

VC-DENSITY FOR TREES

ANTON BOBKOV

ABSTRACT. We show that for the theory of infinite trees we have $vc(n) = n$ for all n .

1. PRELIMINARIES

We use notation $a \in T^n$ for tuples of size n . For variable x or tuple a we denote their arity by $|x|$ and $|a|$ respectively.

We work with finite relational languages. Given a formula we can define its complexity n as the number of quantifiers in its normal form. $S_A^n(x)$ stands for all the types made up of formulas of complexity at most n in a structure A . $tp_B^n(a)$ stands for such a type. For two structures A, B we say $A \equiv_n B$ if two structures agree on all sentences of complexity at most n .

Note 1.1. Saying that $(A, a_1) \equiv_n (A, a_2)$ is the same as saying that a_1 and a_2 have the same n -complexity type in A .

Language for the trees consists of a single binary predicate $\{\leq\}$. Theory of trees states that \leq defines a partial order and for every element a we have $\{x \mid x < a\}$ a linear order. Theory of meet trees requires that in addition tree is closed under meet operation, i.e. for any a, b in the same connected component there exists the greatest upper bound for elements both \leq than a and b . Note that we allow our trees to be disconnected or finite unless otherwise stated.

2. PROPER SUBDIVISIONS: DEFINITION AND PROPERTIES

Definition 2.1. Let A, B, T be models in (possibly different) finite relational languages. If A, B partition T (i.e. $T = A \sqcup B$) we say that (A, B) is a *subdivision* of T .

Definition 2.2. (A, B) subdivision of T is called *n-proper* if for all $p, q \in \mathbb{N}$, for all $a_1, a_2 \in A^p$ and $b_1, b_2 \in B^q$ we have

$$\begin{aligned} (A, a_1) &\equiv_n (A, a_2) \\ (B, b_1) &\equiv_n (B, b_2) \end{aligned}$$

then

$$(T, a_1, b_1) \equiv_n (T, a_2, b_2)$$

Definition 2.3. (A, B) subdivision of T is called *proper* if it is n -proper for all $n \in \mathbb{N}$.

Lemma 2.4. Consider a subdivision (A, B) of T . If it is 0-proper then it is proper.

Proof. Prove the subdivision is n -proper for all n by induction. Case $n = 0$ is given by the assumption. Suppose $n = k + 1$ and we have $\mathbf{T} \models \exists x \phi^k(x, a_1, b_1)$ where ϕ^k is some formula of complexity k . Let $a \in T$ witness the existential claim i.e. $\mathbf{T} \models \phi^k(a, a_1, b_1)$. $a \in A$ or $a \in B$. Without loss of generality assume $a \in A$. Let $\mathbf{p} = \text{tp}_{\mathbf{A}}^k(a, a_1)$. Then we have

$$\mathbf{A} \models \exists x \text{tp}_{\mathbf{A}}^k(x, a_1) = \mathbf{p}$$

Formula $\text{tp}_{\mathbf{A}}^k(x, a_1) = \mathbf{p}$ is of complexity k so $\exists x \text{tp}_{\mathbf{A}}^k(x, a_1) = \mathbf{p}$ is of complexity $k + 1$ by inductive hypothesis we have

$$\mathbf{A} \models \exists x \text{tp}_{\mathbf{A}}^k(x, a_2) = \mathbf{p}$$

Let a' witness this existential claim so that

$$\begin{aligned} \text{tp}_{\mathbf{A}}^k(a', a_2) &= \mathbf{p} \\ \text{tp}_{\mathbf{A}}^k(a', a_2) &= \text{tp}_{\mathbf{A}}^k(a, a_1) \\ (\mathbf{A}, a', a_2) &\equiv_k (\mathbf{A}, a, a_1) \end{aligned}$$

by inductive assumption we have

$$\begin{aligned} (\mathbf{T}, a, a_1, b_1) &\equiv_k (\mathbf{T}, a', a_2, b_2) \\ \mathbf{T} \models \phi^k(a', a_2, b_2) &\quad \text{as } \mathbf{T} \models \phi^k(a, a_1, b_1) \\ \mathbf{T} \models \exists x \phi^k(x, a_2, b_2) \end{aligned}$$

