Chapter 5:
Beyond Sparsity

Compilation date: January 11, 2020

1 Introduction

In the previous chapters we have studied classes of sparse graphs, putting focus on classes of bounded expansion and on nowhere dense classes. In these settings, we often obtained dichotomy theorems for *subgraph-closed* classes; such theorems usually asserted that a subgraph-closed class has some nice structural or algorithmic property if and only if it is nowhere dense. In this chapter we reach a bit beyond sparsity, and focus on graphs that are dense, but are well-structured.

There are two possible ways one may proceed, and in some sense we will try both of them. First, we may study classes of dense graphs derived from sparse graphs, hoping that the existence of a sparse "skeleton" — the pre-image under derivation — imposes a useful structure in the studied graph. Second, we may try to define other abstract properties of graph classes, more general than nowhere denseness, which aim to capture the essence of the tools we use in Sparsity. The hope here would be to extend the applicability of the Sparsity techniques by creating their more general analogues.

During the tutorials we have already seen the simplest example of an operation that from a class of sparse graphs creates a class of possibly dense, but well structured graphs. This operation is taking the *power* of a graph. Formally, for a graph G and $r \in \mathbb{N}$, the r-th power of G, denoted G^r, is the graph on the same vertex set as G where two vertices are adjacent if and only if they are at distance at most r in G. For a class of graphs $\mathcal{C}$, we define its r-th power as follows:

$$\mathcal{C}^r = \{G^r : G \in \mathcal{C}\}.$$

Note that the square of the class of stars is the class of cliques, hence this operator may turn a class of sparse graphs into a class of dense graphs. Further, note that $\mathcal{C}^r$ is generally not closed under taking subgraphs.

The operator of taking the power of a graph is a very simple example of a *first-order interpretation.* We will briefly discuss this more general setting at the end, but for now we actually focus on powers of classes of sparse graphs.

2 Stability

2.1 Excluding ladders and related structures

The initial idea is to seek structures that are forbidden in powers of sparse graphs. If we identify such robust structures, we can use them as obstacles for constructing a theory of dense, well-structured graphs. These structures will be *ladders*, but on the way we will also define *semi-ladders* and *co-matchings.*

Definition 2.1. A pair of sequences $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_\ell$ of vertices of a graph G is called

- a *co-matching* if for all $i, j \in \{1, \ldots, \ell\}$: a_i is adjacent to b_j if and only if $i \neq j$;
- a *ladder* if for all $i, j \in \{1, \ldots, \ell\}$: a_i is adjacent to b_j if and only if $i > j$; and
- a *semi-ladder* if for all $i \in \{1, \ldots, \ell\}$: a_i is not adjacent to b_i, but is adjacent to all b_j satisfying $i > j$.

The *co-matching index* of a class $\mathcal{C}$ is the supremum of the lengths of co-matchings that can be found in graphs from $\mathcal{C}$. Define the *ladder index* and the *semi-ladder index* similarly.

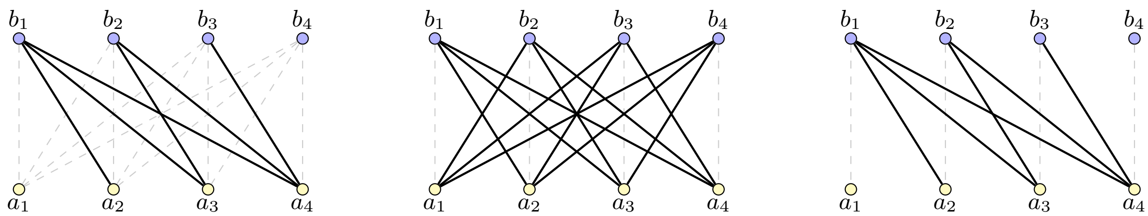

Figure 1: A ladder, a co-matching, and a semi-ladder of length 4.

We note that boundedness of the semi-ladder index is equivalent to the boundedness of both the ladder index and the co-matching index. This follows from an easy Ramsey argument.

Lemma 2.2. *For a class of graphs $\mathcal{C}$, the following conditions are equivalent:*

(a) $\mathcal{C}$ has a finite semi-ladder index.

(b) $\mathcal{C}$ has a finite ladder index and a finite co-matching index.

Proof. The implication from (a) to (b) follows from the observation that every ladder is also a semi-ladder, and also every co-matching is a semi-ladder, hence the semi-ladder index is always larger or equal to both the ladder index and the co-matching index.

For the implication from (b) to (a), suppose that both the ladder index and the co-matching index of $\mathcal{C}$ are smaller than k. Suppose that a graph $G \in \mathcal{C}$ contains a semi-ladder $(a_1, \ldots, a_\ell), (b_1, \ldots, b_\ell)$ of length $\ell \geq R(k, k)$, where $R(k, k)$ is the standard Ramsey number. Construct an auxiliary complete graph H on vertex set $\{1, \ldots, \ell\}$, where between indices $i < j$ we put a blue edge if $a_i b_j$ is an edge in G, and a red edge otherwise. By Ramsey's theorem, in H there is either a blue clique of size k, or a red clique of size k. But the first outcome would give a co-matching of length k in G, while the second outcome would give a ladder of length k in G. In both cases this would be a contradiction, which proves that the semi-ladder index of $\mathcal{C}$ is smaller than $R(k, k)$. $\square$

It turns out that powers of sparse graphs actually have a bounded semi-ladder index.

Theorem 2.3. *For every nowhere dense class of graphs $\mathcal{C}$ and $r \in \mathbb{N}$, the class $\mathcal{C}^r$ has a finite semi-ladder index.*

Proof. Since $\mathcal{C}$ is nowhere dense, it is also uniformly quasi-wide, say with margins $(s_r)_{r\in\mathbb{N}}$ and wideness functions $(N_r(\cdot))_{r\in\mathbb{N}}$. Fix $r \in \mathbb{N}$ and consider any graph $G \in \mathcal{C}$. Suppose in G^r we can find a semi-ladder $(a_1,\ldots,a_\ell),(b_1,\ldots,b_\ell)$ of length ℓ. Letting $s = s_{2r}$ and $t = 2\cdot(r+1)^s$, we shall prove that $\ell < N_{2r}(t+1)$. As G was taken arbitrarily, this gives a corresponding upper bound on the semi-ladder index of $\mathcal{C}^r$.

For the sake of contradiction suppose this is not the case, that is, $\ell \geq N_{2r}(t+1)$. Let $A = \{a_1,\ldots,a_\ell\}$; then $|A| = \ell \geq N_{2r}(t+1)$. By uniform quasi-wideness, we can find disjoint vertex subsets $S \subseteq V(G)$ and $A' \subseteq A - S$ such that $|S| \leq s$, $|A'| > t$, and A' is distance-$2r$ independent in $G - S$.

Recall that for a vertex $u \in V(G)$, the distance-r profile of u on S is defined as the function $\mathrm{profile}_r[u,S]\colon S \to \{0,1,\ldots,r,+\infty\}$ that measures distances from u to all the vertices of S, where values larger than r are replaced by $+\infty$. Note that vertices of A' realize at most $(r+1)^s$ possible different distance-r distance profiles on S. Since $|A| > t = 2\cdot(r+1)^s$, we can find three indices $1 \leq \alpha < \beta < \gamma \leq k$ such that the vertices $x \coloneqq a_\alpha, y \coloneqq a_\beta, z \coloneqq a_\gamma$ belong to A' and have equal distance-r profiles on S. Denote $w \coloneqq b_\gamma$. In particular, we have the following assertions

- the distance between x and y in $G - S$ is larger than $2r$;
- the vertices x, y, z have the same distance-r profiles on S; and
- $\mathrm{dist}_G(w,z) > r$, $\mathrm{dist}_G(w,x) \leq r$, and $\mathrm{dist}_G(w,y) \leq r$.

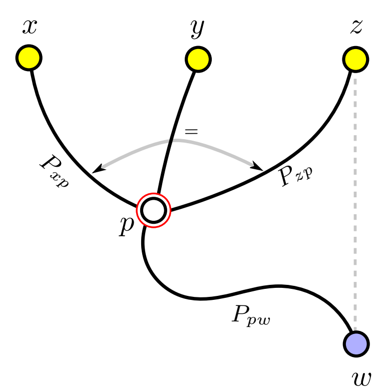

Figure 2: Proof of Theorem 2.3: contradiction looms.

