\dim(B/A) > 0$ as needed. $\square$

As $A \leq W$ and $A \subseteq \mathbb{G}$, we can embed W into $\mathbb{G}$ over A. Abusing notation again, we identify W with its embedding $A \leq W \subseteq \mathbb{G}$. In particular, now we have $b \in \mathbb{G}$.

Also note that

$$\dim(W/A) = Y(\psi) - (n_1\epsilon_L(\psi) + n_2\epsilon')\,,$$
$$|W| - |A| \leq |b| + (n_1+n_2)\sum_{i\in I} S_i.$$

Lemma 4.8. *We have*

$$\{a_j\}_{j\in J_1} \subseteq \psi(A,b) \subseteq \{a_j\}_{j\in J_1\cup J_2}\,.$$

Proof. First inclusion $\{a_j\}_{j\in J_1} \subseteq \psi(A,b)$ is immediate from the construction of W, as W_{ij} witnesses that $\phi_i(a_j,b)$ holds. For the second inclusion, suppose that there is $a \in A - \{a_j\}_{j\in J_1\cup J_2}$ such that $\psi(a,b)$ holds. Let $W' \subseteq \mathbb{G}$ be a witness of $\phi_1(a,b)$. First, note that the case $W' \subseteq W$ is impossible as there are no edges between a and $W - a$, but there are edges between a and $W' - a$. Thus assume $W' \nsubseteq W$. As $(a\cup b, W')$ is minimal, by Lemma 3.6 we have $\dim(W'\cup W/W) < -\epsilon_1$. Therefore

$$\dim(W'\cup W/A) = \dim(W'\cup W/W) + \dim(W/A) < Y(\psi) - (n_1\epsilon_L(\psi) + n_2\epsilon') - \epsilon_1,$$

which is negative by the construction of n_1, n_2. Thus $A \nleq W\cup W'$, as then it would have a positive dimension. Additionally,

$$|W'\cup W| - |A| \leq |W' - W| + |W| - |A| \leq S_1 + |b| + (n_1+n_2)\sum_{i\in I} S_i \leq S,$$

but then this contradicts that A is S-strong, as then we would have $A \leq W\cup W'$. $\square$

In the construction of W we have chosen indices J_1, J_2 arbitrarily. In particular, suppose we let J_2 to be the last n_2 indices of J and J_1 an arbitrary n_1-element subset of the first $N - n_2$ elements of J. Each of those choices would then yield a different trace $\psi(A,b)$ by the lemma above. Thus $\psi(A, M^{|y|}) \geq \binom{N-n_2}{n_1}$ and therefore $\mathrm{vc}(\psi) \geq n_1$. By the definition of n_1 we have $n_1 = \left\lfloor \frac{Y(\psi)}{\epsilon_L(\psi)} \right\rfloor$, so this proves the theorem for ψ.

Now consider a formula which is a conjunction consisting of negative basic formulas $\psi = \bigwedge_{i\in I} \neg\phi_i$. Let $\bar\psi = \bigwedge_{i\in I} \phi_i$. Do the construction above for $\bar\psi$ and suppose its trace is $X \subseteq A$ for some b. Then over b the same construction gives trace $(A - X)$ for ψ. Thus we get as many traces as above, and the same bound.

Finally consider a formula which is a disjunction of formulas considered above $\theta = \bigvee_{k \in K} \psi_k$. Choose the one with the smallest ϵ_L, say ψ_k, and repeat the construction above for ψ_k. Any trace we obtain is automatically a trace for θ, and thus we get as many traces as above, and the same bound. □

Corollary 4.9. *VC-function is infinite in Shelah-Spencer random graphs:*

$$\mathrm{vc}(n) = \infty.$$

Proof. Let A be a graph consisting of $1/\alpha + 2 + n$ disconnected vertices. Fix $\epsilon > 0$. By Lemma 3.3, there exists B such that (A, B) is minimal with dimension $\leq \epsilon$. Consider a basic formula $\psi_{A,B}(x, y)$ where $|x| = 1/\alpha + 2$ and $|y| = n$. Then by the theorem above $\mathrm{vc}(n) \geq \mathrm{vc}(\psi_{A,B}) \geq \frac{n}{\epsilon}$. As ϵ was arbitrary, this number can be made arbitrarily large, giving $vc(n) = \infty$ as needed. □

5. Upper bound

We bound the number of types of some finite collection of formulas $\Psi(\vec{x}, \vec{y}) = \{\phi_i(\vec{x}, \vec{y})\}_{i \in I}$ over a parameter set B of size N, where ϕ_i is a basic formula.