□

We don't require this lemma in full generality. From now on in this paper we'll have \mathbf{T} to be a model of a tree in the language $\mathcal{L} = \{\leq\}$ and \mathbf{A}, \mathbf{B} be in some languages $\mathcal{L}_A, \mathcal{L}_B$ which will be expands of \mathcal{L} , with \mathbf{A}, \mathbf{B} substructures of \mathbf{T} as reducts to \mathcal{L} . We'll refer to (\mathbf{A}, \mathbf{B}) as a *proper subdivision* (\mathbf{T} will be dropped if it is implied from context).

Example 2.5. Suppose the tree consists of two connected components C_1, C_2 . Then (C_1, \leq) and (C_2, \leq) form a proper subdivision.

Example 2.6. Fix \mathbf{T} and $a \in T$. Let $B = \{t \in T \mid a < t\}$, $S = \{t \in T \mid t \leq a\}$, $A = T - B$. Then (A, \leq, S) and (B, \leq) form a proper subdivision, where \mathcal{L}_A has a unary predicate interpreted by S .

Definition 2.7. For $\phi(x, y)$, $A \subseteq T^{|x|}$ and $B \subseteq T^{|y|}$

- Let $\phi(A, b) = \{a \in A \mid \phi(a, b)\} \subseteq A$
- Let $\phi(A, B) = \{\phi(A, b) \mid b \in B\} \subseteq \mathcal{P}(A)$

$\phi(A, B)$ is a collection of subsets of A definable by ϕ with parameters from B . We notice the following bound when A, B are parts of a proper subdivision.

Corollary 2.8. Suppose $\phi(x, y)$ is a formula of complexity n . Let \mathbf{A}, \mathbf{B} be a proper subdivision of \mathbf{T} and $b_1, b_2 \in B^{|y|}$. Then if $\text{tp}_{\mathbf{B}}^n(b_1) = \text{tp}_{\mathbf{B}}^n(b_2)$ then $\phi(A^{|x|}, b_1) = \phi(A, b_2)$. Thus $|\phi(A^{|x|}, B^{|y|})|$ is bounded by $|S_{\mathbf{B}}^n(y)|$

Proof. Take some $a \in A^{|x|}$. We have $(\mathbf{B}, b_1) \equiv_n (\mathbf{B}, b_2)$ and (trivially) $(\mathbf{A}, a) \equiv_n (\mathbf{A}, a)$. Thus by the Lemma 2.4 we have $(\mathbf{T}, a, b_1) \equiv_n (\mathbf{T}, a, b_2)$ so $\phi(a, b_1) \iff \phi(a, b_2)$. Since a was arbitrary we have $\phi(A^{|x|}, b_1) = \phi(A^{|x|}, b_2)$. □

Now we note that the number of such types can be bounded uniformly.

Note 2.9. Fix a (finite relational) language \mathcal{L}_B , and $n, |y|$. Then there is some $N = N(n, |y|, \mathcal{L}_B)$ such that for any structure \mathbf{B} in \mathcal{L}_B we have $|S_B^n(y)| \leq N$

3. PROPER SUBDIVISIONS: CONSTRUCTIONS

First, we describe several constructions of proper subdivisions that are needed for the proof.

Definition 3.1. We say that $E(b, c)$ if b and c are connected

$$E(b, c) \Leftrightarrow \exists x (b \geq x) \wedge (c \geq x)$$

Similarly $E_a(b, c)$ means that b and c are connected through an element above a . More precisely

$$E_a(b, c) \Leftrightarrow \exists x (x > a) \wedge (b \geq x) \wedge (c \geq x)$$

In the following four definitions \mathbf{B} -structures are going to be in the same language $\mathcal{L}_B = \{\leq, U\}$ with U a unary predicate. It is not always necessary to have this predicate but for the sake of uniformity we keep it. \mathbf{A} -structures will have different \mathcal{L}_A languages (those are not as important in later applications).