We show that simultaneous satisfaction of assertions above leads to a contradiction (see Figure 2). Let P_{xw} be a path of length at most r connecting x and w, and let P_{wy} be a path of length at most r connecting w and y. In particular, the concatenation of P_{xw} and P_{wy} has length at most $2r$ and connects x and y. Since the distance between x and y in $G - S$ is larger than $2r$, at least one of the paths P_{xw}, P_{wy} must contain a vertex $p \in S$. Suppose that it is P_{xw}, the other case being analogous. Then P_{xw} is split by p into two subpaths: P_{xp} and P_{pw}. Since x and z have the same distance-r profiles on S, we may find a path P_{zp} connecting z and p whose length is not larger than the length of P_{xp}. Now, the concatenation of paths P_{zp} and P_{pw} has length at most r and connects z with w, a contradiction with $\mathrm{dist}_G(w,z) > r$. □

2.2 Stability

Theorem 2.3 suggests that one may try to construct a robust theory of classes of dense, well-structured graphs by using long semi-ladders as forbidden structures. Indeed, powers of nowhere dense classes would be captured by such a theory. This is, however, not the most robust idea, as another basic operation — *complementation* of a graph — would turn classes that are "easy" from the point of view of theory into "hard" ones. Here, the *complement* of a graph G is the graph $\overline{G}$ on the same vertex set as G where we make two vertices adjacent if and only if they were *not* adjacent in G; for a class $\mathcal{C}$, we write $\overline{\mathcal{C}} = \{\overline{G} \colon G \in \mathcal{C}\}$. Indeed, if $\mathcal{C}$ is the class of all matchings — which undoubtedly is nowhere dense — then $\overline{\mathcal{C}}$ has infinite co-matching index, hence also infinite semi-ladder index.

But this class still has bounded ladder index, which gives us an idea that we should use long ladders as forbidden structures. We also need a notion of turning one graph into another that would encompass both taking powers and taking complements. For this, we will use a more general language of first-order logic and define relations between (tuples) of vertices using first-order formulas. Also, we will work in the setting of relational structures, though it does not add much complexity to the theory.

Definition 2.4. Let $\mathcal{C}$ be a class of structures over a signature Σ. For a formula $\varphi(\bar{x}, \bar{y}) \in \mathsf{FO}[\Sigma]$, where $\bar{x}, \bar{y}$ are sets of variables, and a structure $\mathbb{A} \in \mathcal{C}$, a *$\varphi$-ladder* of length ℓ is a pair of sequences

$$\bar{a}_1, \ldots, \bar{a}_\ell \in \mathbb{A}^{\bar{x}} \qquad \text{and} \qquad \bar{b}_1, \ldots, \bar{b}_\ell \in \mathbb{A}^{\bar{y}}$$

such that for all $i, j \in \{1, \ldots, \ell\}$, we have

$$\mathbb{A} \models \varphi(\bar{a}_i, \bar{b}_j) \qquad \text{if and only if} \qquad i > j.$$

We sometimes also say that φ *interpretes* a linear order of length ℓ in $\mathbb{A}$. For example, a (combinatorial) ladder in the graph G^r corresponds to a δ_r-ladder in G where $\delta_r(x, y)$ is a formula stating that $\mathrm{dist}(x, y) \leq r$. Similarly, a (combinatorial) ladder in the graph $\overline{G}$ corresponds to a α-ladder in G where $\alpha(x, y)$ checks that x and y are *not* adjacent.

We are ready to introduce the main definition of this section.

Definition 2.5. Let $\mathcal{C}$ be a class of structures over a signature Σ. The *ladder index* of a formula $\varphi(\bar{x}, \bar{y}) \in \mathsf{FO}[\Sigma]$ on $\mathcal{C}$ is the supremum of the lengths of φ-ladders in structures from $\mathcal{C}$. We say that $\mathcal{C}$ is *stable* if every first-order formula from $\mathsf{FO}[\Sigma]$ has a finite ladder index in $\mathcal{C}$.

Thus, intuitively, stable classes are exactly those where we cannot define arbitrarily long linear orders using a fixed first-order formula. The following theorem, which we will not prove, provides a fundamental link between Sparsity and stability.

Theorem 2.6. *A subgraph-closed class of graphs is stable if and only if it is nowhere dense.*

Thus, stability encompasses nowhere denseness, but there are classes of graphs that are not subgraph-closed, but are stable, as made explicit next. These are examples of classes of dense graphs where we could expect some algorithmically useful structure.

Lemma 2.7. *For every nowhere dense class of graphs $\mathcal{C}$, the following classes are stable: $\mathcal{C}^r$ for any $r \in \mathbb{N}$, and $\overline{\mathcal{C}}$.*

Proof. The first assertions follows from Theorem 2.6 and the observation that any φ-ladder in a graph $G^r \in \mathcal{C}^r$ is also a φ'-ladder in G, where φ' is obtained by replacing each adjacency check, say between variables x and y, with formula $\delta_r(x, y)$ asserting that $\mathrm{dist}(x, y) \leq r$. The proof for the complementation is analogous. □

Stability is a concept translated from model theory and was a subject of intensive studies since 60s by Morley, Shelah, and others. It is usually applied to infinite models, hence translating concepts from the infinite setting to the finite ones is usually a source of problems. Surprisingly, there are multiple analogies between the two theories: there is a matching analogue of neighborhood complexity, there are tools resembling in spirit uniform quasi-wideness, etc. However, stability theory in the finite, and in particular its algorithmic aspects, is not yet developed. A prime open problem here is to determine whether FO model-checking is fixed-parameter tractable on any stable class of graphs (subject to some minor technical conditions).

2.3 The Erdős-Hajnal property

We will now see some interesting application of the notion of ladder index. The classic Ramsey's theorem states that in an n-vertex graph there is either a clique or an independent set of size at least $\log n$. It is not hard to show using the probabilistic method that in a random n-vertex graph, where every edge appears independently with probability $\frac{1}{2}$, the maximum size of a clique or an independent set is $\Theta(\log n)$ with high probability; thus, the $\log n$ bound cannot be improved in general. The *Erdős-Hajnal conjecture* states that for every fixed graph H there exists a constant $\delta > 0$ such that if G is an n-vertex graph which excludes H as an induced subgraph, then G contains either a clique or an independent set of size $\Omega(n^\delta)$. In other words, the Ramsay number in graphs excluding a fixed H as an induced subgraphs is polynomial instead of exponential. The conjecture is still widely open even for very simple graphs H (the current frontier is $H = C_5$, the cycle on 5 vertices, and $H = P_5$, the path on 5 vertices).

We will now show that in the presence of a bound on the ladder index of a class, the Erdős-Hajnal conjecture for this class holds.

Theorem 2.8. *For every class of graph $\mathcal{C}$ with finite ladder index, there exists $\delta > 0$ such that every graph $G \in \mathcal{C}$, say on n vertices, contains either a clique or an independent set of size $\Omega(n^\delta)$.*

The remainder of this section is devoted to the proof of Theorem 2.8. Let k be the ladder index of $\mathcal{C}$. We prove the following lemma.

Lemma 2.9. *If $G \in \mathcal{C}$ has more than $(a + b)^{2k+2}$ vertices, for some $a, b \in \mathbb{N}$, then G contains either a clique of size a or an independent set of size b.*

Note that Lemma 2.9 implies Theorem 2.8 for $\delta = \frac{1}{2k+2}$. Hence, from now on we concentrate on proving Lemma 2.9.

The idea is to arrange the vertices of G in a binary tree and prove that provided $V(G)$ is sufficiently large, this tree contains a long path. From this path we will extract either a clique or an independent set.