Fix a formula ϕ from our collection. Suppose it defines a minimal chain extension over $\{x, y\}$. Record the size of that extension as $K(\phi)$ and its total dimension $\epsilon(\phi) = \epsilon_U(\phi)$. Define dimension of that formula $D(\phi) = |\vec{y}| \frac{K(\phi)}{\epsilon(\phi)}$ Define dimension of the entire collection as $D(\Psi) = \max_{i \in I} D(\phi_i)$

Fix $S =$??. Suppose we have a finite parameter set $A_0 \subseteq \mathbb{G}^{|x|}$ with $|A_0| = N_0$. We would like to bound $\phi(A_0, \mathbb{G}^{|y|})$. Let $A_1 \subseteq \mathbb{G}$ consist of the components of the elements of A_0. Then $|A_1| \leq |x| N_0$. Using Lemma 3.5 let A be a graph $A_0 \subseteq A \subseteq \mathbb{G}$, S-strong in $\mathbb{G}$. Let $N = |A|$. We have $N \leq |x| N_0 M$ (where M is the constant from the Lemma 3.5). As $A_0 \subseteq A^{|x|}$ we have

Definition 5.1.
- For all $a \in A^{|x|}, b \in \mathbb{G}_2$ if $\phi(a,b)$ holds fix $W_{a,b} \subseteq \mathbb{G}$, a witness of this formula.
- For $b \in \mathbb{G}_2$ let

$$W_b = \bigcup W_{a,b} \mid a \in A^{|x|}, \mathbb{G} \models \phi(a,b).$$

Definition 5.2. For sets $C, B \subset \mathbb{G}$ define the boundary of C over B

$$\partial(C,B) = \{b \in B \mid \text{there is an edge between } b \text{ and a vertex in } C - B\}$$

Definition 5.3.
- For $b \in \mathbb{G}_2$ let ∂_b to be the boundary $\partial(W_b, A)$.
- For $b \in \mathbb{G}_2$ let $\bar{W}_b = (W_b - A) \cup \partial_b$.
- For $b_1, b_2 \in \mathbb{G}_2$ we say that $b_1 \sim b_2$ if $\partial_{b_1} = \partial_{b_2}$ and there exists a graph isomorphism from $\bar{W}_{b_1}$ to $\bar{W}_{b_2}$ that fixes ∂_{b_1} and maps b_1 to b_2. One easily checks that this defines an equivalence relation.
- For $b \in \mathbb{G}_2$ define $\mathscr{I}_b$ to be the $\sim$-equivalence class of b.

Lemma 5.4. *For $b_1, b_2 \in \mathbb{G}_2$ if $b_1 \sim b_2$ then $\phi(A^{|x|}, b_1) = \phi(A^{|x|}, b_2)$.*

Proof. Fix the graph isomorphism $\bar{f} \colon \bar{W}_{b_1} \longrightarrow \bar{W}_{b_2}$. Extend it to an isomorphism $f \colon W_{b_1} \cup A \longrightarrow W_{b_2} \cup A$ by being an identity map on the new vertices. Suppose $\mathbb{G} \models phi(a, b_1)$ for some $a \in A^{|x|}$. Then $f(W_{a,b_1})$ is a witness for $\phi(a, b_2)$ (though not necessarily equal to W_{a,b_2}) and thus $\mathbb{G} \models phi(a, b_2)$. As a was arbitrary, this proves the equality of the traces. $\square$

Thus to bound the number of traces it is sufficient to bound the number of possibilities for $\mathscr{I}_a$.

Theorem 5.5.

$$|\partial_a| \leq D(\Psi)$$

$$|\bar{M}_b - \bar{A}| \leq D(\Psi)$$

Corollary 5.6.

$$\mathrm{vc}(\phi) \le K(\phi)\frac{Y(\phi)}{\epsilon(\phi)}$$

Proof. We count possible ∂-isomorphism classes $\mathscr{I}_b$. Let $W = K(\phi)\frac{Y(\phi)}{\epsilon(\phi)}$. If the parameter set A is of size N then there are $\binom{N}{W}$ choices for boundary ∂_b. On top of the boundary there are at most W extra vertices and $(2W)^2$ extra edges. Thus there are at most

$$W \cdot 2^{(2W)^2}$$

configurations up to a graph isomorphism. In total this gives us

$$\binom{N}{W} \cdot W \cdot 2^{(2W)^2} = O(N^W)$$

options for ∂-isomorphism classes. By Lemma 5.4 there are at most $O(N^W)$ many traces, giving the required bound. □

Proof. (of Theorem 5.5) Fix some b-trace A_b. Enumerate $A_b = \{a_1, \ldots, a_I\}$.