Definition 3.2. Fix $c_1 < c_2$ in T . Let

$$\begin{aligned} B &= \{b \in T \mid E_{c_1}(c_2, b) \wedge \neg(b \geq c_2)\} \\ A &= T - B \\ S_1 &= \{t \in T \mid t < c_1\} \\ S_2 &= \{t \in T \mid t < c_2\} \\ S_B &= S_2 - S_1 \\ T_A &= \{t \in T \mid c_2 \leq t\} \end{aligned}$$

Define structures $\mathbf{A}_{c_2}^{c_1} = (A, \leq, S_1, T_A)$ and $\mathbf{B}_{c_2}^{c_1} = (B, \leq, S_B)$ where \mathcal{L}_A is expansion of $\{\leq\}$ by two unary predicates (and \mathcal{L}_B as defined above). Note that $c_1, c_2 \notin B$.

Definition 3.3. Fix c in T . Let

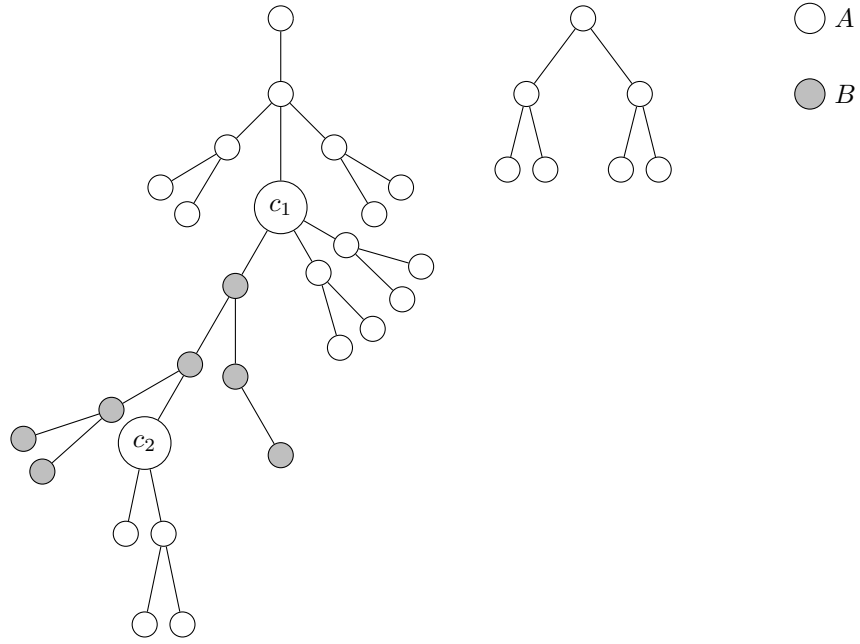
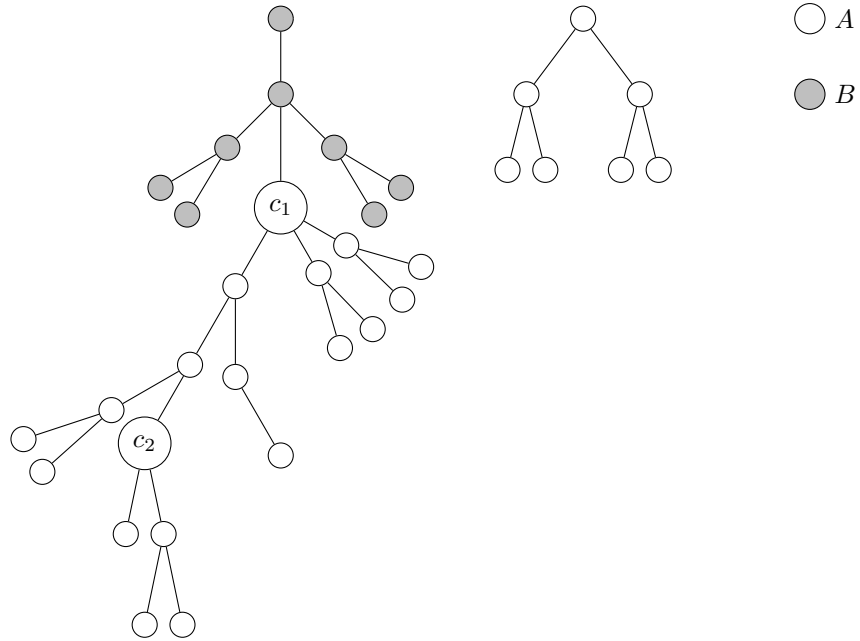
$$\begin{aligned} B &= \{b \in T \mid \neg(b \geq c) \wedge E(b, c)\} \\ A &= T - B \\ S_1 &= \{t \in T \mid t < c\} \end{aligned}$$