We first need to establish some notation to be able to talk about the binary tree into which $V(G)$ will be arranged. We will work with a two-symbol alphabet $\{\mathrm{D}, \mathrm{S}\}$, where D is for *daughter* (left child) and S is for *son* (right child). The nodes of the binary tree will be described by words over this alphabet $\{\mathrm{D}, \mathrm{S}\}^\star$; thus a node is identified with a sequence of left/right turns leading to it from the root. The *depth* of a node w is the length of w. For $w \in \{\mathrm{D}, \mathrm{S}\}^\star$, the nodes $w\mathrm{D}$ and $w\mathrm{S}$ are called, respectively, the *daughter* and the *son* of w, and w is the *parent* of both $w\mathrm{S}$ and $w\mathrm{D}$. A node w' is a *descendant* of a node w if w' is a prefix of w (possibly $w' = w$). A *binary tree* τ is simply a set of nodes, as described above, that is closed under taking descendants. We may also think that the tree τ is labeled with some label set U; in this case, we let $\tau(x) \in U$ be the label of a node x.

Recall that we are working with a graph $G = (V, E)$ for which we assume that G has no ladders longer than k. Let $v_1, v_2, \ldots, v_n$ be any enumeration of vertices of G. We define a V-labelled binary tree τ with n nodes by an iterative procedure as follows. Start with τ being the empty tree. Then, for subsequent v_is, insert v_i into τ as follows. Start with w being the empty word. While w is a node of τ, repeat the following step: if v_i is adjacent to $\tau(w)$, replace w by its daughter, otherwise, replace w by its son. Once w is not a node of τ, extend τ by adding node w and setting $\tau(w) = a$. In this way, we have processed the vertex v_i, and now we proceed to the next vertex v_{i+1}, until all vertices are processed. Thus, τ is a tree labeled with vertices of G, and every vertex of G appears exactly once in τ.

For a word w, an *alternation* in w is any position α, $1 \leq \alpha \leq |w|$, such that $w_\alpha \neq w_{\alpha-1}$; here, w_α denotes the αth symbol of w, and w_0 is assumed to be D. The *alternation rank* of the tree τ is the maximum of the number of alternations in w, over all nodes w of τ. It appears that the assumption on the ladder index of $\mathcal{C}$ gives us an upper bound on the alternation rank of τ.

Lemma 2.10. *The alternation rank of the tree τ is at most $2k + 1$.*

Proof. Let w be a node of τ with at least 2ℓ alternations, for some $\ell \in \mathbb{N}$. Let $\alpha_1, \beta_1, \ldots, \alpha_\ell, \beta_\ell$ be the first 2ℓ alternations of w. Due to the assumption that $w_0 = \mathrm{D}$ we have that w contains symbol S at all positions α_i for $i = 1, \ldots, \ell$, and symbol D at all positions β_i for $i = 1, \ldots, \ell$. For each $i \in \{1, \ldots, \ell\}$, define $a_i \in V(G)$ to be the label in τ of the prefix of w of length $\alpha_i - 1$, and similarly define $b_i \in V(G)$ to be the label in τ of the prefix of w of length $\beta_i - 1$. Then for each $i \in \{1, \ldots, \ell\}$, the following assertions hold:

- the nodes in τ with labels $b_i, a_{i+1}, b_{i+1}, \ldots, a_\ell, b_\ell$ are descendants of the son of the node with label a_i, and
- the nodes with labels $a_{i+1}, b_{i+1}, \ldots, a_\ell, b_\ell$ are descendants of the daughter of the node with label b_i.

By definition of τ, this implies that $a_i b_j \in E$ if and only if $i > j$, for all $1 \leq i \leq \ell$. Since we assumed that G does not have ladders longer than k, this implies that $\ell \leq k$, proving the statement of the lemma. □

The *depth* of a binary tree is the maximal depth of its node. As we show next, having a constant upper bound on the alternation rank of a binary tree implies that its depth has to be polynomial in the number of its nodes, instead of logarithmic.

Lemma 2.11. *If a binary tree σ has alternation rank less than t and depth less than h, then σ has at most h^t nodes.*

Proof. It suffices to prove that the number of words over $\{\mathrm{D}, \mathrm{S}\}$ of length less than h and with less than t alternations is at most h^t. Observe that each such word is uniquely determined by its length and the set of alternations in it, which in turn can be encoded as a choice of a nonempty subset of size at most t over $\{1, \ldots, h\}$: the last element of the subset delimits the end of the word, while the previous ones are positions with alternations. Thus, the number of words over $\{\mathrm{D}, \mathrm{S}\}$ of length less than h and with less than t alternations is at most

$$\binom{h}{1} + \binom{h}{2} + \ldots + \binom{h}{t} \leq h^t. \qquad \square$$

Corollary 2.12. *The tree τ has depth at least $a + b$.*

Proof. If τ had depth less than $a + b$, then by Lemmas 2.10 and 2.11, τ would have at most $(a + b)^{2k+2}$ nodes. However, we assumed that $|V(G)| > (a + b)^{2k+2}$, a contradiction. $\square$

We can now finish the proof of Lemma 2.9. By Corollary 2.12, τ has depth at least $a + b$. Fix a node w of maximum depth in τ, and observe that w either contains at least a letters D or at least b letters S. In the first case, let A be the set of all vertices $\tau(u)$ for which $u\mathrm{D}$ is a prefix of w. Then $|A| \geq a$ and by construction A is a clique in G. In the second case we analogously find an independent set in G of size at least b. This concludes the proof of Lemma 2.9 and of Theorem 2.8.

3 Vapnik-Chervonenkis dimension

We now move to the analysis of graph classes in terms of structural properties of set systems associated with them.

Definition 3.1. A *set system* consists of a *ground set* (or *universe*) U and *family* $\mathcal{F} \subseteq 2^U$ consisting of subsets of U.

Set systems are often called *hypergraphs* in combinatorics, but it is the opinion of the author of this text that this is a really misleading name that does not reflect the nature of these objects, making them mere generalizations of graphs. Note that in the definition of a set system we do not require that the ground set is finite, though in this course we will essentially only work with finite set systems. Similarly as with classes of graphs, we can work with classes of set systems.

Let us make some example. For a graph G and $r \in \mathbb{N}$, we can define the set systems of radius-r balls in G as follows:

$$\mathrm{Balls}_r(G) = \{N_r^G[u] \colon u \in V(G)\} = \{\{v \colon \mathrm{dist}(u, v) \leq r\} \colon u \in V(G)\}.$$

We often write $\mathrm{Balls}(\cdot)$ for $\mathrm{Balls}_1(\cdot)$. For a class of graphs $\mathcal{C}$ and $r \in \mathbb{N}$, we can consider the class of set systems $\mathrm{Balls}_r(\mathcal{C})$.

Similarly as for (classes of) graphs, we can consider structural properties of (classes of) set systems. For instance, consider the following definition of *order dimension*. Here, for a set system $(U, \mathcal{F})$ and a subset of universe $A \subseteq U$, the *set system induced by A* is the set system $(A, \mathcal{F}[A] = \{F \cap A \colon F \in \mathcal{F}\})$.

Definition 3.2. The *order dimension* of a set system $(U, \mathcal{F})$ is defined as the maximum over all $A \subseteq U$ of the length of the longest chain in $\mathcal{F}[A]$, regarded as partially ordered by inclusion. The *order dimension* of a class $\mathcal{D}$ of set systems is the supremum of the order dimensions of the members of $\mathcal{D}$.

Then we have the following interesting connection.

Theorem 3.3. *A class of graphs $\mathcal{C}$ has a finite ladder index if and only if the class of set systems* $\mathrm{Balls}(\mathcal{C})$ *has a finite order dimension.*

Proof. Take any $G \in \mathcal{C}$. Suppose $(a_1, \ldots, a_\ell), (b_1, \ldots, b_\ell)$ is a ladder in G. Then taking $A = \{a_1, \ldots, a_\ell\}$, we find that the set system induced by A in $\mathrm{Balls}(\mathcal{C})$ has a chain of length ℓ, as witnessed by the neighborhoods of $b_1, \ldots, b_\ell$ intersected with A. This proves that the ladder index of $\mathcal{C}$ is at most the order dimension of $\mathrm{Balls}(\mathcal{C})$. For the converse implication, if for some $A \subseteq V(G)$, the set system induced by A has a chain $A_0 \subsetneq A_1 \subsetneq \ldots \subsetneq A_{k-1}$ of length k, where $A_i = N[b_i] \cap A$ for some vertices b_i, then taking any $a_i \in A_i - A_{i-1}$ for $i = 1, 2, \ldots, k-1$ we find a ladder in G of length $k - 1$. This proves tha that the order dimension of $\mathrm{Balls}(\mathcal{C})$ is at most the ladder index of $\mathcal{C}$ plus one. ☐

Now it follows from Theorem 2.3 that for every nowhere dense class of graphs $\mathcal{C}$ and $r \in \mathbb{N}$, the class of set systems

$$\mathrm{Balls}_r(\mathcal{C}) = \mathrm{Balls}(\mathcal{C}^r)$$

has a finite order dimension. Thus, intuitively we may study structural and algorithmic properties of graphs by analyzing structural and algorithmic properties of associated set systems. The algorithmic connection will be explored further in the next section, but we now proceed to the probably most important structural measure of set systems: the *VC dimension.*

The concept of VC dimension was introduced in a seminal paper of Vapnik and Chervonenkis on statistical machine learning. However, related ideas were independently developed by Sauer in combinatorics and Shelah in model theory.