Let $M_i/\{a_i, b\}$ be a witness of $\phi(a_i, b)$ for each $i \le I$. Let $\bar{M}_i = \bigcup_{j<i} M_j$. Let $\bar{M} = \bigcup M_i$, a witness of A_b

Claim 5.7.

$$\left|\partial(M_i M, \bar{A}) - \partial(M, \bar{A})\right| \le |M_i| = K(\phi)$$

$$\dim(M_i M\bar{A}/M\bar{A}) > -\epsilon(\phi)$$

Definition 5.8. $(j-1, j)$ is called a jump if some of the following conditions happen

- New vertices are added outside of $\bar{A}$ i.e.

$$\bar{M}_j - \bar{A} \ne \bar{M}_{j-1} - \bar{A}$$

- New vertices are added to the boundary, i.e.

$$\partial(\bar{M}_j, \bar{A}) \neq \partial(\bar{M}_{j-1}, \bar{A})$$

Definition 5.9. We now let m_i count all jumps below i

$$m_i = |\{j < i \mid (j-1, j) \text{ is a jump}\}|$$

Lemma 5.10.

$$\dim(\bar{M}_i/\bar{A}) \leq -m_i \cdot \epsilon(\phi)$$

$$|\partial(\bar{M}_i, \bar{A})| \leq m_i \cdot K(\phi)$$

$$|\bar{M}_j - \bar{A}| \leq m_i \cdot K(\phi)$$

Proof. (of Lemma 5.10) Proceed by induction. Second and third propositions are clear. For the first proposition base case is clear.

Induction step. Suppose $\bar{M}_j \cap (A \cup b) = \bar{M}_{j+1}$ and $\partial(\bar{M}_j, A) = \partial(\bar{M}_{j+1}, A)$. Then $m_i = m_{i+1}$ and the quantities don't change. Thus assume at least one of these equalities fails.

Apply Lemma 3.7 to $\bar{M}_j \cup (A \cup b)$ and $(M_{j+1}, a_{j+1}b)$. There are two options

- $\dim(\bar{M}_{j+1} \cup (A \cup b)/\bar{M}_i \cup (A \cup b)) \leq -\epsilon_U$. This implies the proposition.
- $M_{j+1} \subseteq \bar{M}_j \cup (A \cup b)$. Then by our assumption it has to be $\partial(\bar{M}_j, A) \neq \partial(\bar{M}_{j+1}, A)$. There are edges between $M_{j+1} \cap (\partial(\bar{M}_{j+1}, A) - \partial(\bar{M}_j, A))$ so they contribute some negative dimension $\leq \epsilon_U$.

This ends the proof for Lemma 5.10. □

(Proof of Theorem 5.5 continued) First part of lemma 5.10 implies that we have $\dim(\bar{M}/\bar{A}) \leq -m_I \cdot \epsilon(\phi)$. The requirement of A to be S-strong forces

$$m_I \cdot \epsilon(\phi) < Y(\phi)$$
$$m_I < \frac{Y(\phi)}{\epsilon(\phi)}$$

Applying the rest of 5.10 gives us

$$|\partial(\bar{M}, A)| \leq m_I \cdot K(\phi) \leq \frac{K(\phi)Y(\phi)}{\epsilon(\phi)}$$
$$|\bar{M} \cap A| \leq m_I \cdot K(\phi) \leq \frac{K(\phi)Y(\phi)}{\epsilon(\phi)}$$

as needed. This ends the proof for Theorem 5.5. □

So far we have computed an upper bound for a single basic formula ϕ.

To bound an arbitrary formula, write it as a boolean combination of basic formulas ϕ_i (via quantifier elimination) It suffices to bound vc-density for collection of formulas $\{\phi_i\}$ to obtain a bound for the original formula.

In general work with a collection of basic formulas $\{\phi_i\}_{i\in I}$. The proof generalizes in a straightforward manner. Instead of $A^{|x|}$ we now work with $A^{|x|} \times I$ separating traces of different formulas. Formula with the largest quantity $Y(\phi)\frac{K(\phi)}{\epsilon(\phi)}$ contributes the most to the vc-density. Thus we have

$$\Phi = \{\phi_i\}_{i\in I}$$
$$\mathrm{vc}(\Phi) \leq \max_{i\in I} Y(\phi_i)\frac{K(\phi_i)}{\epsilon_{\phi_i}}$$

References

[1] M. Aschenbrenner, A. Dolich, D. Haskell, D. Macpherson, S. Starchenko, *Vapnik-Chervonenkis density in some theories without the independence property*, I, Trans. Amer. Math. Soc. 368 (2016), 5889-5949

[2] Michael C. Laskowski, *A simpler axiomatization of the Shelah-Spencer almost sure theories*, Israel J. Math. **161** (2007), 157186. MR MR2350161

[3] P. Assouad, *Densite et dimension*, Ann. Inst. Fourier (Grenoble) 33 (1983), no. 3, 233-282.

[4] P. Assouad, *Observations sur les classes de Vapnik-Cervonenkis et la dimension combinatoire de Blei*, in: Seminaire dAnalyse Harmonique, 1983-1984, pp. 92-112, Publications Mathematiques dOrsay, vol. 85-2, Universite de Paris-Sud, Departement de Mathematiques, Orsay, 1985.

[5] N. Sauer, *On the density of families of sets*, J. Combinatorial Theory Ser. A 13 (1972), 145-147.

[6] S. Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific J. Math. 41 (1972), 247-261.

E-mail address: `bobkov@math.ucla.edu`