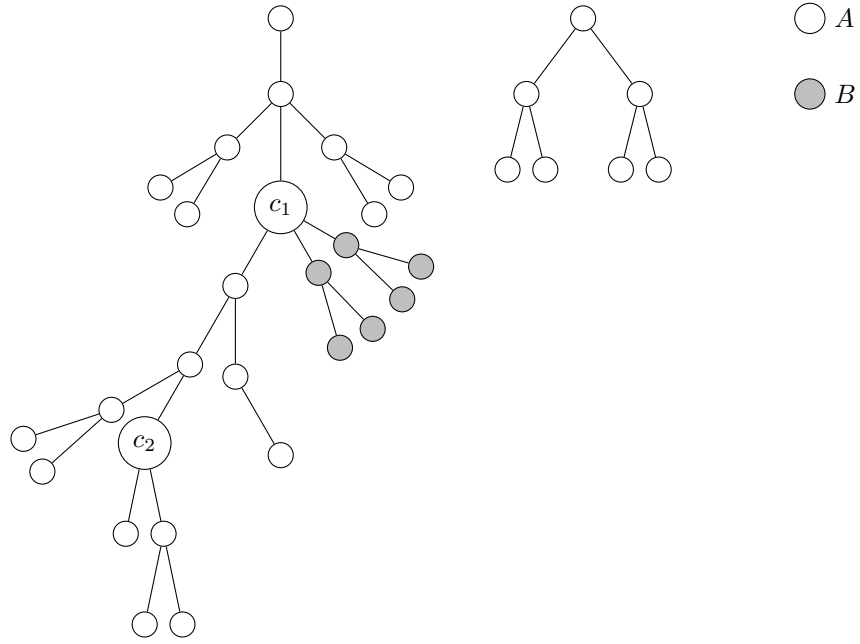
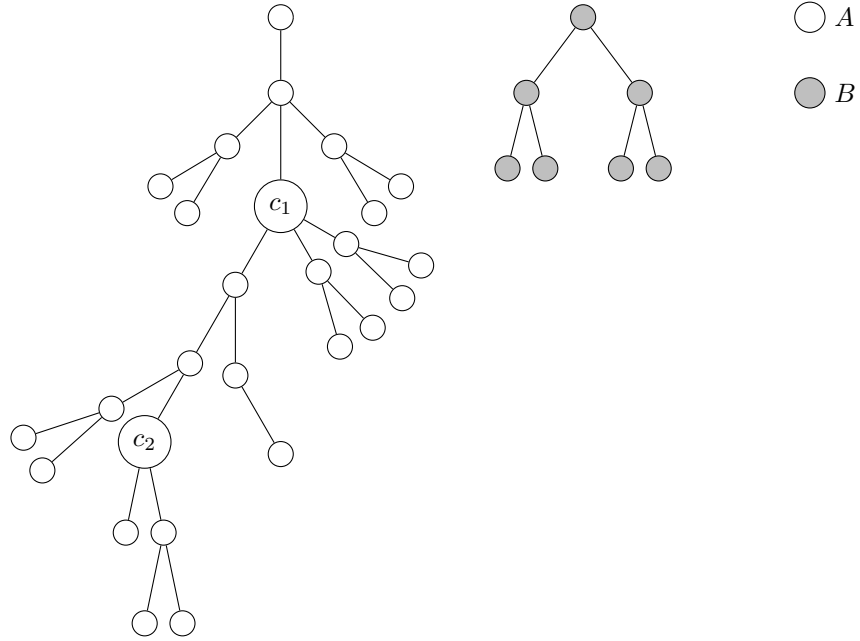
Define structures $\mathbf{A}_c = (A, \leq)$ and $\mathbf{B}_c = (B, \leq, S_1)$ where $\mathcal{L}_A = \{\leq\}$ (and \mathcal{L}_B as defined above). Note that $c \notin B$. (cf example 2.6).