Definition 3.4. Let $(U, \mathcal{F})$ be a set system. We say that a set $A \subseteq U$ is *shattered* by $\mathcal{F}$ if for every $B \subseteq A$ there exists $F \in \mathcal{F}$ such that $B = A \cap F$. The *Vapnik-Chervonenkis dimension*, or short *VC dimension*, of $(U, \mathcal{F})$ is the largest size of a subset of U that is shattered by $\mathcal{F}$.

The following observation follows directly from the definitions.

Lemma 3.5. *Every set system of order dimension at most k has also VC dimension at most $k-1$.*

Proof. It suffices to observe that for a set system $(U, \mathcal{F})$, if $A \subseteq U$ is shattered by $\mathcal{F}$, then $\mathcal{F}[A]$ contains a chain of length $|A| + 1$. ☐

Thus, VC dimension is a more general concept than order dimension, hence it is suited for the analysis of even more general classes of set systems than those induced by balls in sparse graphs. In fact, many set systems originating in geometric settings, e.g. set systems induced by well-behaved objects (half-spaces, balls, boxes) in Euclidean spaces tend to have bounded VC dimension. The example important for us is that of set systems of bounded-radius balls in graphs from nowhere dense classes. Precisely, by Theorem 2.3, Theorem 3.3, and Lemma 3.5 we immediately obtain the following.

Corollary 3.6. *Let $\mathcal{C}$ be a nowhere dense class of graphs. Then for every $r \in \mathbb{N}$ there exists $k(r) \in \mathbb{N}$ such that for every $G \in \mathcal{C}$, the set system* $\mathrm{Balls}_r(G)$ *has order dimension at most k, hence also VC dimension at most k.*

One of the main tools when working with VC dimension is the following lemma, which was independently discovered by several authors. It is usually called the *Sauer-Shelah Lemma.*

Lemma 3.7. *Let $\mathcal{F}$ be a set system of VC dimension k on a ground set U of size n. Then*

$$|\mathcal{F}| \leq \sum_{i=0}^{k} \binom{n}{i} \leq 1 + n^k.$$

Proof. We prove that $\mathcal{F}$ shatters at least $|\mathcal{F}|$ different subsets of the ground set U. This immediately implies the lemma, as only $\sum_{i=0}^{k} \binom{n}{i}$ of the subsets of A have cardinality at most k. We prove the claim by induction on $|\mathcal{F}|$. The base case is clear, as every set family shatters the empty set.

Now assume that $\mathcal{F}$ contains at least two sets and assume that the claim holds for all families of size smaller than $|\mathcal{F}|$. As $\mathcal{F}$ contains at least two sets, there exists $x \in U$ that belongs to some but not all of the sets in $\mathcal{F}$. We split $\mathcal{F}$ into two subfamilies $\mathcal{X} = \{F \in \mathcal{F} \colon x \in F\}$ and $\mathcal{Y} = \{F \in \mathcal{F} \colon x \notin F\}$. By induction assumption, $\mathcal{X}$ shatters at least $|\mathcal{X}|$ sets and $\mathcal{Y}$ shatters at least $|\mathcal{Y}|$ sets. However, it may be the case that some set $A \subseteq U$ is shattered by both subfamilies. Note that such a set A cannot contain x, since a set that contains x cannot be shattered by a subfamily in which all sets contain x, nor by a subfamily in which all sets do not contain x. Hence, both A and $A \cup \{x\}$ are shattered by $\mathcal{F}$. This gives us for each set that is shattered both by $\mathcal{X}$ and $\mathcal{Y}$ two sets that are shattered in $\mathcal{F}$, one of which is shattered neither by $\mathcal{X}$ nor by $\mathcal{Y}$. So the number of sets shattered by $\mathcal{F}$ is at least $|\mathcal{X}| + |\mathcal{Y}| = |\mathcal{F}|$. ☐

From Sauer-Shelah Lemma we can immediately derive a polynomial bound on the number of distance-r profiles in nowhere dense classes.

Corollary 3.8. *Let $\mathcal{C}$ be a nowhere dense class and $r \in \mathbb{N}$. Then there exists polynomial $p(\cdot)$, depending only on $\mathcal{C}$ and r, such that for every $G \in \mathcal{C}$ and $A \subseteq V(G)$, the number of different function from A to $\{0, 1, \ldots, r, \infty\}$ realized as distance-r profiles on A is at most $p(|A|)$.*

Proof. Consider any $p \leq r$. By Theorems 2.3 and 3.3, the class $\mathrm{Balls}_p(\mathcal{C})$ has a finite order dimension, so it also has a finite VC dimension, say $\ell_p \in \mathbb{N}$. Let $G \in \mathcal{C}$ and $A \subseteq V(G)$. Then the set system $\mathrm{Balls}_p(G)$ restricted to A also has VC dimension at most ℓ_p, so by the Sauer-Shelah Lemma it follows that

$$|\mathrm{Balls}_p(G)[A]| \leq 1 + |A|^{\ell_p} \qquad \text{for all } p \leq r.$$

Observe that for any $u \in V(G)$, the distance-r profile of u on A is uniquely defined by the tuple

$$(N_0^G[u] \cap A, N_1^G[u] \cap A, N_2^G[u] \cap A, \ldots, N_r^G[u] \cap A).$$

By the previous bound, the number of different such tuples is bounded by

$$\left(1 + |A|^{\ell_0}\right)\left(1 + |A|^{\ell_1}\right)\left(1 + |A|^{\ell_2}\right) \ldots \left(1 + |A|^{\ell_r}\right),$$

which is a polynomial in $|A|$. ☐

In fact, similarly to the bounded expansion case, in nowhere dense classes one can even prove almost linear bounds on the neighborhood complexity, not just polynomial as asserted by Corollary 3.8. We present a proof of this statement in the next section.

4 Neighborhood complexity in nowhere dense classes

We will prove the following strenghtening of Corollary 3.8.

Theorem 4.1. *Let $\mathcal{C}$ be a nowhere dense class, $r \in \mathbb{N}$, and $\varepsilon > 0$. Then there exists a constant $c \in \mathbb{N}$, depending only on $\mathcal{C}$, r, ε, such that for every $G \in \mathcal{C}$ and nonempty $A \subseteq V(G)$, the number of different function from A to $\{0, 1, \ldots, r, \infty\}$ realized as distance-r profiles on A is at most $c \cdot |A|^{1+\varepsilon}$.*

The main idea of the proof is to follow the same strategy we applied in the bounded expansion case (Theorem 6.3 of Chapter 2), namely to use generalized coloring numbers. In the proof in the bounded expansion case, we bounded the number of distance-r profiles by $c \cdot |A|$ for some constant c that depended on the number $\mathrm{wcol}_{2r}(G)$. While $\mathrm{wcol}_{2r}(G)$ can be bounded by a constant when G belongs to a fixed class of bounded expansion, in nowhere dense classes we can only claim a bound of the form $\mathcal{O}(n^{\varepsilon})$ for any $\varepsilon > 0$. This would be sufficient to obtain a bound of the form $\mathcal{O}(|A| \cdot n^{\varepsilon})$ on the number of realized distance-r profiles on A, provided in the proof of the bounded expansion case, c depended polynomially on $\mathrm{wcol}_{2r}(G)$. Unfortunately, a careful inspection of the proof shows that there are two places where an exponential explosion takes place. Our goal is to replace, in both places, this exponential explosion with a polynomial one using an argument based on the Sauer-Shelah Lemma. For this to work, we need the following important result: in nowhere dense classes, the set systems of weak reachability sets have a bounded VC dimension.