Definition 3.4. Fix c in T and $S \subseteq T$ a finite subset. Let

$$\begin{aligned} B &= \{b \in T \mid (b > c) \text{ and for all } s \in S \text{ we have } \neg E_c(s, b)\} \\ A &= T - B \\ S_1 &= \{t \in T \mid t \leq c\} \end{aligned}$$

Define structures $\mathbf{A}_S^c = (A, \leq, S_1)$ and $\mathbf{B}_S^c = (B, \leq, B)$ where \mathcal{L}_A is expansion of $\{\leq\}$ by a single unary predicate (and $U \in \mathcal{L}_B$ vacuously interpreted by B). Note that $c \notin B$ and $S \cap B = \emptyset$.

FIGURE 1. Proper subdivision for $(A, B) = (A_{c_2}^{c_1}, B_{c_2}^{c_1})$ FIGURE 2. Proper subdivision for $(A, B) = (A_{c_1}, B_{c_1})$

FIGURE 3. Proper subdivision for $(A, B) = (A_G^{c_1}, B_G^{c_1})$ for $G = \{c_2\}$ FIGURE 4. Proper subdivision for $(A, B) = (A_G, B_G)$ for $G = \{c_1, c_2\}$

Definition 3.5. Fix $S \subseteq T$ a finite subset. Let

$$B = \{b \in T \mid \text{for all } s \in S \text{ we have } \neg E(s, b)\}$$

$$A = T - B$$

Define structures $\mathbf{A}_S = (A, \leq)$ and $\mathbf{B}_S = (B, \leq, B)$ where $\mathcal{L}_A = \{\leq\}$ (and $U \in \mathcal{L}_B$ vacuously interpreted by B). Note that $S \cap B = \emptyset$.

Lemma 3.6. *Pairs of structures defined above are all proper subdivisions.*

Proof. We only show this holds for the first definition $\mathbf{A} = \mathbf{A}_{c_2}^{c_1}$ and $\mathbf{B} = \mathbf{B}_{c_2}^{c_1}$. Other cases follow by a similar argument. A, B partition T by definition, so it is a subdivision. To show that it is proper by Lemma 2.4 we only need to check that it is 0-proper. Suppose we have

$$a = (a_1, a_2, \dots, a_p) \in A^p$$

$$a' = (a'_1, a'_2, \dots, a'_p) \in A^p$$

$$b = (b_1, b_2, \dots, b_q) \in B^q$$

$$b' = (b'_1, b'_2, \dots, b'_q) \in B^q$$

with $(\mathbf{A}, a) \equiv_0 (\mathbf{A}, a')$ and $(\mathbf{B}, b) \equiv_0 (\mathbf{B}, b')$. We need to make sure that ab has the same quantifier free type as $a'b'$. Any two elements in T can be related in the four following ways

$$x = y$$

$$x < y$$

$$x > y$$

$$x, y \text{ are incomparable}$$

We need to check that the same relations hold for pairs of $(a_i, b_j), (a'_i, b'_j)$ for all i, j .

- It is impossible that $a_i = b_j$ as they come from disjoint sets.
- Suppose $a_i < b_j$. This forces $a_i \in S_1$ thus $a'_i \in S_1$ and $a'_i < b'_j$
- Suppose $a_i > b_j$. This forces $b_j \in S_B$ and $a \in T_A$, thus $b'_j \in S_B$ and $a'_i \in T_A$ so $a'_i > b'_j$
- Suppose a_i and b_j are incomparable. Two cases are possible:
 - $b_j \notin S_B$ and $a_i \in T_A$. Then $b'_j \notin S_B$ and $a'_i \in T_A$ making a'_i, b'_j incomparable
 - $b_j \in S_B$, $a_i \notin T_A$, $a_i \notin S_1$. Similarly this forces a'_i, b'_j incomparable

□

4. MAIN PROOF

Basic idea for the proof is that we are able to divide our parameter space into $O(n)$ many pieces. Each of q parameters can come from any of those $O(n)$ partitions giving us $O(n)^q$ many choices for parameter configuration. When every parameter coming from a fixed partition the number of definable sets is constant and in fact is uniformly bounded by some N . This gives us $NO(n)^q = O(n^q)$ possibilities for different definable sets.

First, we generalize Corollary 2.8. (This is only required for computing vc-density for formulas $\phi(x, y)$ with $|y| > 1$)

Lemma 4.1. *Consider a finite collection $(A_i, B_i)_{i \leq n}$ where each (A_i, B_i) is a proper subdivision or a singleton: $B_i = \{b_i\}$ with $A_i = T$. Also assume that all B_i have the same language \mathcal{L}_B . Let $A = \bigcap_{i \in I} A_i$. Fix a formula $\phi(x, y)$ of complexity m . Let $N = N(m, |y|, \mathcal{L}_B)$ as in Note 2.9. Consider any $B \subseteq T^{|y|}$ of the form*

$$B = B_1^{i_1} \times B_1^{i_2} \times \dots \times B_n^{i_n} \text{ with } i_1 + i_2 + \dots + i_n = |y|$$

(some of the indexes can be zero). Then we have the following bound

$$\phi(A^{|x|}, B) \leq N^{|y|}$$

Proof. We show this result by counting types. Suppose we have

$$b_1, b'_1 \in B_1^{i_1} \text{ with } b_1 \equiv_m b'_1 \text{ in } B_1$$

$$b_2, b'_2 \in B_2^{i_2} \text{ with } b_2 \equiv_m b'_2 \text{ in } B_2$$

...

$$b_n, b'_n \in B_n^{i_n} \text{ with } b_n \equiv_m b'_n \text{ in } B_n$$

Then we have

$$\phi(A^{|x|}, b_1, b_2, \dots, b_n) \Leftrightarrow \phi(A^{|x|}, b'_1, b'_2, \dots, b'_n)$$

This is easy to see by applying Corollary 2.8 one by one for each tuple. This works if B_i is part of a proper subdivision; if it is a singleton then the implication is trivial as $b_i = b'_i$. This shows that $\phi(A^{|x|}, B)$ only depends on the choice of types for the tuples

$$|\phi(A^{|x|}, B)| \leq |\text{tp}_{B_1}^m(i_1)| \cdot |\text{tp}_{B_2}^m(i_2)| \cdot \dots \cdot |\text{tp}_{B_n}^m(i_n)|$$

Now for each type space we have inequality

$$|\text{tp}_{B_1}^m(i_1)| \leq N(m, i_1, \mathcal{L}_B) \leq N(m, |y|, \mathcal{L}_B) \leq N$$

(For singletons $|\text{tp}_{B_j}^m(i_j)| = 1 \leq N$). Only non-zero indexes contribute to the product and there are at most $|y|$ of those (by equality $i_1 + i_2 + \dots + i_n = |y|$). Thus we have

$$|\phi(A^{|x|}, B)| \leq N^{|y|}$$

as needed. \square

For subdivisions to work out properly we will need to work with subsets closed under meets. We observe that closure under meets doesn't add too many new elements.

Lemma 4.2. *Suppose $S \subseteq T$ is a non-empty finite subset of a meet tree of size n and S' its closure under meets. Then $|S'| \leq 2n - 1$.*

Proof. We prove by induction on n . Base case $n = 1$ is clear. Suppose we have S of size k with closure of size at most $2k - 1$. Take a new point and look at its meets with all the elements of S . Pick the largest one. That is the only element we need to add to S' to make sure the set is closed under meets. \square

Putting all of those results together we are able to compute vc-density of formulas in meet trees.