Theorem 4.2. *For every nowhere dense class $\mathcal{C}$ and $r \in \mathbb{N}$, there exists $k \in \mathbb{N}$ such that for every graph $G \in \mathcal{C}$ and vertex ordering σ, the set system*

$$\mathcal{W}_{r,\sigma} = \{\, \mathrm{WReach}_r[G, \sigma, u] \ :\ u \in V(G) \,\}$$

has VC dimension at most k.

Proof. As G and σ will be fixed throughout the proof, for brevity we write $\mathrm{WReach}_r[u]$ instead of $\mathrm{WReach}_r[G, \sigma, u]$. Since $\mathcal{C}$ is nowhere dense, it is also uniformly quasi-wide, say with margins $(s_p)_{p\in\mathbb{N}}$ and wideness functions $(N_p(\cdot))_{p\in\mathbb{N}}$.

Suppose a set of vertices A is shattered by $\mathcal{W}_{r,\sigma}$, where

$$|A| \geq N_{2r}(m)$$

for some $m \in \mathbb{N}$ to be fixed later. By uniform quasi-wideness, we can find disjoint sets of vertices S and $B \subseteq A$ such that

- $|S| \leq s_{2r} =: s$;
- $|B| = m$; and
- B is distance-$2r$ independent in $G - S$.

Since A is shattered by $\mathcal{W}_{r,\sigma}$ and $B \subseteq A$, we also have that B is shattered by $\mathcal{W}_{r,\sigma}$.

For a vertex u, let the *signature* of u be the pair (X_u, δ_u) defined as follows:

- $X_u \coloneqq B \cap \mathrm{WReach}_r[G - S, \sigma, u]$; here, we treat σ as restricted to the vertices of $G - S$.

- Function $\delta_u\colon \{0,1,\ldots,r-1\}\times S \to B\cup\{\bot\}$ is defined as follows: for $p<r$ and $w\in S$, $\delta_u(p,w)$ is the largest in σ vertex $v\in B$ such that there exists a path of length at most p from u to w whose all vertices are not smaller in σ than v, or $\delta_u(p,w)=\bot$ if no such vertex exists.

We now observe two basic properties. First, the number of possible signatures is polynomial in m. Second, the signature of a vertex u uniquely defines the set $\mathrm{WReach}_r[u]\cap B$.

Claim 1. *The vertices of G have at most $(m+1)^{1+rs}$ different signatures.*

Proof. Observe that since B is distance-$2r$ independent in $G-S$, for every vertex u we have that X_u consists of at most one vertex of B, for otherwise two different vertices of B would be at distance at most $2r$ from each other in $G-S$. Therefore, there are $m+1$ possible choices of X_u. Since the domain and the co-domain of δ_u have sizes at most rs and $m+1$ respectively, there are at most $(m+1)^{rs}$ possible choices of δ_u. ⌟

Claim 2. *If $u,u'\in V(G)$ have the same signatures, then*

$$\mathrm{WReach}_r[u]\cap B = \mathrm{WReach}_r[u']\cap B.$$

Proof. By symmetry, it suffices to prove that if for some $v\in B$ we have $v\in \mathrm{WReach}_r[u]$, then also $v\in\mathrm{WReach}_r[u']$. Let P be a path witnessing that $v\in\mathrm{WReach}_r[u]$. If P does not traverse any vertex of S, then P also witnesses that $v\in\mathrm{WReach}_r[G-S,\sigma,u]$, implying that $v\in X_u$. Since $X_u=X_{u'}$, we also have $v\in X_{u'}$. Hence $v\in\mathrm{WReach}_r[u']$, as claimed.

Therefore, from now on we may assume that P intersects S, say at a vertex w. Let Q and R be the prefix of P from u to w and the suffix of P from w to v, respectively. Since S and B are disjoint, we have that $w\neq v$. Hence Q has length smaller than r, say $p<r$. As all vertices traversed by P are not smaller in σ than v, the same holds also for Q. Since $\delta_u(p,w)=\delta_{u'}(p,w)$, we infer that there exists a path Q' from u to w of length at most p such that all the vertices of Q' are not smaller than v in σ. Now the concatenation of Q' and R witnesses that $v\in\mathrm{WReach}_r[u']$, and we are done. ⌟

Now we can combine Claims 1 and 2 to conclude that

$$|\{\mathrm{WReach}_r[u]\cap B\colon u\in V(G)\}|\leq (m+1)^{1+rs}.$$

On the other hand, we have assumed that B is shattered by $\mathcal{W}_{r,\sigma}$, hence the set system on the left hand side of the above is in fact the power set of B. We conclude that

$$2^m\leq (m+1)^{1+rs}.$$

Since the left hand side of the above inequality grows exponentially with m, while the right hand side grows only polynomially, there is some m_0 depending on r and s (so also only on r and $\mathcal{C}$) such that we necessarily have $|B|=m\leq m_0$. This implies that the shattered set A we started with has to satisfy $|A|\leq N_{2r}(m_0)$. We conclude that the VC dimension of $\mathcal{W}_{r,\sigma}$ is bounded by $N_{2r}(m_0)$. □

With Theorem 4.2 settled, we can proceed to the next step: proving a slightly weaker bound of the form $\mathcal{O}(|A|\cdot n^{\varepsilon})$. The proof of this statement uses the same strategy as the proof of Theorem 6.3 from Chapter 2, so we only sketch the important differences. In essence, the idea is that in the proof in the bounded expansion case contains an exponential explosion in two places, and these are reduced to polynomial explosions using Theorem 4.2 in one case, and Corollary 3.8 in the other.

Lemma 4.3. *Let $\mathcal{C}$ be a nowhere dense class, $r \in \mathbb{N}$, and $\varepsilon > 0$. Then there exists a constant c, depending only on $\mathcal{C}$, r, and ε, such that for every $G \in \mathcal{C}$ and nonempty $A \subseteq V(G)$, the number of different function from A to $\{0, 1, \ldots, r, \infty\}$ realized as distance-r profiles on A is at most $c \cdot |A| \cdot n^{\varepsilon}$, where $n = |V(G)|$.*

Proof sketch. By Theorem 3.4 from Chapter 2, there exists a vertex ordering σ of G such that

$$\mathrm{wcol}_{2r}(G, \sigma) \leq c_0 \cdot n^{\varepsilon}$$

for some constant c_0 depending only on $\mathcal{C}$, r, and ε. The reasoning presented in the proof of Theorem 6.3 from Chapter 2 shows that the number of different distance-r profiles realized on A is bounded by

$$1 + d \cdot 2^{d-1} \cdot (r+2)^d \cdot |A|,$$

where $d = \mathrm{wcol}_{2r}(G, \sigma)$. Let us recall from where the consecutive factors of this bound come from.

The factor $(r+2)^d$ comes from the straightforward upper bound on the number of distance-r profiles on a fixed local separator, which is a set X of size at most $\mathrm{wcol}_r(G, \sigma) \leq d$. Observe, however, that if $p(\cdot)$ is the polynomial provided by Corollary 3.8 for $\mathcal{C}$ and r, then in fact the number of distance-r profiles on X is bounded by $p(d)$.

The factor $d \cdot 2^{d-1} \cdot |A|$ comes from bounding the number of different local separators X. Recall that the local separator of u, namely $X = X[u]$, is defined as

$$X[u] \coloneqq B \cap \mathrm{WReach}_r[G, \sigma, u],$$

where

$$B \coloneqq \bigcup_{a \in A} \mathrm{WReach}_r[G, \sigma, a]$$

is a superset of A consisting of at most $d \cdot |A|$ vertices. In the proof of Theorem 6.3 in Chapter 2, see Claim 5 there, we have argued that each such local separator X satisfies

$$X \subseteq B \cap \mathrm{WReach}_{2r}[G, \sigma, b],$$

where b is the largest in σ vertex of X. This directly gave an upper bound of $|B| \cdot 2^{d-1} \leq d \cdot 2^{d-1} \cdot |A|$ on the number of different local separators.