Theorem 4.3. *Let T be an infinite meet tree and $\phi(x, y)$ a formula with $|x| = p$ and $|y| = q$. Then $\text{vc}(\phi) \leq q$.*

Proof. Pick a finite subset of $S_0 \subset T^p$ of size n . Let $S_1 \subset T$ consist of coordinates of S_0 . Let $S \subset T$ be a closure of S_1 under meets. Using Lemma 4.2 we have $|S_2| \leq 2|S_1| \leq 2p|S_0| = 2pn = O(n)$. We have $S_0 \subseteq S^p$, so $|\phi(S_0, T^q)| \leq |\phi(S^p, T^q)|$. Thus it is enough to show $|\phi(S^p, T^q)| = O(n^q)$.

Label $S = \{c_i\}_{i \in I}$ with $|I| \leq 2pn$. For every c_i we construct two partitions in the following way. We have c_i is either minimal in S or it has a predecessor in S (greatest element less than c). If it is minimal construct $(\mathbf{A}_{c_i}, \mathbf{B}_{c_i})$. If there is a predecessor p construct $(\mathbf{A}_{c_i}^p, \mathbf{B}_{c_i}^p)$. For the second subdivision let G be all elements in S greater than c_i and construct $(\mathbf{A}_G^c, \mathbf{B}_G^c)$. So far we have constructed two subdivisions for every $i \in I$. Additionally construct $(\mathbf{A}_S, \mathbf{B}_S)$. We end up with a finite collection of proper subdivisions $(\mathbf{A}_j, \mathbf{B}_j)_{j \in J}$ with $|J| = 2|I| + 1$. Before we proceed we note the following two lemmas describing our partitions.

Lemma 4.4. *For all $j \in J$ we have $S \subseteq A_j$. Thus $S \subseteq \bigcap_{j \in J} A_j$ and $S^p \subseteq \bigcap_{j \in J} (A_j)^p$*

Proof. Check this for each possible choices of partition. Cases for partitions of the type $\mathbf{A}_S, \mathbf{A}_G^c, \mathbf{A}_c$ are easy. Suppose we have partition $(\mathbf{A}, \mathbf{B}) = (\mathbf{A}_{c_2}^{c_1}, \mathbf{B}_{c_2}^{c_1})$. We need to show that $B \cap S = \emptyset$. By construction we have $c_1, c_2 \notin B$. Suppose we have some other $c \in S$ with $c \in B$. We have $E_{c_1}(c_2, c)$ i.e. there is some b such that $(b > c_1), (b \leq c_2)$ and $(b \leq c)$. Consider the meet $(c \wedge c_2)$. We have $(c \wedge c_2) \geq b > c_1$. Also as $\neg(c \geq c_2)$ we have $(c \wedge c_2) < c_2$. To summarize $c_2 > (c \wedge c_2) > c_1$. But this contradicts our construction as S is closed under meets, so $(c \wedge c_2) \in S$ and c_1 is supposed to be a predecessor of c_2 in S . \square

Lemma 4.5. *$\{B_j\}_{j \in J}$ partition $T - S$ i.e. $T = \bigsqcup_{j \in J} B_j \sqcup S$*

Proof. This more or less follows from the choice of partitions. Pick any $b \in S - T$. Take all elements in S greater than b and take the minimal one a . Take all elements in S less than b and take the maximal one c (possible as S is closed under meets). Also take all elements in S incomparable to b and denote them G . If both a and c exist we have $b \in \mathbf{B}_c^a$. If only upper bound exists we have $b \in \mathbf{B}_G^a$. If only lower bound exists we have $b \in \mathbf{B}_c$. If neither exists we have $b \in \mathbf{B}_G$. \square