Now observe that for a fixed $b \in B$, if we denote $Z_b \coloneqq B \cap \mathrm{WReach}_{2r}[G, \sigma, b]$, then the number of different sets of the form

$$Z_b \cap \mathrm{WReach}_r[G, \sigma, u] \qquad \text{for some } u \in V(G)$$

is bounded by $1 + |Z_b|^k \leq 1 + d^k$ for some constant k depending only on $\mathcal{C}$ and r. Indeed, this follows from Theorem 4.2 combined with the Sauer-Shelah Lemma (Lemma 3.7): the VC dimension of the set system $\{\mathrm{WReach}_r[G, \sigma, u] \colon u \in V(G)\}$ is bounded by a constant k depending only on $\mathcal{C}$ and r, hence the cardinality of the set system $\{\mathrm{WReach}_r[G, \sigma, u] \cap Z_b \colon u \in V(G)\}$ is bounded by $1 + |Z_b|^k \leq 1 + d^k$. We conclude that the bound on the number of local separators X can be replaced with a sharper estimation of

$$|B| \cdot (1 + d^k) \leq d \cdot (1 + d^k) \cdot |A|.$$

We conclude that the total number of different distance-r profiles realized on A is bounded by

$$1 + d \cdot (1 + d^k) \cdot p(d) \cdot |A| = 1 + q(d) \cdot |A|,$$

where $q(d) \coloneqq d \cdot (1 + d^k) \cdot p(d)$ is a polynomial that depends only on $\mathcal{C}$ and r. Since $d \leq c_0 \cdot n^\varepsilon$, we conclude that the considered number of profiles is bounded by

$$1 + |A| \cdot q(c_0 \cdot n^\varepsilon).$$

It now remains to rescale ε by the degree of $q(\cdot)$, that is, redo the whole reasoning for $\frac{\varepsilon}{\ell}$ instead of ℓ, where ℓ is the degree of $q(\cdot)$. □

The bound provided by Lemma 4.3 is almost as we wanted, with the exception that instead of factor $|A|^\varepsilon$ we have obtained factor n^ε, for any $\varepsilon > 0$. Note that these factors are essentially the same provided that n would be polynomially bounded in $|A|$. Indeed, if ℓ would be the degree of such a polynomial bound, then we would have $n^\varepsilon \leq \mathcal{O}(|A|^{\ell\varepsilon})$, so again rescaling ε by a factor of ℓ would suffice. The following simple lemma, which is implied by Corollary 3.8, shows that we can reduce the problem to this case.

Lemma 4.4. *Let $\mathcal{C}$ be a nowhere dense class and $r \in \mathbb{N}$. Then there exists a polynomial $q(\cdot)$ such that for every graph $G \in \mathcal{C}$ and $A \subseteq V(G)$, there exists a set of vertices $D \supseteq A$ satisfying the following properties:*

- *$|D| \leq q(|A|)$; and*
- *every function from A to $\{0, 1, \ldots, r, \infty\}$ that realized as a distance-r profile on A in G is also realized as a distance-r profile on A in $G[D]$.*

Proof. Let $p(\cdot)$ be the polynomial provided by Corollary 3.8 for $\mathcal{C}$ and r. For every function $\pi \colon A \to \{0, 1, \ldots, r, \infty\}$ that is realized as a distance-r profile on A in G, let us pick any vertex u_π such that

$$\mathrm{profile}_r[u_\pi] = \pi.$$

Let D_π be a set of vertices constructed as follows: for every $a \in A$ satisfying $\alpha(a) < \infty$, select any shortest path from u_π to a and add all its vertices in D_π. Clearly, $|D_\pi| \leq 1 + r|A|$.

We define

$$D \coloneqq A \cup \bigcup_{\pi \text{ is realized}} D_\pi,$$

where the union is over all distance-r profiles on A realized in G. Since their number is bounded by $p(|A|)$ by Corollary 3.8, we have

$$|D| \leq |A| + (1 + r|A|) \cdot p(|A|),$$

which is a polynomial in $|A|$. It now remains to observe that for every profile π realized in G, the profile of u_π on A is the same in $G[D]$ as it was in G. □

We can now wrap up the proof of Theorem 4.1.

Proof of Theorem 4.1. By closing $\mathcal{C}$ under taking subgraphs, we may assume that $\mathcal{C}$ is subgraph-closed; note that this does not spoil the nowhere denseness of $\mathcal{C}$. Given G and $A \subseteq V(G)$, apply Lemma 4.4 to G and A, thus obtaining a suitable superset $D \supseteq A$ satisfying $|D| \leq q(|A|)$. Noting that $G[D] \in \mathcal{C}$, we may apply Lemma 4.3 to the graph $G[D]$ and the set A in it. Thus we infer that the number of different distance-r profiles on A in $G[D]$ is at most

$$c' \cdot |A| \cdot |D|^{\varepsilon} \leq c' \cdot |A| \cdot q(|A|)^{\varepsilon},$$

for some constant c'. By Lemma 4.4, the above quantity is also an upper bound on the number of different distance-r profiles on A in G. It now remains to rescale ε by the degree of $q(\cdot)$ and choose the constant c appropriately. □

5 Approximating hitting sets

A *hitting set* in a set system $\mathcal{F}$ over a ground set A is a set $H \subseteq A$ such that $H \cap F \neq \emptyset$ for all $F \in \mathcal{F}$. In this section we will consider the Hitting Set problem for set systems: given a set system $\mathcal{F}$ over A, compute the smallest hitting set for $\mathcal{F}$. To see the motivation of this problem, observe that if G is a graph and $\mathrm{Balls}_r(G)$ is the set system of balls of radius r in G, which is a set system over $V(G)$, then hitting sets in $\mathrm{Balls}_r(G)$ are exactly distance-r dominating sets in G.

The Hitting Set problem in general is NP-complete and so-called W[2]-complete when parameterized by the solution size, which means that an algorithm deciding whether there is a hitting set of size k in time $f(k) \cdot (|A| + |\mathcal{F}|)$ is unlikely, for any function k. In this section we will consider approximation algorithms for Hitting Set: given a set system $\mathcal{F}$ over A, we wish to find a hitting set that maybe is not optimum, but whose size is not far from the optimum. We will first discuss a very simple greedy algorithm which achieves an approximation factor of $\ln n$. We will then consider set systems of bounded VC dimension, where a better approximation factor can be obtained.

In the following we use the following notation: for a set system $\mathcal{F}$, by $\tau(\mathcal{F})$ we denote the smallest size of a hitting set in $\mathcal{F}$.

5.1 Greedy approximation

Consider the following greedy algorithm. Starting with an empty hitting set H, iteratively add elements of A to H according to the following greedy rule: in each round, choose the element $a \in A$ that hits the largest number of sets in $\mathcal{F}$ which still have to be hit.

Theorem 5.1. *Let $\mathcal{F}$ be a set system of size $|\mathcal{F}| = m$. Then the greedy algorithm outputs a hitting set of $\mathcal{F}$ of size at most $\tau(\mathcal{F}) \cdot \ln m$.*

Proof. Let A the ground set of $\mathcal{F}$, let $k = \tau(\mathcal{F})$, and let $H \subseteq A$ be a hitting set of $\mathcal{F}$ of size k. Then there exists an element $a \in H$ which hits at least m/k sets of $\mathcal{F}$ (otherwise $\mathcal{F}$ cannot be hit by k elements). Hence, in the first round the greedy algorithm will choose an element $b_1 \in A$ which hits at least m/k sets of $\mathcal{F}$. Hence, after the first round of the algorithm there remain at most

$$m_1 = m - \frac{m}{k} = m \cdot \left(1 - \frac{1}{k}\right)$$

sets to be hit. Of course, H is also a hitting set for the sets that remain to be hit, so we can argue just as above that there exists an element $b_2 \in A$ which hits at least m_1/k of the remaining sets.

Hence, after the second round, it remains to hit at most $m_2 = m_1 - m_1/k$ sets. Now observe that $m_2 = m_1 - m_1/k \leq m_1 \cdot (1 - 1/k) \leq m \cdot (1 - 1/k)^2$. We can repeat this argumentation and conclude that after executing i steps of the greedy algorithm it remains to hit at most

$$m_i = m \cdot \left(1 - \frac{1}{k}\right)^i$$

sets. Let us determine for what value of i we have $m_i < 1$, as then we are sure that in fact all sets are hit and the algorithm has already terminated. We have

$$m_i = m \cdot \left(1 - \frac{1}{k}\right)^i < m \cdot e^{-\frac{i}{k}},$$

where the last inequality follows from the bound $1 - x < e^{-x}$, which holds for all $x > 0$. Thus, for $i \geq k \ln m$ we have $m_i < m \cdot e^{-\ln m} = 1$. We conclude that the greedy algorithm terminates after at most $k \ln m$ steps, in particular, it computes a hitting set of size at most $k \ln m$. □

We remark that in general set systems, the approximation ratio of Theorem 5.1 is essentially tight: under $\mathsf{P} \neq \mathsf{NP}$, there is no polynomial-time approximation algorithm achieving approximation factor $\alpha \ln n$ for any $\alpha < 1$. In the next section we show that if we assume that the set system in question has bounded VC dimension, then the approximation factor can be drastically improved.

5.2 Approximating hitting sets in set systems of bounded VC dimension

In this section we will prove the following theorem.

Theorem 5.2. *Let $d \geq 2$ be a fixed integer. There exists a randomized polynomial-time algorithm which given a set system $\mathcal{F}$ of VC dimension at most d, computes a hitting set for $\mathcal{F}$ that has size at most $\mathcal{O}(d \cdot \tau(\mathcal{F}) \ln \tau(\mathcal{F}))$ with probability at least $\frac{1}{2}$.*

In other words, we improve the approximation ratio from $\ln m$ to $\mathcal{O}(d \cdot \ln \tau(\mathcal{F}))$. Observe that the error probability can be made arbitrarily close to 0 by repeating the algorithm several times and choosing the smallest output.

We now proceed with building up tools for the proof of Theorem 5.2. A *probability distribution* on a set A is a mapping $\mu \colon A \to [0,1]$ such that $\sum_{a \in A} \mu(a) = 1$. For a set $B \subseteq A$, let $\mu(B) = \sum_{b \in B} \mu(b)$. The following definition of an ϵ-*net* is vital for our approach.

Definition 5.3. Let $\mathcal{F}$ be a set family over a ground set A, let μ be a probability distribution on A and let $\epsilon > 0$. A set $H \subseteq A$ is an ϵ-*net* with respect to μ if $H \cap F \neq \emptyset$ for all $F \in \mathcal{F}$ with $\mu(F) \geq \epsilon$.

Thus, an ϵ-net is the same as the hitting set for the subsystem of $\mathcal{F}$ consisting of the sets having measure (probability) at least ε.

We will now show that a small ϵ-net can be found by simply sampling enough elements according to the probability distribution μ. Precisely, a *sample* of size ℓ from μ is an ℓ-tuple of elements of A sampled independently at random from the distribution μ. Note that we allow to draw an element multiple times in order to to make calculations simpler.

Theorem 5.4. *Let $\mathcal{F}$ be a set system of VC dimension $d \geq 2$ and let μ be a probability distribution on the ground set A. There exists a universal constant c such that for every $0 < \epsilon \leq 1/2$, a random sample from μ of size at least $c \cdot \frac{d}{\epsilon} \ln \frac{1}{\epsilon}$ is an ϵ-net with probability at least $\frac{1}{2}$.*

In the proof we will use the following variant of Chernoff's bound.

Lemma 5.5. *Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables with success probability p. Let $X = X_1 + \ldots + X_n$ and let $\mu = np$ be the expected value of X. Then for every $\delta \in [0,1]$, we have*

$$\mathbb{P}[X \geq (1-\delta)\mu] \geq 1 - e^{-\frac{\delta^2 \mu}{2}}.$$

In particular, for $\delta = \frac{1}{2}$ we have

$$\mathbb{P}[X \geq \mu/2] \geq 1 - e^{-\frac{\mu}{8}},$$

which is larger than $\frac{1}{2}$ for $\mu \geq 8$.

Proof of Theorem 5.4. Let $s = c \cdot \frac{d}{\epsilon} \ln \frac{1}{\epsilon}$ for a universal constant c to be defined at the end of the proof. We assume without loss of generality that s is an integer. Let N be a sample of size s from the probability distribution μ. We shall prove prove that N is an ϵ-net with probability at least $\frac{1}{2}$.

Without loss of generality we may assume that all $F \in \mathcal{F}$ satisfy $\mu(F) \geq \epsilon$, since we only want to hit those sets F which satisfy the above inequality. For a fixed $F \in \mathcal{F}$, the probability that the random sample N does not hit F is at most $(1-\epsilon)^s \leq e^{-s\epsilon}$. Let E_1 be the event that N fails to be an ϵ-net, that is, that N does not hit some $F \in \mathcal{F}$. We bound $\mathbb{P}[E_1]$ from above based on the following second random experiment.

We draw another sample M of size s from μ, and let it be independent of N. Let $k = s\epsilon/2$; again, assume without loss of generality that k is an integer. Let E_2 be the event

there exists $F \in \mathcal{F}$ such that no element of N belongs to F,
while at least k elements of M belong to F.

Note that we treat M as tuple of s elements from A, so if an element $e \in F$ is sampled i times in M, it contributes i to the number of elements of M belonging to F. By somehow abusing the notation, by $|M \cap F|$ we will denote the number of elements of M belonging to F, counted in the manner described above. Clearly, $\mathbb{P}[E_2] \leq \mathbb{P}[E_1]$, since E_2 in particular requires E_1 to occur. We are first going to show that $\mathbb{P}[E_2] \geq \frac{1}{2}\mathbb{P}[E_1]$.

Consider the conditional probability $\mathbb{P}[E_2 \mid N]$, i.e., the probability that E_2 occurs for N fixed and M random[1]. If N is an ϵ-net then E_2 cannot occur, hence in this case $\mathbb{P}[E_1 \mid N] = \mathbb{P}[E_2 \mid N] = 0$. So suppose that there exists $F \in \mathcal{F}$ with no element of N belonging to F. There may be several such sets, fix one of them and denote it by F_N. We have $\mathbb{P}[E_2 \mid N] \geq \mathbb{P}[|M \cap F_N| \geq k]$. Now, the quantity $|M \cap F_N|$ is a sum of s independent Bernoulli random variables with success probability at least ϵ, so by applying Lemma 5.5 for $n = s$, $p = \epsilon$, and $\mu = s\epsilon = cd \ln \frac{1}{\epsilon}$, we have

$$\mathbb{P}[|M \cap F_N| \geq k] \geq 1 - e^{-\frac{\mu}{8}}.$$

By requiring c to satisfy $c \cdot 2 \cdot \ln \frac{1}{2} \geq 8$, we ensure that $\mu \geq 8$ and hence $\mathbb{P}[|M \cap F_N| \geq k] \geq 1/2$. Hence $\mathbb{P}[E_1 \mid N] \leq 2 \cdot \mathbb{P}[E_2 \mid N]$ for all fixed N and thus

$$\mathbb{P}[E_1] \leq 2\mathbb{P}[E_2],$$

[1] Formally, $\mathbb{P}[E_2 \mid N]$ is an N-measurable random variable, but since we are dealing with discrete probabilistic spaces, the reader may think of it as a function that assigns to each potential outcome of sampling N the probability that E_2 occurs conditioned on this outcome.

as claimed.

Now we are going to bound $\mathbb{P}[E_2]$ differently. Instead of choosing N and M at random directly as above, we first draw a sample $S = (a_1, \ldots, a_{2s})$ of size $2s$ from the distribution μ. Then we randomly choose s positions between 1 and $2s$ (without repetition; however, note that S may contain the same element multiple times) and define N as the s-tuple of elements at these positions in the sequence S, and M as the s-tuple of remaining elements. Hence there are exactly $\binom{2s}{s}$ choices for N and M for a fixed sequence S, and the resulting distribution of N and M is exactly the same as in our first experiment. We now prove that for every fixed sequence S, the conditional probability $\mathbb{P}[E_2 \mid S]$ is small. This implies that $\mathbb{P}[E_2]$ is small, and therefore $\mathbb{P}[E_1]$ is small as well.