Note 4.6. Those two lemmas imply $S = \bigcap_{j \in J} A_j$

Note 4.7. For one-dimensional case $q = 1$ we don't need to do any more work. We have partitioned parameter space into $|J| = O(n)$ many pieces and over each piece the number of definable sets is uniformly bounded. By Note 2.9 we have that $|\phi((A_j)^p, B_j)| \leq N$ for any $j \in J$ (letting $N = N(n_\phi, q, \{\leq, S\})$ where n_ϕ is

complexity of ϕ and S is a unary predicate). Compute

$$\begin{aligned}
|\phi(S^p, T)| &= \left| \bigcup_{j \in J} \phi(S^p, B_j) \cup \phi(S^p, S) \right| \leq \\
&\leq \sum_{j \in J} |\phi(S^p, B_j)| + |\phi(S^p, S)| \leq \\
&\leq \sum_{j \in J} |\phi((A_j)^p, B_j)| + |S| \leq \\
&\leq \sum_{j \in J} N + |I| \leq \\
&\leq (4pn + 1)N + 2pn = (4pN + 2p)n + N = O(n)
\end{aligned}$$

Basic idea for the general case $q \geq 1$ is that we have q parameters and $|J| = O(n)$ partitions to pick each parameter from giving us $|J|^q = O(n^q)$ choices for parameter configuration, each giving uniformly constant number of definable subsets of S . (If every parameter is picked from a fixed partition, Lemma 4.1 provides a uniform bound). This yields $\text{vc}(\phi) \leq q$ as needed. The rest of the proof is stating this idea formally.

First, we extend our collection of subdivisions $(\mathbf{A}_j, \mathbf{B}_j)_{j \in J}$ by the following singleton sets. For each $c_i \in S$ let $B_i = \{c_i\}$ and $A_i = T$ and add $(\mathbf{A}_i, \mathbf{B}_i)$ to our collection with \mathcal{L}_B the language of B_i interpreted arbitrarily. We end up with a new collection $(\mathbf{A}_k, \mathbf{B}_k)_{k \in K}$ indexed by some K with $|K| = |J| + |I|$ (we added $|S|$ new pairs). Now we have that B_k partition T , so $T = \bigsqcup_{k \in K} B_k$ and $S = \bigcap_{j \in J} A_j = \bigcap_{k \in K} A_k$. For $(k_1, k_2, \dots, k_q) = \vec{k} \in K^q$ denote

$$B_{\vec{k}} = B_{k_1} \times B_{k_2} \times \dots \times B_{k_q}$$

Then we have the following identity

$$T^q = \left(\bigsqcup_{k \in K} B_k \right)^q = \bigsqcup_{\vec{k} \in K^q} B_{\vec{k}}$$

Thus we have that $\{B_{\vec{k}}\}_{\vec{k} \in K^q}$ partition T^q . Compute

$$\begin{aligned}
|\phi(S^p, T^q)| &= \left| \bigcup_{\vec{k} \in K^q} \phi(S^p, B_{\vec{k}}) \right| \leq \\
&\leq \sum_{\vec{k} \in K^q} |\phi(S^p, B_{\vec{k}})|
\end{aligned}$$

We can bound $|\phi(S^p, B_{\vec{k}})|$ uniformly using Lemma 4.1. $(\mathbf{A}_k, \mathbf{B}_k)_{k \in K}$ satisfies the requirements of the lemma and $B_{\vec{k}}$ looks like B in the lemma after possibly permuting some variables in ϕ . Applying the lemma we get

$$|\phi(S^p, B_{\vec{k}})| \leq N^q$$

with N only depending on q and complexity of ϕ . We complete our computation

$$\begin{aligned}
|\phi(S^p, T^q)| &\leq \sum_{\vec{k} \in K^q} |\phi(S^p, B_{\vec{k}})| \leq \\
&\leq \sum_{\vec{k} \in K^q} N^q \leq \\
&\leq |K^q| N^q \leq \\
&\leq (|J| + |I|)^q N^q \leq \\
&\leq (4pn + 1 + 2pn)^q N^q = N^q (6p + 1/n)^q n^q = O(n^q)
\end{aligned}$$

□

Corollary 4.8. *In the theory of infinite meet trees we have $vc(n) = n$ for all $n \in \mathbb{N}^+$.*

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