So fix a sequence S as above. Let $F \in \mathcal{F}$ be a fixed set and consider the conditional probability

$$p_F = \mathbb{P}[N \cap F = \emptyset \text{ and } |M \cap F| \geq k \mid S].$$

If $|S \cap F| < k$ then $p_F = 0$. Otherwise, we have $p_F \leq \mathbb{P}[N \cap F = \emptyset \mid S]$. The latter is the probability that a random sample of s positions out of $2s$ positions from S avoids the at least k positions occupied by elements of F. This probability is bounded from above by

$$\frac{\binom{2s-k}{s}}{\binom{2s}{s}} \leq \left(1 - \frac{k}{2s}\right)^s \leq e^{-(k/2s)s} = e^{-k/2} = e^{-cd \ln \frac{1}{\epsilon}/4} = \epsilon^{cd/4}.$$

Finally we use that $\mathcal{F}$ has bounded VC dimension. According to the Sauer-Shelah lemma, Lemma 3.7, the number of distinct intersections of $\mathcal{F}$ with the sequence S is at most

$$\binom{2s}{0} + \binom{2s}{1} + \ldots + \binom{2s}{d} \leq (d+1) \cdot \binom{2s}{d} \leq 2d \left(\frac{2es}{d}\right)^d;$$

here we used the fact that $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ for all $1 \leq k \leq n$. Since the event that $N \cap F = \emptyset$ and $|M \cap F| \geq k$ depends only on $S \cap F$, it suffices to consider at most $2d\left(\frac{2es}{d}\right)^d$ distinct sets $F \in \mathcal{F}$ — those with pairwise distinct interesections with S — and by applying the union bound we infer

$$\mathbb{P}[E_2 \mid S] \leq 2d \left(\frac{2es}{d}\right)^d \cdot \epsilon^{cd/4} = 2d \left(2ce \cdot \frac{1}{\epsilon} \cdot \ln \frac{1}{\epsilon} \cdot \epsilon^{c/4}\right)^d \leq 2d \left(2ce \cdot \epsilon^{c/4-2}\right)^d.$$

Now, since $\epsilon \leq \frac{1}{2}$, for a sufficiently large constant c we have

$$2ce \cdot \epsilon^{c/4-2} \leq \frac{1}{16}.$$

Then the above probability is bounded from above by $2d \cdot 16^{-d}$, which is always smaller or equal to $\frac{1}{4}$ for $d \geq 2$. Since the above reasoning applies to any fixed S, it follows that

$$\mathbb{P}[E_2] \leq \frac{1}{4}$$

for such choice of c, implying

$$\mathbb{P}[E_1] \leq 2\mathbb{P}[E_2] \leq \frac{1}{2}.$$

This finishes the proof of the theorem. $\square$

Theorem 5.4 allows us to well approximate hitting sets using the approach of *linear programming*. Consider the following linear program which for a given set system $\mathcal{F}$ seeks for a probability distribution on the ground set A maximizing ϵ for which every set from $\mathcal{F}$ has probability at least ϵ.

<u>ϵ-net LP</u>

- **Variables:** μ_a for all $a \in A$, and ϵ
- **Objective:** maximize ϵ
- **Constraints:**
  - $\sum_{a \in F} \mu_a \geq \epsilon$ for all $F \in \mathcal{F}$;
  - $\mu_a \geq 0$ for all $a \in A$; and
  - $\sum_{a \in A} \mu_a = 1$.

Linear programming is polynomial-time solvable, hence we may solve the above ϵ-net LP in polynomial time, obtaining a probability distribution $\mu^\star$ on A and a value $\epsilon^\star > 0$ such that the following holds: $\mu^\star(F) \geq \epsilon^\star$ for each $F \in \mathcal{F}$. For now assume $\epsilon^\star \leq \frac{1}{2}$; we will treat the remaining corner case $\epsilon^\star > \frac{1}{2}$ later. Now we apply Theorem 5.4 to the distribution $\mu^\star$, yielding that a sample from $\mu^\star$ of size at least $c \cdot \frac{d}{\epsilon^\star} \ln \frac{1}{\epsilon^\star}$ is an $\epsilon^\star$-net with probability at least $\frac{1}{2}$. Since $\mu^\star(F) \geq \epsilon^\star$ for each $F \in \mathcal{F}$, being an $\epsilon^\star$-net is equivalent to being a hitting set for $\mathcal{F}$, so we have obtained a hitting set for H of size at most $c \cdot \frac{d}{\epsilon^\star} \ln \frac{1}{\epsilon^\star}$.

It now remains to relate $\epsilon^\star$ to $\tau(\mathcal{F})$, the optimum size of a hitting set of $\mathcal{F}$, to show that this sample is in fact bounded in terms for $\tau(\mathcal{F})$. For this, we consider the following linear program.

<u>Hitting set LP</u>

- **Variables:** x_a for all $a \in A$
- **Objective:** minimize $\sum_{a \in A} x_a$
- **Constraints:**
  - $\sum_{a \in F} x_a \geq 1$ for all $F \in \mathcal{F}$; and
  - $x_a \geq 0$ for all $a \in A$.

Observe that a minimum integral solution to the hitting set LP, that is, one where each variable takes only values in $\{0, 1\}$, corresponds exactly to a minimum hitting set. An optimal fractional solution for the hitting set LP is denoted by $\tau^\star(\mathcal{F})$. Clearly, we have $\tau^\star(\mathcal{F}) \leq \tau(\mathcal{F})$. We now observe that the ϵ-net LP and the hitting set LP are in fact the same LP, just scaled. This is made explicit in the following proposition, whose proof is straightforward.

Proposition 5.6. *If $(\mu_a)_{a \in A}$ and ϵ is a solution to the ϵ-net LP, then setting $x_a = \mu_a/\epsilon$ for $a \in A$ yields a solution to the hitting set LP of cost $1/\epsilon$. Conversely, if $(x_a)_{a \in A}$ is a solution to the hitting set LP, then setting $\mu_a = x_a / \sum_{a \in A} x_a$ and $\epsilon = 1/\sum_{a \in A} x_a$ yields a solution to the ϵ-net LP.*

Consequently, the optimum $\epsilon^\star$ of the ϵ-net LP is equal to $1/\tau^\star(\mathcal{F})$, the inverse of the optimum for the hitting set LP.

Thus we completed the proof of Theorem 5.2: the sample size $c \cdot \frac{d}{\epsilon^\star} \ln \frac{1}{\epsilon^\star}$ is in fact bounded by

$$cd \cdot \tau^\star(\mathcal{F}) \ln \tau^\star(\mathcal{F}) \leq cd \cdot \tau(\mathcal{F}) \ln \tau(\mathcal{F}),$$

as we wanted, and we have already argued that this sample is a hitting set for $\mathcal{F}$ with probability at least $\frac{1}{2}$. One missing detail that we did not discuss is what happens if it turns out that $\epsilon^\star > \frac{1}{2}$. Then we may apply Theorem 5.4 for $\epsilon = \frac{1}{2}$ instead of $\epsilon^\star$, yielding that a constant-size sample is a hitting set for $\mathcal{F}$ with probability at least $\frac{1}{2}$.

Note that we have proved in fact a stronger fact: the gap between integral and fractional hitting sets is bounded roughly by the logarithm of the latter in set systems of bounded VC dimension.

Corollary 5.7. *There exists a universal constant c such that for every set system $\mathcal{F}$ of VC dimension d and $\tau^\star(\mathcal{F}) \geq 2$, it holds that*

$$\tau(\mathcal{F}) \leq cd \cdot \tau^\star(\mathcal{F}) \ln \tau^\star(\mathcal{F}).$$

Finally, by applying Theorem 5.2 to the set system of radius-r balls in a graph from a nowhere dense class (which has bounded VC dimension by Corollary 3.6) we infer the following.

Corollary 5.8. *For every nowhere dense class of graphs $\mathcal{C}$ and every $r \in \mathbb{N}$ there exists a randomized polynomial-time algorithm that given a graph $G \in \mathcal{C}$ outputs a distance-r dominating set in G of size at most $\mathcal{O}(k \log k)$, where k is the minimum size of a distance-r dominating set in G.*

We remark that the algorithm of Theorem 5.2 can in fact be derandomized, yielding the same asymptotic bound on the approximation ratio. This is done by replacing Theorem 5.4 with a deterministic counterpart with the same asymptotic bounds.